

METSZETEK

TÁRSADALOMTUDOMÁNYI FOLYÓIRAT

CROSS-SECTIONS
SOCIAL SCIENCE JOURNAL

Vol. 10 (2021) No. 3



METSZETEK
Társadalomtudományi folyóirat
Vol. 10 (2021) No. 3

Főszerkesztő:

Dr. Czibere Ibolya

Szerkesztők:

Dr. Balogh László Levente, Dr. Csoba Judit, Dr. Fényes Hajnalka,
Dr. Mohácsi Márta, Dr. Rácz Andrea, Sipos Flórián

Szerkesztőbizottság:

KATZ, Katalin – The Hebrew University of Jerusalem
KOLOZSVÁRI, Orsolya – College of Coastal Georgia
NISTOR, Laura – Sapientia Erdélyi Magyar Tudományegyetem, Kolozsvár
PICCIONE, Vincenzo – Roma Tre University, Italy
URIZ, Maria Jesús – Public University of Navarre, Pamplona
VIRTO, Lucia Martínez – Public University of Navarre, Pamplona
WOODS, Ronald – University of Technology Sydney, Australia

Tanácsadó testület:

BOBKOV, Vyacheslav – Russian Science Academy, Moscow, Russia
FOX, Chris – Manchester Metropolitan University, United Kingdom
GEELHOED, Sandra – University of Applied Sciences Utrecht, Netherlands
HERRMANN, Peter – University of Eastern Finland, Kuopio, Finland; Corvinus University, Budapest, Hungary
HORAKOVA HIRSCHLER, Nicole – University of Ostrava, Czech Republic
JANKY, Béla – Budapest University of Technology, Hungary
KERÜLŐ, Judit – University of Nyíregyháza, Hungary
RAZUMOVA, Tatiana – Moscow State Lomonosov University, Russia
SIK, Endre – Eötvös Loránd University, Budapest, Hungary; TÁRKI Social Research Institute, Budapest, Hungary
SZABÓ, Márton – University of Miskolc, Hungary
SZÁNTÓ, Zoltán – Corvinus University of Budapest, Hungary
TIMÁR, Judit – Hungarian Academy of Sciences Centre for Economic and Regional Studies Research Center; Eötvös Loránd University, Hungary
WIKTORSKA-ŚWIĘCKA, Aldona – University of Wrocław, Poland

Olvasószerkesztő:

Dr. Szekeres Melinda

Tördelőszerkesztő:

Juhászné Marosi Edit

Webdesign:

Fábián Márton

Szerkesztőségi munkatársak:

Béres Zsuzsa
Dr. Molnár Éva

ISSN 2063-6415

Kiadja: a Debreceni Egyetem Bölcsészettudományi Kar
Felelős kiadó: Prof. Dr. Keményfi Róbert dékán
Webcím: metszetek.unideb.hu
E-mail: metszetek@arts.unideb.hu





Tartalomjegyzék

TEMATIKUS TANULMÁNYOK – Digitális szociológia

Vendégszerkesztők: DESSEWFFY TIBOR, SZABÓ MIKLÓS

Digitális szociológia – Kihívások és távlatok. Szerkesztői előszó (Dessewffy Tibor).....	3
KMETTY ZOLTÁN: Új utak a társadalom megismerésében. <i>A donáció alapú digitális adatgyűjtésben rejlő lehetőségek</i>	6
NÉMETH RENÁTA: A felügyelt gépi tanulás kihívásai a szociológiai alkalmazásokban.....	27
HEGEDŰS DÁNIEL: Magyar videoblogger hálózatok online	43
OLÁH ESZTER: A feltétel nélküli alapjövedelem megtételének empirikus elemzése YouTube hozzászólásokon keresztül.....	68
BENE MÁRTON: Mi a téma? Politikai témák és felhasználói reakciók szövegbányászati vizsgálata a 2018-as országgyűlési kampány politikusi Facebook-oldalain.....	94
SÜTŐ ANNA: Merre tovább, netgeneráció? <i>A magyar fiatalok életstílus-alapú szegmensei</i>	124
SIMON SÁRA: Közösségi média kommunikáció a digitális egészségügyi térben <i>A #cysticfibrosis és a #Asthma Big Data összehasonlító elemzése</i>	143
MÁTÉ FANNI: A depresszió keretezési módjainak vizsgálata online fórumokon természetesnyelv-feldolgozással	181

RECENZÍÓ

A. GERGELY ANDRÁS: Studia Cingarorum – romológia a gyakorlati tudásforgalmazásban	209
---	-----

OLVASS FELESLEGESET!

SIK ENDRE ÁLLANDÓ ROVATA

JUHÁSZNÉ FAZEKAS ZSUZSANNA: Halmazódó hátrányok? – avagy cseh és szlovák roma bevándorlók mindennapjai Glasgowban	214
Abstracts	221



Contents

TEMATICS ARTICLES – Digital Sociology

KMETTY, ZOLTÁN: New ways in exporting Society. The potential of donation <i>based digital data collection</i>	6
NÉMETH, RENÁTA: The challenges of supervised machine learning in sociological applications.....	27
HEGEDŰS, DÁNIEL: Hungarian Videoblogger Networks Online.....	43
OLÁH, ESZTER: Empirical analysis of the judgment of unconditional basic income through YouTube comments.....	68
BENE, MÁRTON: What’s the matter? A text mining analysis of political topics and user engagement on politicians’ Facebook pages during the 2018 Hungarian general election campaign	94
SÜTŐ, ANNA: Where to go, net generation? Lifestyle-based segments of the Hungarian youth	124
SIMON, SÁRA: Social media communication in the digital medical space	143
MÁTÉ, FANNI: Classification of depression-related online forums using <i>Natural Language Processing</i>	181

REVIEW

A. GERGELY, ANDRÁS: Studia Cingarorum – Romology in practice in the distribution of knowledge.....	209
---	-----

READ UNNECESSARY!

ENDRE SIK’S PERMANENT COLUMN

JUHÁSZNÉ FAZEKAS, ZSUZSANNA: Cumulative disadvantages? – or the everyday life of Czech and Slovak Roma immigrants in Glasgow	214
---	-----

Abstracts	221
-----------------	-----



Digitális szociológia – Kihívások és távlatok

Szerkesztői előszó

DESSEWFFY TIBOR

Ha igaz az, hogy a szociológia feladata a körülöttünk örvénylő valóság megértése és értelmes magyarázata, akkor aligha van mit csodálkoznunk a digitális szociológia utóbbi években tapasztalható robbanásszerű előretörésén. A területhez tartozó publikációk, kiadványok, kutatások számának szinte követhetetlen ütemű terjedése arra a kettősségre vezethető vissza, amit digitális hermeneutikának neveztem egy korábbi könyvemben (Dessewffy 2019). Arra az egymással összefüggő, egymásra folyamatos hatással bíró kettősségre gondolok, hogy egyrészt a digitális adatokon alapuló technológiák megváltoztatják az életünket, illetve – szociológusok számára hasonlóan fontos módon – radikálisan bővítik a megismerés eszköztárát, és így a feltehető, vizsgálható értelmes kérdések körét is.

A világ, a társadalmak átalakulását jól érzékelteti, hogy míg egy évtizede lehetséges volt a digitális szociológiát elkülönült szakszociológiaként kezelni, ma már alig lehet olyan területet említeni, ahol a digitalizáció következményeinek ne lenne komoly szerepe, a településszociológiától az érzelmek szociológiájáig, a kulturális fogyasztástól az egyenlőtlenségek vizsgálatán át az öregedés- vagy az ifjúságkutatásig. Még az olyan, látszatra immunis területek, mint a társadalom- vagy politikátörténet is erősen érintettek, hiszen a digitalizált archívumok és releváns módszertanok (pl. szövegbányászat) számtalan új kutatási lehetőséget nyitnak meg az e témák iránt érdeklődők számára. Ha pedig a múltból a jelen felé fordítjuk figyelmünk fókuszát, akkor valóban lajstromozhatatlanul tág az új jelenségek köre.

Ehhez a kavalkádhoz két megjegyzés kívánkozik. Egyrészt ma már egyre többen ismerik fel – hozzátenném a szociológián kívül is –, hogy nem technológiáról van szó önmagában. Az a kérdés, hogy a digitális adatok és az ezek használatára irányuló technológiák elterjedése hogyan változtatja meg az életünket, pontosabban mi mindannyian hogyan alkalmazzuk és alakítjuk ezeket a technológiai lehetőségeket. Ezért kérdezheti Daniel Miller és szerzőtársai (Miller és mtsai 2016) szellemes könyvük címében: Mit tesz a világ az Internettel? A felhasználások, újragondolások és kreatív átalakítások listája szinte végtelen. Ezek némelyike felemelő (mint a betegségekkel élők önszorgító Facebook-csoportjai), némelyike borzalmas (mint a tömeggyilkosságokat élőben közvetítő gyilkosok), de olyan felhasználása ez a technológiának, amire aligha gondoltak kialakításakor a fejlesztők.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A másik fontos megjegyzés ennek az átalakulásnak dinamikájára vonatkozik. A kevés bizonyosság egyike, amivel szolgálhatunk, az, hogy bármilyen brutálisnak is érzékeljük ezt az átalakulást, ami a születéstől a halálig, a munkától a szerelmen át a háborúig átalakítja életünk folyamatait, ennek a változásnak még csak az elején vagyunk. Nem a beláthatatlan jövővel szeretnék riogatni, hiszen egyikünk sem tudhatja, mit hoznak majd a következő évek, évtizedek, inkább arra szeretném felhívni a figyelmet, hogy milyen gyakran válunk a „nyelv cselének” áldozatává: hiszen én is úgy beszéltem a „digitalizációról”, „internetről”, mintha ezek a fogalmak, akár a múlt-ra vetítve, egységes folyamatokat jeleznének. Vannak olyan kontextusok, amikor ez a homogenizáció elfogadható – remélem, a fenti bekezdések is ilyenek voltak –, de digitális folyamatok kapcsán fájdalmasan jelentkezik az időhöz kötöttség problémája. Egész mást jelentett, más társadalmi gyakorlatokat implikált az internet a betárcsázós hőskorban a múlt évezred végén, mást a korlátlan, majd wifihasználattal az azt követő 2000-es években, majd megint más felhasználási szokásokat erősített fel a mobilhasználattal jellemezhető jelenben tartó időszak, és akkor még nem beszéltünk a platformok ma véget nem érőnek tűnő burjánzásáról. Ebben a nagyon gyorsan változó, gyakran radikális átrendeződéseket hozó világban a folyamatok, jelenségek „időbélyegezésének”, a kutatási eredmények reflexív kontextusba helyezésének különösen nagy szerepe van.

Ezzel át is térhetünk a digitális hermeneutika megismerési oldalára. Ennek egy örömdetes aspektusa, hogy a megismerési eszköztár átalakulása követi a valóság változásának sebességét: az új analitikai eszközökkel az adatok gyűjtése, kiértékelése, vizualizációja és publikációja is hihetetlen mértékben felgyorsult. Ez azonban, ne kerteljünk, korántsem ártatlan folyamat, sem intellektuális, sem pozicionálási szempontból. Bruno Latour mondását számtalanszor idéztem az elmúlt években: ha megváltoztatod a megismerés eszközeit, megváltoztatod a társadalomelméletet is (Latour 2010). Nem egyszerűen mást látunk akár ugyanabból a valóságdarabból, ha szemüvegen, vagy távcsővel, vagy mikroszkóppal vizsgáljuk, hanem egészen más jelenségek és összefüggések tárulnak fel előttünk, amelyek új elméletek iránti igényeket is artikulálnak. Ennek az új nyelvnek, elméleti keretnek a kidolgozása lesz a következő évtized egyik legnagyobb intellektuális kihívása.

A másik megkerülhetetlen kihívás a pozicionálással kapcsolatos. Az elmúlt évszázadban a társadalomtudományok pozícióját – belső rivalizálásokat leszámítva – a társadalmi folyamatok értelmezésében tulajdonképpen nem kérdőjelezték meg. Mára ez radikálisan megváltozott: fizikusok, matematikusok, informatikusok, programozók, a data science zászlója alatt csoportosuló legkülönbözőbb bandériumok jelentik be igényüket nem egyszerűen alternatív valóságmagyarázatok prezentálására, de adatfelhasználási kompetenciájuk legitimitási bázisán érvényesebb elemzések bevezetésére is.

Ez nem a szociológia pozícióvesztése miatt aggályos elsősorban – bár nem kell szégyellni egyik szakma képviselőinek sem, ha működik bennük az életösztön és túl-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

élni szeretnének. Ami ennél azonban fontosabb, az a szociológiai tudás és képzelet elmúlt évszázadokban felhalmozott fogalomkészlete és megértési apparátusa, amely meggyőződésem szerint elengedhetetlen a digitális kor társadalmi folyamatainak megértéséhez. Azért szolgált különös örömmre a *Metszetek* felkérése a *Digitális szociológia* különszám szerkesztésére, mert a kollégák alábbiakban olvasható tanulmányai ezt a szakmai rezilienciát, intellektuális nyitottságot, képzelőerőt és módszertani alaposságot jelenítik meg, amely szükséges ahhoz, hogy a szociológia a 21. században is sikeres maradjon. Fontos tanulmánygyűjtemény született, amely egyszerre mutatja a lehetséges megközelítések sokszínűségét és ad kurrens képet a feldolgozott témákról. Köszönöm a különszám szerzőinek a remek tanulmányokat, a felkért opponenseknek az írásokra fordított időt és a konstruktív kritikákat, valamint a *Metszetek* szerkesztőségének a felkérést és a szövegek gondozását. Azt remélem, hogy az Olvasónak is annyi öröme lesz ezekben a tanulmányokban, mint a szerzőknek és a szerkesztőknek.

Irodalom

- Dessewffy, T. (2019): *Digitális szociológia. Budapest, Typotex Elektronikus Kiadó Kft.*
- Miller, D. – Sinanan, J. – Wang, X. – McDonald, T. – Haynes, N. – Costa, E. – Nicolescu, R. (2016): *How the world changed social media?* UCL press.
- Latour, B. (2010): *Tarde's idea of quantification.* Routledge.



Új utak a társadalom megismerésében¹

A donáció alapú digitális adatgyűjtésben rejlő lehetőségek

KMETTY ZOLTÁN²

ABSZTRAKT

Napról napra egyre több digitális adat keletkezik és egyre több társadalomtudományi elemzés használ Twitter, Instagram vagy akár Facebook-adatokat. A „big data” jelenség kapcsán felmerülő társadalomtudományi lehetőségeket és dilemmákat számos nemzetközi és hazai tanulmány végigjárta már – de az „adathoz jutás” kérdésével csak érintőlegesen foglalkoztak ezek a tanulmányok. Az adathoz jutás pedig egyre nehezebbé válik. Mit tehetünk abban az esetben, ha a piaci szereplők lezárják a platformjaikat és ha mégis találunk elérhető adatot, akkor a kutatás-etikai tanács parancsol nekünk megálljt? A válasz egyszerű: forduljunk a felhasználókhoz és tőlük kérjük el az adatokat. Ezt a megközelítést nevezi a szakirodalom adatdonációnak. A tanulmányban részletesen bemutatjuk az adatdonációs megközelítést külön kitérve arra, hogy a jelenlegi nagy nyugati platformok esetében milyen adatokhoz férhetnek hozzá a kutatók a felhasználókon keresztül. Az adatdonációs hozzáférés gyakorlati megvalósíthatóságát egy hazai pilot kutatás alapján mutatjuk be.

KULCSSZAVAK: adatgyűjtés, adatdonáció, közösség média, big data, Facebook

ABSTRACT

New ways in exporting Society The potential of donation-based digital data collection

More and more digital data is being generated every day, and more and more social science analyses are using Twitter, Instagram, or Facebook data. Many international and national studies have already explored the social science opportunities and dilemmas raised by the phenomenon of „big data” - but the issue of „access to data” has only been touched upon tangentially. And access to data is becoming increasingly difficult. What can we do if market players close the access to their data, and, if we find data available, the Research Ethics Board tells us to stop? The answer is simple: go to the users and ask them for the data. This approach is what the literature calls data donation. This paper will describe the data donation approach in detail, focusing on how researchers can access data through users on the current major Western platforms. The practical feasibility of data donation access will be illustrated based on a domestic pilot study.

KEYWORDS: data collection, data donation, social media, big data, Facebook

¹ A kutatás az NKFI-től nyert támogatást a Fialat Kutató Témapályázaton. A kutatás azonosítója: FK128981.

² Eötvös Loránd Tudományegyetem, Társadalomtudományi Kar, Szociológia Intézet; Társadalomtudományi Kutatóközpont, CSS-RECENS kutatócsoport, e-mail: kmetty.zoltan@tat.k.elte.hu



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Bevezetés

A kérdőíves survey kutatások uralták a kvantitatív társadalomtudományi kutatások elmúlt 50-70 évét. A kutatók mindig tudatában voltak ennek a módszernek a gyengeségeivel, de mint a rendelkezésre álló legjobb technikát, semmi sem törte meg a hegemoniáját. A közelmúltbeli változások azonban megkérdőjelezték a survey kutatások vezető szerepét. E változások egyik része az egyre nehezebbé váló terepmunka, valamint a csökkenő válaszadási arány; a változások másik motorja pedig új típusú digitális adatforrások megjelenése. A digitális adatok részét képezik azok a nem szándékolt adatok egyrészt amit a különböző általunk használt eszközök felvesznek rólunk, vagy lakásunkról (például: mobiltelefonos adatok helylokációja). A digitális adatok egy másik része a felhasználók által megosztott tartalmak, például tweetek vagy bejegyzések az általuk kedvelt helyekről, vagy más, közösségi médiában megjelenő reakciók, interakciók. Társadalomtudományi szempontból a közösségi média adja a digitális adatok egyik legérdekesebb típusát. A nagy üzleti cégek számára a „Social Listening” része a mindennapos üzleti folyamatnak, mivel a felhasználók igényeinek gyors megválaszolása vagy a fogyasztói vélemények megértése elengedhetetlen részei a fogyasztói elkötelezettség mérésének. Az „ipar” igényei implicálják a közösségi média adatok feldolgozásának gyors fejlődését, de ez az igény a kutatói oldalról is jelentkezik. Egyre több társadalomtudományi elemzés használ Twitter, Instagram vagy akár Facebook-adatokat.

A „big data” jelenség kapcsán felmerülő társadalomtudományi lehetőségeket és dilemmákat számos nemzetközi és hazai tanulmány végigjárta már (Lazer – Radford 2017, Csepeli 2015, Dessewffy – Láng 2015, Kmetty 2018, Ságvári 2017) – de az „adathoz jutás” kérdésével csak érintőlegesen foglalkoztak ezek a tanulmányok. Ez részben azzal magyarázható, hogy látszólagosan rengetek digitális adat érhető el szinte bárki számára – mindenholon ömlenek ránk az adatok. De ha jobban körbejárjuk a kérdést, akkor egyértelművé válik, hogy a helyzet korántsem egyszerű. Egyrészt egyre több (piaci) szereplő ismeri fel a saját adatvagyonának értékességét és korlátozza a külső szereplők adat hozzáférését. Másrészt a(z Európában) szigorodó adatvédelmi szabályozások miatt is szűkül a kutatási célból elérhető digitális adatok köre. Mit tehetünk abban az esetben, ha a piaci szereplők lezárják a platformjaikat és ha mégis találunk elérhető adatot, akkor a kutatás-etikai tanács parancsol nekünk megálljt? A válasz egyszerű: forduljunk a felhasználókhoz és tőlük kérjük el az adatokat (Halavais 2019). Ezt a megközelítést nevezi a szakirodalom adatdonációnak. Ebben a megközelítésben követlenül fordulunk a felhasználókhoz – kihagyjuk a piaci cégeket (platform szolgáltatókat) és a felhasználók beleegyezésével kezdjük el használni az adataikat. Ezzel a megoldással nemcsak az adatfelvételünknek teremtünk egy nagyon tiszta jogi környezetet, hanem azt is lehetővé tesszük, hogy akár további kérdőíves adatokat is gyűjtsünk a kutatásba bevont személyektől.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A tanulmány első felében bemutatjuk a digitális adathozzáférés jellemző csatornáit a megközelítések előnyeivel és hátrányaival. Ezt követően részletesen bemutatjuk az adatdonációs megközelítést külön kitérve arra, hogy a jelenlegi nagy nyugati platformok esetében milyen adatokhoz férhetnek hozzá a kutatók a felhasználókon keresztül. A tanulmány utolsó részében pedig bemutatunk egy olyan egyedülálló hazai pilot kutatást, amely adatdonációs stratégiát használt az adatgyűjtés során.

Digitális adathozzáférés

JELLEMZŐ UTAK

A digitális adatokhoz való hozzáférés jellemző módját az API-k (Application Programming Interface) jelentették. Az API-k alapvetően a számítógépek egymás közötti kommunikációját/adatcseréjét segítik azzal, hogy autentikált csatornákon kaput nyitnak egy adatbázis bizonyos részéhez. Ha például szeretnénk feltüntetni a honlapunkon az éttermünk „google review” értékelését, akkor a honlapunkba tudunk egy olyan kódot integrálni, amely bizonyos időközönként lekéri az aktuális értékeléseket a Google erre dedikált API-án keresztül. Ezt a hozzáférést azonban nemcsak gépek közötti kommunikációra lehet használni, hanem adatok kutatási célú kinyerésére is. A nyilvános API-k egyszerű hozzáférést biztosítanak nagy mennyiségű adathoz, de az adatok minősége változó és az is változó, hogy melyik platform mennyi adatot ad az API-kon keresztül. Bár bizonyos platformok esetekben – mint például a Twitter – ez a hozzáférési mód továbbra is az egyik leghatékonyabb adatelérési út, de más közösségi oldalak, például a Facebook vagy az Instagram esetében ezt a hozzáférési módot leállították vagy drasztikusan megnehezítették a platformok tulajdonosai (Breuer et al. 2021). Az API-k lezárása elsősorban a Cambridge Analytica botrány következménye, de a szigorodó adatvédelmi környezet mindentől függetlenül is abba az irányba terelte a platformokat, hogy szűkítsék az adathozzáférést. Ennek kapcsán Freelon (2018) azt írta, hogy a „Computational Social Science” belépett a „Post-API” korba, Bruns (2019) pedig ezt az egész helyzetet „APIcalypse”-nek nevezte. Mások, mint Tromble (2021) vagy Puschmann (2019) viszont kiemelik ennek a folyamatnak a pozitív hatását, miszerint végre véget ért ezzel a „közösségi média kutatásának vadnyugata”.

A nehezedő adathozzáférési környezetben új modelleket kellett kidolgozni a digitális adatokhoz való hozzáférés érdekében.

A NetGain Partnership (Shapiro et al. 2021) által publikált tanulmány két nagy ágat különbözteti meg a digitális adathozzáférésnek: azok a megközelítések, amelyek együttműködnek a platformokkal és azok, amelyek nem.

A platform együttműködés a következő modellekre jellemző (Shapiro et al. 2021: 28):



TEMATIKUS TANULMÁNYOK – Digitális szociológia

1. **API:** A Cambridge Analytica botrányig ez a hozzáférési mód volt a legelterjedtebb. Bár nagy mennyiségű adathoz lehetett egyszerű módon hozzáférni az API-kon keresztül, de a megközelítésnek voltak hátrányai is: általában kevés adatunk volt a megfigyelt emberekről és arról is, hogy azok az adatok, amihez hozzáférünk hogyan viszonyulnak a teljes adatmennyiséghez. Szintén gondot jelent az API-alapú hozzáférés esetében, hogy a megváltozó adatgeneráló algoritmusok jelentősen gyengíthetik a kapott eredmények érvényességét.
2. **Differenciált adatvédelem (Differential Privacy – DP):** A differenciált adatvédelemnek több megközelítése van. A digitális adatok kapcsán azokat az adathozzáférési módokat értik ez alatt, amikor az adatokra valamilyen zajt tesznek, vagy feltöltik az adatokat véletlenszerű válaszokkal vagy válaszolókkal. Ez valamelyest rontja az adatminőséget és megnehezíti azokat az elemzéseket, amelyekben alapvetően gyenge összefüggés van a változók között. Tipikus példája ennek a megközelítésnek a „Social Science One”³ project, amelynek keretében kutatók elfogadott kutatási tervekkel hozzáférhetnek zajjal terhelt Facebook-adatokhoz (King – Persily 2020).
3. **Platformok direkt adatmegosztása publikálási megkötésekkel:** Ez egy viszonylag ritka adathozzáférési mód, nagyon szoros kapcsolatot feltételez a kutató és a platform között. Az ilyen típusú hozzáférésekből publikált eredmények nem reprodukálhatók és etikailag erősen megkérdőjelezhetők.
4. **Kontrollált környezetben zajló hozzáférés:** Ezt a hozzáférési mód elsősorban az intézményi szereplőkre (pl. kormányzatok, statisztikai hivatalok) jellemző. Az adatszolgáltató kialakít egy olyan kontrollált hozzáférési környezetet (pl. adatszoba), amiben ellenőrizheti, hogy a kutatók milyen adatokat dolgoznak fel és ezekből milyen elemzéseket készítenek. Ha a kutatási eredmények adatvédelmi szempontból nem aggályosak, akkor engedheti az adatszolgáltató azok közzétételét. Ez az adatgyűjtési megközelítés nem jellemző a közösségi média platformok esetében.

A platformtól független adatgyűjtési módok a következők (Shapiro et al. 2021: 28):

1. **Web-scrapelés:** A web-scrapelés egy nagyon jellemző adatszerzési megoldás a digitális adatok esetében. A web-scrapelés gyakorlatilag a honlapokon és online oldalakon fent lévő információk nyers letöltését és adatbázisba rendezését jelenti. A közösségi média adatok scrapelése kapcsán megfogalmazódnak technikai és jogi dilemmák is (Mancosu – Vegetti 2020, Boeschoten et al. 2020). A közösségi oldalak egyrészt rendszeresen változtatják az oldal kinézetét, szerkezetét, ezért az oldalak leszedését végző kódokat folyamatosan karban kell tartani. Ez a konzisztens adatgyűjtést nagyon költségessé tudja

³ <https://socialscience.one>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tenni. Másrészt a platform tulajdonosok különböző jogi eljárásokat indíthatnak a kutatókkal szemben. Amerikában a számítógépes visszaélésekkel (CFAA) kapcsolatos törvény nyújt erre lehetőséget, Európában a GDPR szabályozás. A platform tulajdonosok arra hivatkozhatnak ezekben az eljárásokban, hogy a kutatók megsértik a felhasználó feltételeket (Terms of Service – TOS). A scrapelés történhet a felhasználók engedélye nélkül és engedélyével. Utóbbira jó példa a GESIS kutatása, amelyben a kutatásban résztvevők feltelapították a gépükre egy böngésző kiegészítőt, ami folyamatosan scrapelte a Facebook-falukat, amikor használták az oldalt (Haim – Nienierza 2019). Ez a megoldás GDPR szempontból maximálisan megfelelő, de ettől függetlenül szembe mehet a platform felhasználási feltételeivel.

2. **App-scrapelés:** Az appok scrapelése még bonyolultabb, mivel az appok általában egy zárt környezetben futnak, amiben nehéz hozzáférni az adatokhoz. Ehhez speciális programokat kell telepíteni a felhasználóknak a gépre/telefonjára, az adatgyűjtéshez. Ez nagyon szoros együttműködést feltételez a felhasználóval.
3. **Adatdonáció:** Ebben a modellben a kutatók megkérlik a résztvevőket, hogy osszák meg velük a platformon tárolt adatokat. A nagy platformoknak a GDPR törvény megfelelés miatt lehetőséget kell adni a felhasználóiknak, hogy elérjék és letöltsék a róluk tárolt adatokat, adatsomagok keretében (data download packages – DDP). A nagy nyugati platformok, mint a Google, a Facebook, az Instagram a WhatsApp vagy a Netflix felhasználóbarát módon adnak lehetőséget az adataink elérésére és letöltésére. Ezeket az adatokat amellet, hogy letöltheti a felhasználó akár további is oszthatja. Ez lehetőséget teremt a kutatóknak, hogy teljesen tiszta jogi környezetben férjenek hozzá közösségi média adatokhoz.

ADATDONÁCIÓ

A vállalatok helyett a felhasználókkal való együttműködés fő előnye, hogy ez átláthatóbbá teszi az adatgyűjtési folyamatot és az adatok felhasználását azok számára, akiknek az adatait használják. Ebben a kutatási megközelítésben nem okoz problémát a felhasználóktól beleegyező nyilatkozatokat gyűjteni. Pontosan elmagyarázhatjuk nekik, miről szól a kutatásunk, mire fogjuk használni az adataikat, hogyan tároljuk majd az adatokat és hogyan biztosítjuk az anonimitás. Az adatgyűjtés során láthatják és akár kontrolálhatják azt, hogy milyen adatokat osztanak meg a kutatókkal (Breuer et al. 2021).

Az adatdonációs megközelítés mivel szoros együttműködést igényel a résztvevőkkel azt is lehetővé teszi, hogy a digitális adatgyűjtés mellett „klasszikus” kérdőíves adatokat is gyűjtsünk a felhasználóktól. A két adatgyűjtési mód ötvözése kuta-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tói szempontból egy nagyon értékes adatbázist eredményezhet (Stier et al. 2020). A survey kutatások jellemző problémája, hogy a társadalmi elvárások miatt a kérdezettek gyakran torzítottan válaszolnak a kérdőíves kutatás kérdéseire, de akár felmerülhetnek visszaemlékezési problémák is bizonyos szituációkban (Scharkow 2016). A digitális adatok esetében bár szintén felmerülhetnek bizonyos torzítások, mint például az algoritmusok szerepe – mivel beavatkozás mentesek, nem befolyásolják őket a kutatói percepciók. Ezzel szemben hátránya lehet a digitális adatoknak, hogy sokszor csak a szereplők cselekedeteit látjuk, nem tudunk semmit róluk – hiányoznak a demográfia információk. Ezek a demográfia adatok viszont pótolhatók a survey kutatásból így át tudjuk hidalni a digitális adatoknak ezt a gyengeségét. A két adatforrás kombinálása pedig olyan összefüggéseket is fel tud tárni, ami külön-külön nehéz, vagy akár lehetetlen is lenne.

Az adatdonáció további előnye az, hogy a kutatókat nem kötik az API-k. Az API-k bár egyszerű adatgyűjtést biztosítanak más szempontból korlátozzák az elérhető adatok körét. Csak azokat az információkat érjük el az API-kon keresztül, amit a platform szolgáltatók fontosnak gondolnak. Mivel a legtöbb API nem kutató céllal készült ezért sokszor kapunk számunkra irreleváns adatokat és az igazán releváns adatok pedig hiányoznak. Az adat donációs modellben viszont látunk mindent, amit a felhasználó engedélyez. Az API-alapú közösségi média adatgyűjtések ráadásul általában csak a nem privát tartalmakhoz adnak hozzáférést, ezzel szemben az adat donációs modellben hozzáférhetünk a privát tartalmakhoz is.

További előny az is, hogy akár kis mintanagyság mellett is, az ezzel a módszerrel gyűjtött adatok meglehetősen részletesek és gazdagok lehetnek. Képesek lehetünk akár egy felhasználó teljes közösségi média „életútját” visszakövetni. Ez lehetővé teszi a felhasználók viselkedésének részletes időbeli elemzését akár az egyén és akár a csoport szintjén. Megvizsgálhatjuk például a személyes életesemények (például az egyetem megkezdésének) hatását, vagy vizsgálhatunk jelentős helyi/országos/világ szintű történések online lenyomatait.

A digitális adatokkal szemben gyakori kritika, hogy nehezen általánosíthatók az eredmények. Ennek elsődlegesen az az oka, hogy ritkán tudjuk azt, hogy az általunk gyűjtött adatok mennyiben reprezentálják a vizsgálni kívánt sokaságot. Az adatdonációs megközelítés viszont lehetővé teszi, hogy pontos képet kapjunk a mintánk szerkezetéről és ezáltal az is lehetővé válik, hogy akár a teljes sokaságra is kiterjeszhető eredményeket tudjunk publikálni.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

1. tábla. Az adatdonációs megközelítés előnyei és hátrányai

<i>Előnyök</i>	<i>Hátrányok</i>
Részletes adatok a felhasználókról Időbeliség Kakár privát megosztások Kombináhatóság kérdőíves adatokkal Jogilag tiszta környezet Nincs leszűkítve a hozzáférés, mint az API-ban Kontrolálhatjuk, ki kerül be a mintánkba	Aktív közreműködés szükséges a résztvevőktől Kiseb adatbázisok készíthetők, mint egy API-alapú kutatásban Bonyolult az adatokat anonimizálni A kontextust általában nem látjuk, amiben az adat keletkezik A platform szolgáltató megváltoztathatja időnként, hogy milyen adatok érhetőek el a DDP-ben.

Forrás: Saját szerkesztés: Breuer et al. 2021: 18

Az előnyök mellett fontos megemlítenünk a módszer kapcsán felmerülő limitációkat is. Az API-alapú adatgyűjtés gyors és egyszerű. Az adat donációs adatgyűjtést ezzel szemben alaposan meg kell szervezni. Meg kell győzni a potenciális mintagombokat, hogy vegyenek részt a kutatásban. A résztvevőknek pedig aktívan közre kell működni az adatfelvételben, segíteni kell a kutatóknak az adathozzáférésben, ami nem minden felhasználónak triviálisan egyszerű. Az adatgyűjtés nem csak a felhasználókat állíthatja kihívások elé, hanem a kutatókat is. Leginkább az anonimizálás igényel extra odafigyelést, de maga az adatszerkezet bonyolultsága is kihívás elé állíthatja a kutatókat. Magas szintű programozói tudás szükség ahhoz, hogy az egyedi felhasználóktól érkező adatokat össze tudjuk rendezni egy (vagy több) elemezhető adatbázisba. Ez még inkább igaz akkor, ha az adatfelvételi időszakon belül változtat a platform szolgáltató az adathozzáférési protokollján. Ez további külön adatintegrációs lépéseket indukál a kutatók részéről.

A közösségi média adatok bár nagyon részletesek, de a kontextusuk sokszor hiányzik. Azt például láthatjuk, hogy kedvel valamilyen tartalmat a felhasználó, de a konkrét tartalmat már nem láthatjuk, mert az az adat már a másik felhasználó „tulajdona” ezért a résztvevő adatletöltési csomagjának ez már nem része. A megfigyelés teljeskörűségének hiánya abban is jelentkezik, hogy ’csak’ azokat a tevékenységeket láthatjuk, amikor a felhasználó aktívan tesz valamit, amikor csak „böngész” és „megfigyel” eltűnik a kutatói radarról. Ez például fontos probléma lehet olyan kutatásoknál, amelyek a felhasználók hírközlését vizsgálják. Ilyen esetekben lehet érdekes a kutatási dizájn kombinálása web böngésző alapú adatgyűjtéssel is⁴.

A limitációk ellenére a DDP-n keresztüli adatgyűjtés az egyik legígéretesebb új módszer a közösségi média tevékenység vizsgálatára. A következő részben rövid át-

⁴ Ilyen kombinált adatgyűjtés tudomásunk szerint még nem készült soha.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tekintést adunk arról, hogy milyen típusú adatokhoz férhetünk hozzá egy adat donációs projektben a Facebook az Instagram és a Google esetében.⁵

FACEBOOK

A Facebook jelenleg a legnagyobb közösségi média oldal és a növekedése töretlen volt az elmúlt 10 évben. 2020 végére 2,8 milliárd aktív használója volt a Facebooknak havonta, és 1,8 milliárd naponta.⁶ Minden percben több mint 500 ezer komment születik és 300 ezer státusz frissítés.⁷ Magyarországon több mint 5,2 millió Facebook-használó van, ami azt jelenti, hogy az aktív internetezők közel 90% aktív a közösségi oldalon, vagy legalábbis rendelkezik profillal.⁸ A magyar felhasználók elsősorban a barátokkal, ismerősökkel való kapcsolattartásra használják az oldalt, de több mint a felhasználók fele a híreket is itt fogyasztja, ami jól mutatja az oldal jelentőségét.

A saját Facebook-adatok a beállítások menüben érhető el, és ezen belül „A Facebook-adataid” almenüre kell kattintani. Ha valaki csak böngészni szeretné a Facebook-adatait akkor a „Hozzáférés az adataidhoz” opciót lehet választani, az adatok letöltéséhez viszont a „Saját információ letöltése” linkre kell kattintani. Itt hozzáférhetünk minden rólunk tárolt adathoz. Az adatkörök folyamatosan bővülnek, ahogy a Facebookon új funkciók kerülnek kialakításra. Elérhetőek olyan alap információk, mint a barátlista, reakciók, kommentek, posztok, vagy a csoporttagság. De itt vannak a profil adataink is, illetve olyan információk például, hogy a Facebook milyen érdeklődési kategóriába sorol minket. A letöltésnél kiválaszthatjuk, hogy milyen tartalmakat szeretnénk letölteni – nem kell a teljes információmezőt letölteni. Három további opció van a letöltésnél. Kiválaszthatjuk az időszakot, a multimédiás tartalom minőségét és a letöltési formátumot. Utóbbi esetében kettő közül választhatunk: html vagy JSON. Utóbbi javasolt inkább, a JSON fájlokat nagyon hatékonyan tudja olvasni a Python és az R is. Az R és Python utalás arra is rávilágít, hogy ahhoz, hogy elemzési formátumra hozzassuk ezeket az adatokat kell a projekt csapatba legalább egy nagyon jól programozó kolléga.

A tanulmány harmadik felében bemutatunk majd egy olyan projektet, amelyben Facebookról gyűjtöttünk adatokat. A Facebook-adatokra épülő adatgyűjtés további technikai részeket ezért itt most nem bontjuk ki, a későbbi részekben fogjuk ezeket ismertetni.

⁵ Olyan platformokat mutatunk be, amelyeket a magyarok rendszeresen használnak. A kör bővíthető lenne más platformokkal, mint a Twitter vagy a WhatsApp. Ezeknek azonban alacsony a hazai bázisa.

⁶ <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

⁷ <https://zephoria.com/top-15-valuable-facebook-statistics/>

⁸ http://nmhh.hu/dokumentum/187704/lakossagi_internethasznalat_2016.pdf



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Másolat kérése		Elérhető másolatok	
Dátumtartomány:	Az összes adatom	Formátum:	JSON
Médiatartalom minősége:	Jó	Fájl létrehozása	
Az adataid		Az összes kijelölése	
<input checked="" type="checkbox"/>	Csoportok Csoportok, amelyeknek a tagja vagy, illetve amelyeket kezelsz, valamint a bejegyzéseid és hozzászólásaid azokban a csoportokban, amelyeknek a tagja vagy	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	Ismerősök Emberek, akikkel kapcsolatban állsz a Facebookon	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	Facebook Gaming A Facebook Gaming-profilod	<input checked="" type="checkbox"/>	
<input type="checkbox"/>	Aktivitás A Facebookon általad végrehajtott műveletek.	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	Történetek A történetednél megosztott fényképek és videók	<input checked="" type="checkbox"/>	
<input type="checkbox"/>	A helyeid Az általad létrehozott helyek listája	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	Fizetési előzmények A Facebookon keresztül lebonyolított fizetéseid előzményadatai	<input checked="" type="checkbox"/>	

1. ábra. A Facebook adatletöltés ablakából egy részlet
Forrás: www.facebook.com

INSTAGRAM

Az Instagram a második legnagyobb közösségi oldal Magyarországon. Az Instagram penetrációja 36% körüli az internetező lakosságon belül.⁹ De a fiatalabbak esetében ez az arány még magasabb. Az Instagramon a Facebookhoz képest kevesebb funkció érhető el, több az audiovizuális elem és kevesebb a szöveges adat. Bár az Instagram nem olyan népszerű adatforrás, mint a Twitter a társadalomtudományokban, de bőven vannak olyan tanulmányok, amelyek az Instagram adatokat használják társadalomtudományi elemzésekben (Moreno et al. 2016, Reece – Danforth 2017, Brown et al. 2018, Koltai – Kmetty – Bozsónyi 2021). A fő vizsgált témák a mentális egészséggel, a cyberbullyinggal vagy az önképpel, önreprezentációval kapcsolatosak voltak. Az utóbbi években egyre több tanulmány kezdett el foglalkozni az audiovizuális elemek elemzésével is.

⁹ https://nmhh.hu/dokumentum/212534/internet_2019_tanulmany.pdf



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Instagram

q. Keresés

Profil módosítása

Professzionális fiók

Jelszó megváltoztatása

Alkalmazások és webhelyek

E-mail és SMS

Push értesítések

Névjegyzék kezelése

Adatvédelem és biztonság

Bejelentkezési tevékenység

E-mailek az Instagramtól

Fiókadatok

Csatlakozás dátuma
2020. április 15. 17:19

Fiók adatvédelmi beállításának változásai
[Az összes megtekintése](#)

Jelszóváltoztatások
[Az összes megtekintése](#)

Korábbi e-mail-címek
[Az összes megtekintése](#)

Korábbi telefonszámok
[Az összes megtekintése](#)

Születési dátum
A fiókodhoz nem tartoznak itt megjeleníthető információk.

Az alkalmazások közötti üzenetküldésre való átállítás dátuma
A fiókodhoz nem tartoznak itt megjeleníthető információk.

Információk a profilban

Korábbi felhasználónevek
[Az összes megtekintése](#)

Korábbi teljes nevek
[Az összes megtekintése](#)

Korábbi bemutatkozások szövegei
[Az összes megtekintése](#)

Korábbi hivatkozások a bemutatkozásban
[Az összes megtekintése](#)

...

Kapcsolatok

Aktuális követési kérések
[Az összes megtekintése](#)

Téged követő fiókok
[Az összes megtekintése](#)

Általad követett fiókok
[Az összes megtekintése](#)

Általad követett keresőcímkék
[Az összes megtekintése](#)

Az általad letiltott fiókok
[Az összes megtekintése](#)

Fiókok, amelyek elől elrejtetted a történeteket
[Az összes megtekintése](#)

Fiók tevékenysége

Bejelentkezések
[Az összes megtekintése](#)

Kijelentkezések
[Az összes megtekintése](#)

Keresési előzmények
[Az összes megtekintése](#)

Történetekhez kapcsolódó aktivitás

Szavazások
[Az összes megtekintése](#)

Hangulatjel-csúszkák
[Az összes megtekintése](#)

2. ábra. Az Instagram adatletöltés ablakából egy részlet
Forrás: www.instagram.com

A legtöbb tanulmány az Instagram API-t használta az adatgyűjtéshez, de ez az adathozzáférés gyakorlatilag megszűnt, miután a Facebook felvásárolta az Instagramot. Tudomásunk szerint az adatdonációs megközelítést eddig egyetlen társadalomtudományi kutatásban sem alkalmazták. Boeschoten és szerzőtársai (2020) tanulmányukban egy fiktív Instagram-alapú kutatással illusztrálják az adatdonációs megközelítés előnyeit.

A Facebookhoz hasonlóan az Instagram is lehetővé teszi, hogy a felhasználók – a GDPR-rendeletnek megfelelően – hozzáférjenek saját adataikhoz és letöltsék azokat. Az adatvédelmi al-oldalon ellenőrizhetjük, hogy az Instagram milyen típusú adatokat rögzít rólunk. Letölthetők technikai metaadatok (bejelentkezési adatok), profilinformációk, különböző információk a kapcsolatokról, korábbi keresésekről, történetekről vagy megosztott posztokról és képekről. Az időbélyegző és a földrajzi helymeghatározás is sok esetben elérhető.

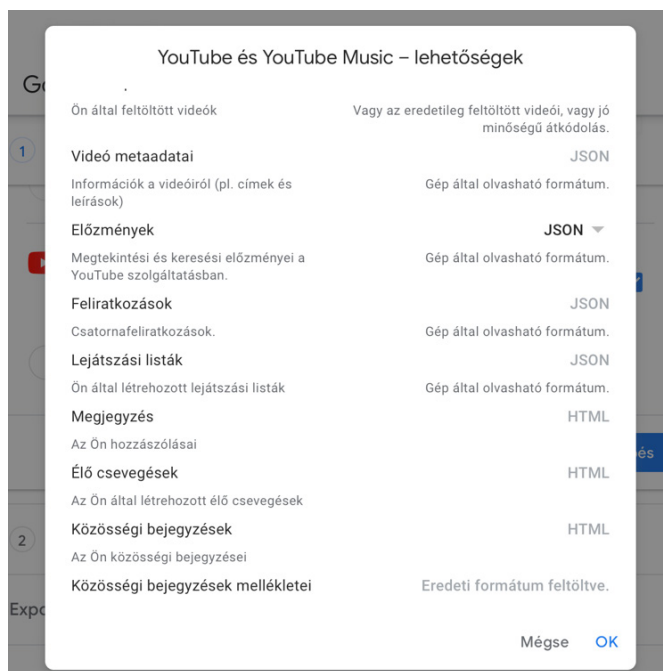


TEMATIKUS TANULMÁNYOK – Digitális szociológia

Az adatletöltő oldalon csak az adatformátumot tudjuk kiválasztani (html/JSON), de azt nem, hogy milyen adatokat szeretnénk letölteni. Az Instagram a felhasználó teljes adatarchívumát exportálja. Az Instagram által biztosított DDP jelentősen különbözik a Facebook által biztosított DDP-től. Mivel a felhasználó által megosztott összes audiovizuális tartalmat megkapjuk, nagyobb lesz az adatmért, ami megnehezíti bizonyos adatmegosztási sémák alkalmazását.

GOOGLE – YOUTUBE

A Google nem egy önálló platform, hanem inkább egy szolgáltató. A Google szolgáltatásait keresésre, e-mailezésre, adatmegosztásra vagy zenehallgatásra használják. Arról nincs információnk, hogy hányan használják Magyarországon a Gmailt e-mail kliensként, de azt tudjuk, hogy Magyarországon az internetezők több mint 80%-a hallgat zenét vagy néz videókat a YouTube-on.¹⁰



3. ábra. A Google Takeout adatletöltés ablakából egy részlet
Forrás: www.google.com

¹⁰ https://nmhh.hu/dokumentum/212534/internet_2019_tanulmany.pdf



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Szolgáltatásain keresztül a Google különböző típusú adatokat tárol a felhasználókról. Egyes szolgáltatásokat csak kevesen használják, így ezek az adatok kutatási szempontból kevésbé érdekesek. Más szolgáltatások nagyon érzékeny adatokat (például e-maileket) tartalmaznak, így ezen adatok gyűjtése nem reális erős lemorzsolódás nélkül. Az egyedi kutatási igények szabhatják meg, hogy milyen adatokra érdemes célozni a Google univerzumban. Ilyenek lehetnek például a következők:

- A felhasználó által kattintott vagy megtekintett hirdetések
- YouTube-aktivitás
- Google keresési előzmények
- Helymeghatározási előzmények.

A Google létrehozta a „Google Takeout” szolgáltatást az adatok letöltésére. Szemben az Instagrammal, itt a felhasználó kiválaszthatja, hogy milyen adatokat kíván letölteni. Az adattípusra kattintás után további opciók válnak elérhetővé. Az adatok formátuma nem egységes, az adattípustól függ. Szöveges adatok esetén a html/JSON a leggyakoribb formátum, de lista típusú adatok esetén a CSV is gyakori. Ha audio-vizuális adatokat szeretnénk letölteni akkor is több formátum közül választhatunk.

Egy hazai kísérleti kutatás bemutatása

ADATGYŰJTÉS

A tanulmány eddigi részében bemutatottuk, hogy milyen módszerekkel lehet digitális adatokhoz hozzáférni, ezen módszereken belül melyek az előnyei az adatdonációs megközelítésnek és körbejártuk, hogy a három nagy hazai penetrációjú platformon (Facebook, Instagram, Google) milyen adatokhoz és hogyan férünk hozzá. Az adatdonációs megközelítést azonban nem csak elméletben tudjuk bemutatni, hanem egy hazai pilot kutatás tapasztalataival is rendelkezünk. Ez a pilot kutatás 2018 őszén indult az NKFI támogatásával.¹¹

A kutatásban azt tűztük ki célul, hogy leteszteljük egyrészt a technikai megvalósíthatóságát egy adatdonációs projektnek, másrészt felmérjük azt, hogy milyen típusú társadalomtudományi elemzésekre alkalmas egy így nyert adat. Ebben a fejezetben az előbbi célokat vizsgáljuk, az elemzési lehetőségekre később térünk ki. A projekt első lépésében megvizsgáltuk azt, hogy milyen adatok érhetőek el egyáltalán a Facebookos adatletöltést használva. Ehhez a kiinduló profilokat a kutatócsoport¹² kutatóinak saját profiljai adták. Ez a lépés segített nekünk megérteni, hogy

¹¹ FK 128981

¹² A kutatócsoportban a következő kutatók vettek részt: Kmetty Zoltán (kutatásvezető), Németh Renáta, Boros Krisztián, Mogyorósi Pálma, Tarnói Csenge, Vancsó Anna, Váry Dániel.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

egyres tevékenységek hogyan reprezentálódnak a Facebook-adatsorban, mi az, amit láthatunk egy tevékenység kapcsán és mi az, amire nincs rálátásunk. Ezen a ponton döntöttünk arról is, hogy melyek azok az adatok, amelyek mindenképp szeretnénk gyűjteni és melyek azok az adatok, amelyekről le tudunk mondani. Kutató szempontból elsősre jó ötletnek tűnhet minden adatot elkérni. De ez három szempontból is problémás lehet.

- Vannak olyan funkciók, amit kevesen használnak vagy már nem használnak a felhasználók. Az ilyen funkcióra épülő adatokkal nem tudunk mit kezdeni, kvantitatív elemzésben nem használhatók értelmesen. Tipikus példa erre a Facebook-piactér adatsora, vagy a böké (poke) funkció.
- Vannak olyan adatok, amik nagyon érzékeny információkat tartalmazhatnak. Ilyenek például az üzenetek. Ha olyan tartalmakat kérünk el a felhasználóktól, amit nem szívesen adnak oda, azzal növeljük a kutatás visszautasításának valószínűségét. Mivel a kutatás alában is nagyon sokat kér a résztvevőktől, fontos találnunk ebben egy működő egyensúlyt. Ezt a logikát követve a privát üzeneteket és a keresési előzményeket nem kértük el végül a résztvevőktől.
- Adatvédelem szempontjából fontos elvárás minden hasonló kutatás kapcsán, hogy maximálisan biztosítsa a résztvevők anonimitását. Az anonimitás biztosítása sokkal nehezebb egy ilyen kutatásban, nem elég a válaszadótól külön fájlban kezelni a nevét és a címét. Az audiovizuális elemek anonimizálása nagyon komplex feladat, amely magában rengeteg erőforrást igényel. Ezért úgy döntöttünk, hogy az audiovizuális tartalmakat nem kérjük el a felhasználóktól. Ez például a pozitív következménnyel is járt, hogy nem kellett extra nagy tárhelyt biztosítani a kutatási adatoknak.

A kutatás lebonyolításával egy professzionális piackutató céget bízunk meg. A tervezett (és végül meg is valósult) minta nagyság 150 fő volt. A kutatásban egy gyenge életkori kvótát használtunk, annyit kötöttünk ki, hogy a minta legalább 40 százaléka 30 év feletti kell, hogy legyen. A minta nem valószínűségi minta, tehát nem reprezentálja a Facebook-használókat. A célunk a pilot kutatással elsősorban az volt, hogy megvizsgáljuk a kutatás technikai lebonyolításának lehetőségeit és képet kapjunk arról, mire lehet használni a kinyert adatokat. A résztvevőknek rendszeres Facebook-felhasználóknak kellett lenniük (azaz legalább heti rendszerességgel használniuk kellett a platformot). A válaszadók mindegyike magyar volt, az ország keleti részén élt, többnyire nagyvárosban. A terepmunka 2019 áprilisa és szeptembere között zajlott. A kutatási alanyok 3000 forintot kaptak a részvételért. A piackutató cég a saját kutatási adatbázisát használta fel a toborzáshoz, illetve diákszövetkezeteken keresztül is folyt a rekrutáció. Azoknak a résztvevőknek, akik beleegyeztek a vizsgálatban való részvételbe, elküldtük a projekt leírását, és kaptak egy meghívót a piackutató cég helyi irodájába.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Fontos technikai részlet, hogy a Facebook-adatok, a letöltés elindítása után nem válnak rögtön elérhetővé. A Facebook alacsony prioritást rendel az adatletöltés mellé, ezért általában lassan készül el az adat. Néha perceket kell várni, de van, amikor órákat is (főleg ha nagy az adattömb). Ezért a kutatási felkérést elfogadókat arra kérte a piackutató cég, hogy a tervezett találkozó előtt egy nappal indítsák el az adatexportálási folyamatot. A folyamat megkönnyítése érdekében a résztvevők részletes leírást kaptak arról, hogyan férhetnek hozzá a Facebook-adataikhoz, és hogyan indíthatják el az exportálást.

Az adatdonációs megközelítés egyik legnagyobb előnyének emeltük ki, a tiszta jogi környezetet és a jól érvényesíthető adatvédelmi szempontokat. Ennek fontos alapja a résztvevők beleegyező nyilatkozata a kutatásba, illetve az adatkezelési elvek elfogadása. Ehhez igazodva a vizsgálat megkezdése előtt a résztvevőknek alá kellett írniuk egy részvételi beleegyező nyilatkozatot.

Ahogy korábban már említettük a kutatási alanyoknak személyesen be kellett menni a piackutató cég egyik irodájába. Elviekben elképzelhető egy olyan kutatási dizájn is, ahol a résztvevők feltöltik a Facebook-adataikat egy dedikált szerverre és nem mennek be az irodába. További lehetőség lehet az is, hogy kérdezőbiztosok az otthonukban keresik fel a résztvevőket és a kérdezőbiztos gépére mentik el az adatokat. Azonban mi általunk alkalmazott kutatási megközelítésben fontos elem volt, hogy nem a nyers JSON fájlokat kaptuk meg a kutatást lebonyolító cégtől, hanem egy előfeldolgozott és lehetőségek szerint anonimizált adatsort. Ez az előfeldolgozás viszont csak jól előkészített környezetben működött stabilan.

A munka során kifejlesztettünk egy R kódot, ami a letöltött JSON fájlokat előfeldolgozta, átalakította CSV fájlkká és eltávolította az adatokból a megtalált neveket. Két dolog motivált minket, amikor elkezdtük ennek a kódnak a megírását. Egyrészt fontos volt, hogy biztosíthassuk a résztvevőknek azt a jogot, hogy belenézhessenek a letöltött adataikba és akár ki is törölhessenek bizonyos tartalmakat. JSON fájlok esetén ez nagyon nehézkesen lett volna megvalósítható, viszont a CSV fájlok egyszerűen olvashatók (akár excelben is). Ezzel a megoldással ezt a lehetőséget biztosítottuk a résztvevőknek. Fontos ennek kapcsán megjegyezni, hogy végül senki nem törölt ki adatok a letöltött archívumából. A másik motiváció az volt, hogy ne kelljen személyes adatot kezelni az adatfelvétel után, a személyes adatok kezelése teljes egészében a piackutató cég feladata legyen. Ehhez a kutatási alanyok nevét, és a Facebook-barátaiknak a nevét elfedtük az algoritmussal, hashelés segítségével (lásd hasonló kapcsán: Mancosu – Vegetti 2020). Ez a hashelés sokkal jobb megközelítés, mint az egyszerű törlés, mivel így anonim módon, de nyomon követhetjük egy adott Facebook-baráttal, mikor és milyen interakció történt. Az összes adatot ugyanazzal a hashelő módszerrel fedtük le – ez biztosította azt, hogy ugyanannak a személynek minden adatbázisban ugyanaz legyen a kódja is.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Az átalakító és anonimizáló R kód tesztelésekor számos problémát azonosítottunk a karakterkódolásokkal és a különböző programverziókkal kapcsolatban. A hasonló problémák elkerülésének érdekében a piackutató cég egy külön ezt a célt szolgáló laptopot használt a vizsgálathoz. Az irodában a résztvevőket arra kértük, hogy a kijelölt laptopon jelentkezzenek be a Facebook-fiókjukba, majd töltsék le az előkészített Facebook-adatprofil-archívumukat JSON formátumban. Ez választ ad, arra a korábban feltett kérdésünkre, hogy miért volt szükség arra, hogy a kutató cég irodájában megjelenjenek a résztvevők. A letöltött adatok átalakítása és anonimizálása nagyon nehezen lett volna kivitelezhető, ha más dizájnt választunk.

Egy érdekes következménye volt a hashelésnek. A kutatásban résztvevők Facebook-barátairól nagyon korlátozott információk állnak csak rendelkezésre. A résztvevővel zajló interakciókat látjuk, de arra csak egy gyenge becslésünk lehet, hogy milyen korú, milyen iskolai végzettségű, vagy milyen településtípuson lakik a barát. Az emberek neve viszont nagyon árulkodó információ a nemmel kapcsolatban. De a hashelés után ez az információ eltűnik. Ezért még a hashelés előtt, az összes Facebook-barátnak megbecsültük a nemét az elérhető magyarországi keresztnév listák alapján. Ez a megoldás több mint 90%-át besorolta az ismerősöknek. Ott volt gondban az algoritmus, ahol becéző nevek jelentek meg a Facebookon (pl. Kovács Eszti Kovács Eszter helyett), vagy az ismerősnek a családi és keresztnévben keveredtek a fiú és női nevek (pl. Ágoston Sára) vagy nem magyar Facebook-ismerőse is volt valakinek. Ez a példa jól mutatja, hogy már az adatfelvétel tervezése során érdemes átgondolni azt, hogy milyen változókra lehet majd szükségünk a későbbi elemzésekben.

Ahogy a fejezet felvezetésében részletesen foglalkoztunk vele, nem a teljes archívumot kértük el a résztvevőktől, hanem szűkítettük az adatok körét. Többek között nem kértük el a multimédia tartalmakat, a Facebook Messenger-üzeneteket, vagy a keresési előzményeket. A szűkített adatsor ellenére a kutatásban gyűjtött adatok a Facebook-tevékenységek széles körét fedik le, mint a bejegyzések, hozzászólások, kedvelések és reakciók, kedvelt oldalak, barátok, profilinformációk és hirdetésekre vonatkozó adatok. Az adatok tartalmi kibontásával a következő fejezetben foglalkozunk.

A Facebook-adataik megosztása mellett a résztvevőknek egy online kérdőívet is ki kellett tölteniük amíg az irodában tartózkodtak. A kérdőív nagyon változatos témákat fedett le, tartalmazott kérdéseket a politikáról, a médiahasználatról, az ön-reprezentációról, a mentális egészségről, a szabadidős tevékenységekről és a zenei preferenciákról is. A két adatforrás összekapcsolása érdekében ugyanazt az azonosító kódot használtuk az online kérdőívben és a Facebook-adatok tárolásakor. A nyers Facebook-adatokat az anonimizálási folyamat után törölte a kutató cég. A piackutató cég az anonimizált adatokat csak a vizsgálat vezető kutatójával osztotta meg.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

GYŰJTÖTT ADATOK KÖRE

A viszonylag kicsi, 150 fős minta ellenére egy nagyon komplex és gazdag adatsor keletkezett a kutatásban. Nem célunk, hogy jelen tanulmányban minden apró részletét kibontsuk a gyűjtött adatoknak, de egy átfogó képet mindenképp szeretnénk adni az adatsorról. A tanulmányban azt a struktúrát követjük, ami mentén a Facebook is rendezi az adatokat.

Barátok: A résztvevők Facebook-barátaira vonatkozó adatok tartalmazzák az összes barátjuk listáját, a barátság kezdetének időbélyegét, a barát hashelt nevét és a barát becsült nemét (lásd korábban). Az elutasított és függőben lévő baráti kérésekről, valamint az eltávolított barátokról is rendelkezünk információkkal. A 150 résztvevőnek összesen 115 955 élő barátkapcsolata volt, ami átlagban 773 barátot jelent. A legmagasabb Facebook-barátszám 2840 volt.

Oldalak: Az adatok tartalmazzák a résztvevők által követett összes oldal listáját, valamint egy időbélyegyet, amely jelzi, hogy mikor kezdték el követni az adott oldalt. Az oldalak egy részét (a mintánkban szereplő oldalak mintegy kétharmadát) a Facebook 18 különböző kategóriába sorolta (pl. könyvek vagy zenei oldalak). Összességében a vizsgálatunkban részt vevő 150 résztvevő 83 232 oldalt követett. Az egyedi oldalak száma 52 700. A résztvevők átlagosan 562 oldalt követtek.

Reakciók: Ez az adathalmaz tartalmazza a felhasználók összes reakcióját, időbélyegzővel és a reakció típusával (a leggyakoribb a tetszés). Tartalmazza továbbá a reakció célpontját (azaz, hogy az egy barát vagy egy oldal volt-e). A barátok posztjaira adott reakciók messze gyakoribbak (76,5%), mint az oldalak tartalmára adott reakciók. Fontos, hogy csak a metaadatok szintjén van információnk arról a tartalomról, amelyre a felhasználó reagált. Az adatsor nem azt tartalmazza, hogy pontosan mire reagált a résztvevőnk, hanem csak a tartalom típusára vonatkozó információkat. A tartalom 11 típusba sorolható, plusz egy „egyéb” opció. Adatainkban a legtöbb reakciót a posztokra (45,0%) és a képekre (43,8%) adott reakciók jelentették. Az adatbázisban szereplő reakciók száma összesen 1 802 430.

Egyéb Facebook-tevékenységek: A reakción kívül számos más típusú tevékenységet is végezhetnek az emberek a Facebookon. A leggyakoribbak a hozzászólás és a kommentelés, de a felhasználók fényképeket, videókat is feltölthetnek, linkeket oszthatnak meg, játszhatnak, szavazásokat indíthatnak, eseményeken vehetnek részt stb. Összességében 346 407 tevékenység rekordjai szerepel az adathalmazunkban. A leggyakoribb tevékenység a posztolás. A posztoláshoz rendelkezünk a tevékenység időbélyegzőjével, valamint a poszt tényleges tartalmával, a tevékenységben részt vevő barátok (maszkolt) nevével, a felhasználók által esetlegesen megosztott linkekkel és a poszthoz fűzött kommentek számával (ha van ilyen). Ezen túlmenően az adatok további információkat is tartalmaznak, például annak az eseménynek vagy csoportnak a nevét, amelyre/amelyre vonatkozóan a felhasználó posztolt/feltöltött/megosztott.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Érdeklődési kategóriák: A Facebook minden felhasználót kategorizál és ezt a kategorizálást használja a hirdetési optimalizálásra. A besorolás teljesen algoritmikus és nem pontosan ismert, hogy milyen adatokat használ fel. Az elérhető információkból arra lehet következtetni, hogy a besorolásban szerepe, van a saját kedveléseknek (p.l: oldalak), a tevékenységeknek, a használt keresési szavaknak és még a barátok preferenciáinak is (DeVito 2017). De mivel az algoritmus egy fekete doboz, így csak a kategorizálás eredményét tudjuk pontosan megfigyelni. Ebben az adathalmazban nincs időbélyegző, csak a kategórianevek elérhetők felhasználónként. A 150 résztvevő 105 642 érdeklődési kategóriához kapcsolódott (18 689 egyedi kategória). A mintánkban szereplő résztvevők átlagosan 704 kategóriához kapcsolódtak. A kategóriák között vannak általánosabbak (pl. étel) és specifikusabbak is (Pizza).

Az adatok rövid bemutatása után arra térünk ki, hogy milyen elemzéseket tud támogatni egy ilyen jellegű adatsor:

ELEMZÉSI LEHETŐSÉGEK

Az előző fejezet jól illusztrálja, hogy milyen gazdag adatbázis keletkezik egy adatnációs megoldást használó kutatásban. A Facebook-archívum csak az adatok egyik felét fedik le, ezekhez jönnek még a klasszikus survey adatok is. Utóbbi persze nem kötelező, de a dizájn miatt gyakorlatilag értelmetlen kihagyni.

Módszertani oldalról legalább három lehetséges elemzési irányt felrajzolhatunk. Az egyik irány a Facebook és survey adatok összevetése érvényességi oldalról. Erre jó példa az a tanulmányunk, amelyben a Facebook-tevékenységek alapján azonosítható zenei ízlést vetettük össze a kérdőívben mért zenei preferenciákkal (Kmetty – Németh 2020). Az ilyen jellegű kutatások fontos ismereteket nyújtanak a különböző adatsorok érvényességi korlátairól, de segíthetnek abban is, hogy milyen elvek mentén érdemes a Facebook-tevékenységek alapján operacionalizálni egyes kérdéseket.

Egy másik triviális megközelítés, ha a két adatsorral kiegészítjük egymást. A közösségi média aktivitás és depresszió összefüggését vizsgáló tanulmányunkban (Kmetty – Bozsonyi 2021) hasonló logikát követtünk. A kutatásban a depresszió szintjét egy a survey adatbázisban egy kérdőív blokkal mértük. Ez volt a függő változó a modelljeinkben. A Facebook-adatokból pedig olyan indikátorokat alakítottunk ki, amelyek feltételezésünk szerint összefügghetnek a depresszióval, mint például a posztolás gyakorisága, a posztolás időbeli változása vagy a kutatási résztvevők érdeklődési körei. Minden módszernek megvan a saját limitációja arra vonatkozóan, hogy mit lehet és mit nem lehet vele mérni. Ha jól kombináljuk a módszereket, akkor olyan összefüggésekig tudunk eljutni, ami a korábbi kutatási megközelítésekben elérhetetlen volt.

A harmadik irány, amit kiemelnénk elsősorban a közösségi média adatokat használja és nem nyúl a survey adatokhoz, vagy csak nagyon korlátozott mértékben.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Ezeknek az elemzéseknek a fő fókuszja a közösségi média adatok dinamikai szempontú elemzése. Mivel a legtöbb tevékenységhez időbélyeg társul, ezért az adatsort nem csak statikus modellekben lehet vizsgálni. Ezt a logikát követi például az a munkánk (Kmetty et al. 2020) amelyben a közösségi média aktivitás időbeli dinamikáját elemezzük. A Facebook-használókra jellemző, hogy egy szűk időintervallumon belül nem keverik a tevékenységeket, tehát az aktivitásuk „vonatszerű”. Elemzésünkben ezeket a vonatokat azonosítjuk és vizsgáljuk azt, hogy mitől függ az, hogy valaki tevékenységi típust vált.

A három felsorolt példakutatás természetesen nem fedi le az összes lehetséges irányt, de jól illusztrálja, hogy milyen elemzésekben gondolkozhatunk akkor, ha hozzáférünk donáció alapú közösségi média adatokhoz. Ha egy kutatáson belül több adatforrást is megoszt velünk a kutatási alany (pl. Facebook + Instagram), akkor az elemzési lehetőségek még egy további dimenzióval bővülnek. Várhatóan a következő években meg fognak jelenni ilyen komplex adatfelvételek is.

Összegzés

A digitalizáció folyamatosan alakítja a társadalmunkat. Megváltoznak a kapcsolattartási szokásaink, a médiafogyasztásunk, a szabadidő eltöltési szokásaink, vagy akár a munkánk jellege is. Ez a változás nemcsak tartalmi szempontból érdekes a társadalomtudományoknak, hanem módszertani szempontból is. A digitalizáció új típusú adatokat generál, a digitális kutatások infrastrukturális környezetének dinamikus fejlődése pedig új módszertani arzenált biztosít a kutatóknak. A digitális adatok azonban a lehetőségek mellett rengeteg kihívással és kérdéssel is járnak. Hogy tudunk sok adatot feldolgozni, mire lesz „reprezentatív” az elemzésünk, működnek-e a standard módszereink, hogyan tudjuk megtalálni a szociológia szempontból érdekes tartalmat egy nagyon zajos adatban? De talán a legfontosabb kérdés az, hogyan tudunk jó minőségű adathoz jutni? Erre a dilemmára vázoltam fel egy lehetséges megoldást, a donáció alapú adatgyűjtést. Ebben a megközelítésben a kutató nem a platform szolgáltatókkal áll kapcsolatban, hanem közvetlenül a kutatási alanyokkal és tőlük „szerzi” be az adatokat. Ez egyrészt egy jogilag sokkal tisztább adathozzáférési mód, hiszen adott benne a résztvevő beleegyezése. Másrészt a hozzáférhető adatok köre is sokkal gazdagabb, mint például egy standard API-alapú hozzáférésnél. És akár arra is van lehetőségünk, hogy egy survey kutatással kombináljuk ezeket az adatfelvételeket. Cserébe sokkal több energiát kell egy ilyen adatfelvétel megszerzésébe befektetni és a mintanagyságok is jóval kisebbek lesznek.

A GDPR szabályozás gyakorlatilag kikényszerítette az összes nyugati világban aktív közösségi média platformtól, hogy egyszerű módon tegye elérhetővé a felhasználóknak a róluk gyűjtött adatokat. Ez megteremtette annak is a lehetőségét, hogy kutatóként hozzáférjünk Facebook, Instagram vagy Google adatokhoz, ha meg tud-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

juk győzni a felhasználókat arról, hogy átadják nekünk ezeket az adatokat. A donáció alapú adatgyűjtések még gyerekcipőben járnak, ezért jelenleg kevés gyakorlati példa van ezekre a kutatásokra. Ritka kivétel az az ELTE Társadalomtudományi Karán zajló pilot kutatás, amelyben 150 aktív Facebook-használótól gyűjtöttünk adatokat. Ennek a kutatásnak a technikai lebonyolítását részletesen bemutattuk jelen tanulmányban.

A donáció alapú adatgyűjtés jövője részben a platform szolgáltatókon múlik. Ha engednek a szigorú API hozzáférési szabályokon akkor kisebb lesz az igény donáció alapú adatgyűjtésre. Bár az adatdonációnak számos előnye van, a gyors és nagy adatmennyiséget generáló API-kal nehéz felvenni a versenyt. Az utóbbi évek tendenciái azonban inkább a hozzáférés további szigorítását vetítik előre, nem az enyhítését. Az adatdonációs modell elterjedését szintén befolyásolhatja, hogy az adott országban milyen közösségi média oldalakat használnak. A Twitter továbbra is nagyon nyitott a kutatók felé adathozzáférési szempontból, tehát azokban az országokban, ahol a Twitter nagyon elterjedt (pl. USA) kisebb a nyomás a kutatókon, hogy más platformokat is monitorozzanak. Az is fontos szempont lehet, hogy az adott ország állampolgárai mennyire adattudatosak és mennyire félnek az adataikkal való visszaéléstől. Ha valahol alacsony az adatmegosztási bizalom, akkor könnyen kudarcba fulladhat egy adatdonációs projekt. A magyar pilotban nem gyűjtöttünk adatot a visszautasításról. A német GESISS kutatóközpontnak viszont vannak ilyen jellegű adatai. A GESIS kutatói egy olyan modellt teszteltek, amiben egy böngésző kiegészítőt kellett telepítenie a résztvevőnek a gépére, ami gyűjtötte, hogy milyen tartalmak jelennek meg a résztvevő Facebook-falán. Ez bár nem DDP, hanem böngésző alapú adatgyűjtés, de ugyanúgy szükséges hozzá a kutatási alany aktív hozzájárulása. Az online Facebookozó minta 60%-a engedélyezte a böngésző kiegészítő feltelepítést, de a gyakorlatban csak 40%-a telepítette az eszközt. Az elutasítók felének voltak adatvédelmi aggályai (Breuer et al. 2021). Felmerülhet alternatív megoldásként, hogy ezek az adatfelvételek ráépülnek olyan online panelekre, ahol már amúgy is magas az adatmegosztási hajlandóság. Ez azonban nem feltétlen segít a donáció alapú adatgyűjtések külső érvényességén.

Az itt felsorolt nehézségek ellenére azt gondolom, hogy a közeljövőben egyre több donáció alapú adatgyűjtés fog indulni. A közösségi média kikerülhetetlen szerepet tölt be a modern társadalmak életében. A társadalomtudomány pedig nem kerülheti meg, hogy vizsgálja azt, hogy ezeken a platformokon mi történik. Ezt pedig akkor tudja legjobban megtenni, ha rálát a felhasználók tényleges tevékenységére. Ebben pedig nagy segítséget tud lenni az adatdonációs megközelítés.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Irodalom

- Boeschoten, L. – Ausloos, J. – Moeller, J. – Araujo, T. – Oberski, D. L. (2020): Digital trace data collection through data donation. arXiv preprint arXiv:2011.09851.
- Breuer, J. – Kmetty, Z. – Haim, M. – Stier, S. (2021): User-focused approaches for collecting Facebook data in the “post-API age. Kézirat
- Brown, R. C. – Fischer, T. – Goldwisch, A. D. – Keller, F. – Young, R. – Plener, P. L. (2018): # cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychological medicine*, 48(2), 337-346. (doi.org/10.1017/S0033291717001751)
- Bruns, A. (2019): After the ‘APocalypse’: Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544–1566. (doi.org/10.1080/1369118x.2019.1637447)
- Csepeli Gy. (2015): Szociológia és a Big Data. *Replika*, (92-93), 171-176.
- Dessewffy T. – Láng L. (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika*, (92-93), 157-170.
- DeVito, M. A. (2017): From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753-773. (doi.org/10.1080/21670811.2016.1178592)
- Freelon, D. (2018): Computational research in the post-API age. *Political Communication*, 35, no. 4: 665-668. (doi.org/10.1080/10584609.2018.1477506)
- Haim, M. – Nienierza, A. (2019): Computational observation. *Computational Communication Research*, 1(1), 79-102. (doi.org/10.5117/CCR2019.1.004.HAIM)
- Halavais, A. (2019): Overcoming terms of service: a proposal for ethical distributed research. *Information, Communication & Society*, 22: no. 11, 1567-1581. (doi.org/10.1080/1369118X.2019.1627386)
- King, G. – Persily, N. (2020): A new model for industry-academic partnerships. *PS: Political Science & Politics*, 53(4), 703-709. (doi.org/10.1017/S1049096519001021)
- Kmetty Z. (2018): A szociológia helye a Big Data paradigmában és a Big data helye a szociológiában. *Magyar Tudomány*, 179(5), 683-692.
- Kmetty Z. – Bozsonyi, K. (2021): Identifying depression-related behavior on Facebook – an experimental study. Kézirat
- Kmetty Z. – Karsai M. – Koltai J. – Kertész J. (2020): Context of burstiness in Facebook activities. IC2S2 Conference, 17-20 July, 2020, poszter.
https://fbpilot.tatk.elte.hu/media/e4/18/4f92a3175714df1320bfd520c44d9ecd12f61d455866b06af7b23b616a/Burstiness_ic2s2_poster.pdf (utolsó letöltés: 2021. 11. 20.)



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Kmetty Z. – Németh R. (2020): Which is your favorite music genre? A validity comparison of Facebook data and survey data. arXiv preprint arXiv:2002.00501.
- Koltai, J. – Kmetty Z. – Bozsonyi K. (2021): From Durkheim to Machine Learning: Finding the Relevant Sociological Content in Depression and Suicide-Related Social Media Discourses. In *Pathways Between Social Science and Computational Social Science* (pp. 237-258). Springer, Cham.
(doi.org/10.1007/978-3-030-54936-7_11)
- Lazer, D. – Radford, J. (2017): Data ex machina: introduction to big data. *Annual Review of Sociology*, 43, 19-39. (doi.org/10.1146/annurev-soc-060116-053457)
- Mancosu, M. – Vegetti, F. (2020): What you can scrape and what is right to scrape: A proposal for a tool to collect public Facebook data. *Social Media+ Society*, 6(3), 2056305120940703. (doi.org/10.1177/2056305120940703)
- Moreno, M. A. – Ton, A. – Selkie, E. – Evans, Y. (2016): Secret society 123: Understanding the language of self-harm on Instagram. *Journal of Adolescent Health*, 58(1), 78-84. (doi.org/10.1016/j.jadohealth.2015.09.015)
- Puschmann, C. (2019): An end to the wild west of social media research: A response to Axel Bruns. *Information, Communication & Society*, 22(11), 1582-1589. (doi.org/10.1080/1369118X.2019.1646300)
- Reece, A. G. – Danforth, C. M. (2017): Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 1-12.
(doi.org/10.1140/epjds/s13688-017-0110-z)
- Ságvári B. (2017): Társadalomtudomány a Big Data korában. *Statistikai szemle*, 95(5), 491-504 (doi.org/10.20311/stat2017.05.hu0491)
- Scharkow, M. (2016): The accuracy of self-reported internet use – A validation study using client log data. *Communication Methods and Measures*, 10(1), 13-27. (doi.org/10.1080/19312458.2015.1118446)
- Shapiro, E. H. – Sugarman, M. – Bermejo, F. – Zuckerman, E. (2021): New approaches to platform data research. Netgain partnership. (last accessed: 15.03.2021) <https://www.netgainpartnership.org/resources/2021/2/25/new-approaches-to-platform-data-research> (utolsó letöltés: 2021. 11. 20.)
- Stier, S. – Breuer, J. – Siegers, P. – Thorson, K. (2020): Integrating survey data and digital trace data: Key issues in developing an emerging field. (doi.org/10.1177/0894439319843669)
- Tromble, R. (2021): Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age. *Social Media+ Society*, 7(1), 2056305121988929. (doi.org/10.1177/2056305121988929)



A felügyelt gépi tanulás kihívásai a szociológiai alkalmazásokban

NÉMETH RENÁTA¹

ABSZTRAKT

Az ipari/üzleti alkalmazásokban már sokszorosan bizonyított felügyelt gépi tanulás szociológiai alkalmazásai sajátos kérdéseket vetnek fel. A sajátosság oka, hogy ezekben az alkalmazásokban komplex fogalmak megtanulása az algoritmus feladata (pl. hogy gyűlöletbeszédet tartalmaz-e egy tweet). A felügyelt tanulás lényege, hogy előre bekódolt (gyűlöletbeszéd/nem gyűlöletbeszéd) szövegek címkézését tanulja meg az algoritmus, jellegzetes szövegmintázatokat keresve. A felmerülő kérdések: hogyan jön létre a címkézés? Hogyan lehet betanított kódolókkal elvégeztetni egy olyan hermeneutikai kihívást, mint a gyűlöletbeszéd felismerése? Segítenek-e ezen a rutinszerűen alkalmazott, részletezett annotálási irányelvek? A cikk arra is kitér, hogyan végzik crowdsourcing platformokon a kódolást a nagy cégek, illetve ismertetem az MI-torzítást is, aminek itt az a lényege, hogy a kódolók maguk viszik be a diszkriminációt az adatokba. E kérdéseket kutatási tapasztalatainkkal illusztrálok.

KULCSSZAVAK: felügyelt gépi tanulás, annotálás, crowdsourcing, MI-torzítás

ABSTRACT

The challenges of supervised machine learning in sociological applications

The sociological applications of supervised machine learning, already well proven in industrial/business applications, raise specific questions. The reason for this specificity is that in these applications, the algorithm is tasked with learning complex concepts (e.g. whether a tweet contains hate speech). Supervised learning consists of learning to classify previously annotated (hate speech/non-hate speech) texts by the algorithm, looking for characteristic text patterns. The questions that arise are: how to prepare annotation? How can a hermeneutic challenge such as hate speech recognition be performed by annotators? Are routinely applied, detailed annotation guidelines helpful? The article also discusses how large companies perform coding on crowdsourcing platforms, and describes AI bias, which in this case means that annotators themselves introduce bias into the data. I illustrate these issues with our own research experiences.

KEYWORDS: supervised machine learning, annotation, crowdsourcing, AI bias

¹ Eötvös Loránd Tudományegyetem, Társadalomtudományi Kar, Research Center for Computational Social Science, email:nemeth.renata@tatk.elte.hu. A szerzőt a K-134428 azonosítójú NKFIH pályázat támogatta kutatási munkájában.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Bevezetés

A weboldalakon, közösségi média felületeken, szöveggé alakított videókon, digitalizált könyvtárakban felhalmozódott szövegek a társadalmi valóság sosem látott szélességű elérését teszik lehetővé. Individuális és kollektív cselekedeteinkről, szerepeinkről és csoportközi viszonyainkról, mindezek időbeli dinamikájáról és strukturális heterogenitásáról naponta többmillió terrabytenyi digitális szöveg keletkezik, s a meglévő adatvagyon néhány évente megsokszorozódik.

A digitális szövegeknek ez az univerzuma párhuzamosan olyan szöveganalitikai technológiákat hívott életre, melyek ennek az elérésnek ma már a szociológia számára is releváns mélységét nyújtják. Savage és Burrows 2007-ben az évtized egyik legtöbbet idézett szociológiai tanulmányában az empirikus szociológia közelgő válságáról írt. Azt jóslták, hogy válság következik be, ha a korábban saját módszertani szakértelméről ismert szociológia nem tud megfelelni a big data által támasztott kihívásoknak, és így elveszíti vezető szerepét. Ez nem következett be. Nyolc évvel később a British Sociological Association által kiadott „Sociological Futures” című könyvsorozat első tagja (Ryan – McKie 2015) már a címében is utalt a válság végére, és fontos lehetőségeket látott a big data kutatásban, valamint az szövegbányászatban is. Ugyanakkor a korábban a szociológia által uralt empirikus szakértelmet tekintve egyértelmű elmozdulás figyelhető meg az akadémiai szférából az ipar felé, hiszen a terület hatalmas üzleti lehetőségeket generál és az ipar finanszírozni is képes a szükséges fejlesztéseket.

Ez a cikk a szövegbányászat egy alterületét mutatja be. A szövegbányászat izgalmas, és a szociológia számára is nagyon perspektivikus terület az informatika, mesterséges intelligencia-kutatás és nyelvészet határán. Utóbbi években tapasztalt fel-futása mögött egyszerre áll a digitális szöveges források elérhetősége és a hatalmas adatbázisok feldolgozásához szükséges technológia fejlődése. Ambiciózus projektek használják üzleti alkalmazásokban, marketingben, kormányzati és védelmi területen egyaránt. Az elmúlt néhány évben a szövegbányászat a társadalomtudományokban is kezdett teret nyerni, az antropológiától a közgazdaságtanon át a szociológiáig². Természetesen, míg a szövegbányászat egy viszonylag új interdiszciplináris terület, maga a szövegelemzés több évtizedes tradícióra tekint vissza a társadalomkutatásban: kifinomult kvalitatív módszerek fejlődtek ki. A kvantitatív társadalomkutatási szövegelemzés inkább csak bizonyos kifejezések vagy kvalitatíve azonosított kódok megjelenését számszerűsítette a szövegekben, ehhez képest óriási előrelépést jelent a szövegbányászat eszköztára, ami olyan feladatok elvégzését teszi lehetővé, mint tematikus struktúrák vagy látens szemantikai relációk automatizált azonosítása (Németh – Katona – Kmetty 2020).

² Társadalomkutatónak szóló kiváló összefoglaló a módszerről: Ignatow – Mihalcea 2017.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Zárójelben megjegyezném, hogy a szövegbányászattal kapcsolatban sem az angol, sem a magyar terminológia nem kiforrott még, a természetesnyelv-feldolgozás (NLP, natural language processing), számítógépes nyelvészet, automatizált szövegelemzés diffúz körvonalakkal bír, rokon, de nem szinonim elnevezések (Németh – Katona – Kmetty 2020).

Ebben a tanulmányban azokat a lehetőségeket tekintem át, amelyeket a felügyelt gépi tanulás nyújthat a szociológia számára. Ezeknek az ipari/üzleti alkalmazásokban már sokszorosan bizonyított algoritmusoknak a szociológiai alkalmazásai sajátos kérdéseket vetnek fel. A sajátosság oka, hogy ezekben az alkalmazásokban komplex fogalmak megtanulása az algoritmus feladata. A felügyelt tanulás lényege, hogy előre bekódolt szövegek címkézését tanulja meg az algoritmus, jellegzetes szövegmintázatokat keresve. A felmerülő kérdések: hogyan jön létre a címkézés? Hogyan lehet betanított kódolókkal elvégeztetni egy olyan hermeneutikai kihívást, mint a gyűlöletbeszéd felismerése? Segítenek-e ezen a rutinszerűen alkalmazott, részletezett annotálási irányelvek? Jobban végzi-e a kutató a besorolást, mint az egyszerű kódoló, vagyis magasabbrendű-e az egyik interpretáció, mint a másik? A cikk arra is kitér, hogyan végzik crowdsourcing platformokon a kódolást a nagy cégek, hogy működik iparszerűen ez a humán/gép együttműködés, és milyen kérdések merülnek fel a crowdsourcing interpretáció kapcsán. Végül röviden kitérek az AI-torzításra, aminek itt az a lényege, hogy a kódolók maguk viszik be a diszkriminációt az adatokba.

Felügyelt vs. felügyelet nélküli gépi tanulás jellemzői

Az alábbiakban röviden ismertetem a felügyelt és felügyelet nélküli tanulás logikáját, majd a felügyelt tanulás gyakorlati kivitelezését. Nem térek ki a módszer szöveganalitikai jellegzetességeire, arra, hogy mi a „szöveg”, mint input adat sajátossága a numerikus inputtal szemben, vagy hogy milyen módon keresnek a modellek mintázatokat a szövegekben. Utóbbi kérdések iránt érdeklődő olvasóknak a Németh – Koltai (2021) szerzőpárosához, illetve Németh, Katona és Kmetty (2020) szerzők szociológusokhoz szóló szövegbányászati bevezetőjét ajánlom.

A felügyelt és felügyelet nélküli tanulás mögött álló egzakt statisztikaelméletet Vapnik (2000) alaposan tárgyalja. Intuitíve, a felügyelt és felügyelet nélküli tanulás közötti különbség azon alapszik, hogy már létező elmélet/meglevő háttérismeret empirikus megnyilvánulásait keressük (felügyelt tanulás, ahol a „felügyelet” maga a háttérelmélet), vagy induktív módon egy még nem vizsgált téma feltárása a cél (felügyelet nélküli tanulás). A dichotómia hasonlít a klasszikus szociológia konfirmatív/exploratív módszerek közötti különbségtételéhez.

A nem felügyelt módszerek némelyikét a klasszikus társadalomkutatásban is gyakran alkalmazzák, ilyen például a szövegekre alkalmazott klaszterelemzés. Ez olyan vektortérben definiálható, ahol a tengelyek a szavakat jelölik, a dokumentu-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

mok vektortérbeli helyét pedig az egyes szavak dokumentumbeli gyakorisága határozza meg. De nem felügyelt módszer a legújabb szövegbányászati megoldások alapját adó szóbeágyazás³ is. Ez a módszer az egyes szavakhoz rendel jelentést, még pedig un. disztribúciós szemantikai megközelítésben, amikor is a jelentést kizárólag a használati környezet határozza meg. A jelentésről tehát kizárólag címkézetlen adatokból tanul az algoritmus, nincs szükség kívülről bevitt szakértői tudásra. A disztribúciós szemantika használható a legkülönbözőbb területek hatalmas skáláján, így például a gépi fordításnál is.

Ezzel szemben a felügyelt tanulás lényege, hogy (elméletünk/háttérismeretünk-re támaszkodva) előre bekódoljuk a szövegeket, majd ezeket a címkéket az algoritmus megpróbálja megtanulni. A kutató elméleti megfontolásai befolyásolják az elemzést, hiszen a címkék, mint kategóriák meghatározása megelőzi az elemzést. Egy társadalomkutatói példa Poletti és társai (2017) munkája, akik olasz nyelvű Twitter-üzeneteken igyekeztek automatikus gyűlöletbeszéd-felismerőt létrehozni. A gyűlöletbeszéd általuk alkalmazott definíció szerint valamely kisebbségi csoport ellen irányul és tartalmaz egyfajta illokúciós erőt, amely alkalmas a célcsoporttal szembeni erőszak terjesztésére, népszerűsítésére, alátámasztására vagy erre történő uszításra. Ezért a fogalom pontos megragadása érdekében a kódoknak tartalmaznia kellett a célcsoportot (vallási, etnikai kisebbségek vagy migránsok), és a kódolók a tweetet olyan jegyekkel is felruházták, mint hogy sztereotipizál-e, tartalmaz-e agressziót, támadó-e, vagy tartalmaz-e az agressziót kendőző iróniát.

A megtanulás itt azt jelenti, hogy (viszonylag alacsony hiba mellett) az algoritmus maga is képes lesz kódokat (gyűlöletbeszéd/nem gyűlöletbeszéd) rendelni még címkézetlen szövegekhez. A tanulási folyamat pedig a kutatók által címkézett két szöveghalmaz eltérését leginkább megfogó jellegzetes szövegmintázatok keresésén alapul, ahol az előre címkézett szövegek halmazát tanuló-halmaznak nevezzük. Nagyon leegyszerűsített példán: használhatunk logisztikus regressziót, ahol a függő változó a bináris címke (gyűlöletbeszéd/nem gyűlöletbeszéd), numerikus magyarázó változóink pedig azt jelölik, hogy az adott nyelv szavai hányszor fordulnak elő a szövegben; a cél a legjobban illeszkedő modellhez tartozó együttthatók megtalálása. A gyakorlatban persze ennél jóval komplexebb modelleket alkalmaznak, komplexebb (például stiláris jegyeket reprezentáló vagy a szöveg mélyebb szemantikai viszonyait megragadó) magyarázó változókkal, a nagy mennyiségű (több ezer) magyarázó változó miatt dimenziócsökkentő megoldásokkal és a túlillesztést elkerülni hivatott megoldásokkal, de az algoritmus létrejöttének logikája ott is hasonló.

A felügyelt osztályozás gyakran használt típusa a szentiment – vagy érzelem-elemzés. Egy szöveg szentimentje a szerző attitűdje egy tárgyhoz (pozitív, negatív vagy semleges), míg az érzelmek a boldogságtól a haragig terjedő érzések.⁴ Ez az

³ Angolul word embedding, a módszerről lásd: Németh – Katona – Kmetty 2020.

⁴ Összefoglaló az alkalmazott módszerekről: Nahili – Rezeg – Kazar 2020.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

alkalmazás mind üzleti, mind tudományos alkalmazásokban nagyon elterjedt. Szöveges adatok szentiment- vagy érzelem szerinti címkézésével vizsgálhatják például marketingesek, hogy hogyan reagálnak a felhasználók reklámokra, szolgáltatásokra vagy termékekre, digitális bölcsészek, hogy hogyan változnak a regényben megjelenített érzelmek, vagy szociológusok, hogy hogyan terjednek érzelmek és vélemények társadalmi hálózatokban.

A felügyelt módszerek abban támogatják tehát a kutatókat, hogy nagyobb szövegkorpuszt rövidebb idő alatt dolgozhassanak fel, mint amire az emberi kapacitás képes lenne. További (inkább tudományos elvárásokhoz igazodó) cél, hogy megértsük az automatikus címkézés mögött álló szabályokat, egyszerű példán: hogy lássuk, mely szóhasználat valószínűsíthető leginkább a gyűlöletbeszéd jelenlétét. Ez utóbbi cél tulajdonképpen azoknak a tartalmi-szociológiai interpretációra/magyarázatra lehetőségét adó jegyeknek a feltárását jelenti, amelyek az egyébként „fekete dobozként” működő kódoló algoritmus mögött állnak⁵.

A felügyelt tanuláshoz több új továbbfejlesztése létezik, ahol az eredeti logika kissé módosul. Így pl. az aktív tanulás esetén nem egy fix tanuló-halmazunk van, hanem az algoritmus maga kéri menet közben konkrét, még címkézetlen, de a tanulásban fontosnak tűnő szövegek címkézését, vagy a transzfer tanulás, amikor a címkézett adatoktól eltérő besorolási problémát kell a tanuló algoritmusnak megoldania (Eisenstein 2019).

A felügyelet nélküli módszerek a felügyelt technikákkal együtt is használatosak. Például a felügyelet nélküli módszerek hatékony alkalmazása, ahogy fentebb említettük, gyakran dimenziócsökkentést igényel, s ezek a dimenziócsökkentési eljárások gyakran felügyelet nélküliek. Így egy felügyelt tanulási regressziós modellben magyarázó változóként használhatjuk a klaszterelemzés által adott besorolást. De ugyanígy gyakori megoldás az, amikor felügyelt modellben a szöveget alkotó szavak helyett a szóbeágyazási modell által adott szójelentéseket használják magyarázó változóként.

A felügyelt gépi tanulás inputja: humán annotálás

Két karakteresen különböző megoldás létezik a tanuló-halmaz létrehozására. Az egyikre példa a fent már idézett, Poletti és társai (2017) által publikált kutatás: ők kódolókat (szövegbányászati terminussal: annotátorokat) tanítottak be részletes irányelvek alkalmazásával arra, mikor minősítsenek egy szöveget agresszívnek vagy támadónak, mikor sztereotipizál és mikor irányul egy kisebbségi csoport ellen a szö-

⁵ Erről a magyarázati igényről és általában a magyarázatra/predikcióra való törekvés kettőségéről lásd: Németh 2021: 110–121.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

veg. Itt tehát humán annotátorok vannak, akik olvasnak és interpretálnak, adott annotálási irányelveket követve.

A másik lehetőségre Jelveh, Kogut és Naidu (2014) írása példa, akik amerikai közgazdászok ideológiai pozíciójának (jobboldali/baloldali) gépi tanulását végezték el a szerzők tudományos írásai alapján. A tanuló-halmaz azokból a közgazdászoknak az írásaiból állt, akiknek pozíciója megállapítható volt politikai kampány-támogatásokat, illetve petíció aláírásokat tartalmazó nyilvántartások alapján. Itt tehát nincsenek humán annotátorok, nem olvasunk és nem interpretálunk, mert kész címkéink vannak.

A gyakorlatban leggyakrabban egy kutató (vagy egy kutatócsoport) kézzel végzi a kódolást, a szöveg annotálását, akárcsak a „klasszikus” kvalitatív szövegelemzés esetében. Az annotálás minőségének jelentősége kiemelkedő: a felügyelt tanuló algoritmus jó minőségben annotált tanuló-halmazból tud hatékonyan tanulni. Az annotálás persze időigényes és gyakran nem is egyszerű feladat. Az elsődleges cél a replikálhatóság, ami azt jelenti, hogy egy másik annotátor nagyon hasonló annotációkat készítené. Hovy és Lavid (2010) a következő strukturált eljárást javasolja az annotációk előállítására:

1. Határozzuk meg, hogy milyen kategóriákba sorolva kell annotálni. Ez általában valamilyen elmélet alapján történik (lásd: milyen jegyekkel definiálható a gyűlöletbeszéd). Itt megfelelő egyensúlyt kell találni a részletezettség/pontoság és a skálázhatóság/idővonzat között.

2. Választhatunk olyan platformot, amely támogatja az annotálás adminisztrálását. Több általános célú annotációs eszköz érhető el.

3. Az annotálási feladatra vonatkozó utasítások formalizálása annotálási irányelvek formájában. Amennyiben az utasítások nem explicitek, a kapott annotációk szubjektív benyomásokon alapulnak majd, ami a replikálhatóságot veszélyezteti.

4. Az adatok egy kis részhalmazának kísérleti annotálása (pilot), több annotátorral. A pilot előzetes benyomást ad mind a megismételhetőségről, mind az annotálási irányelvek alkalmazhatóságáról. A megismételhetőséget az annotátorok döntései közötti egyezés (inter-annotator agreement) mutatóival jellemezhetjük. A konkrét annotálási eltérések vizsgálata segíthet az utasítások pontosításában, és az annotálási feladat módosításához is vezethetnek.

5. Fő annotálás. Érdemes legalább az adatbázis egy részét párhuzamosan annotálni, azaz két vagy több annotátorral egymástól függetlenül besorolítani, hogy az annotátorok közötti egyezés kiszámítható legyen. Sok projektben a szövegek több címkét is kapnak, amelyek aztán összesítve egy „konszenzusos” címkévé állnak össze.⁶

⁶ Példaként lásd: Danescu – Niculescu – Mizil et al. 2013.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

6. A felügyelt tanulás középpontjában az annotátorok közötti egyetértés áll: ha a kódok nem megbízhatóak, a tanuló algoritmus nem tud hatékonyan tanulni belőlük, és besorolásai sem lesznek megbízhatóak. Ezért elengedhetetlen az annotálás értékeléseként az annotátorok közötti egyezés mutatójának kiszámítása. Ha ez a mutató alacsony, az az annotátorok megbízhatóságát vagy magát a teljes annotációs rendszert (a besorolás értelmességét) kérdőjelezi meg. A mutató definíciójára több matematikai megoldás létezik a konkrét feladat függvényében. Az annotátorok besorolásainak egyszerű százalékos egyezése, annak ellenére, hogy nagyon széles körben használják, nem veszi figyelembe a véletlenszerűen előforduló egyezést. Ugyanis, ha két kommentátor véletlenszerűen választ két címke között, akkor a köztük lévő egyetértés várhatóan 50%-os lesz. Ezért egy jó mutató a nyers egyezési százalékot a véletlen egyezés arányához viszonyítja. Egy ilyen széles körben használt mutató a Cohen-féle kappa is.

Érdeemes megjegyeznünk, hogy ha az általunk vizsgált eset nem túl specifikus, nem feltétlenül kell saját annotálást végrehajtanunk, használhatunk mások által annotált adatbázisokat is. Számptalan annotált, nyílt elérésű korpusz található az interneten, melyek elsősorban nyelvészeti feladatokra alkalmazhatók, de találhatóunk Twitter bot-detektálásra szolgáló címkézett adatbázisokat is, sőt társadalomkutatási célúakat is, mint a Manifesto projekt, amely ötven ország politikai pártjainak választási programjainak annotált adatbázisát nyújtja 1945-től napjainkig.

Egy saját kutatási példa

Sik Domonkossal és Máté Fannival végzett kutatásunkban (Németh – Sik – Máté 2020) mi is a fentihez hasonló eljárást dolgoztunk ki. Kutatásunk célja az volt, hogy különböző felügyelt tanulói algoritmusok alkalmazásával automatikusan osztályozzuk ismert nemzetközi online depresszió-fórumok bejegyzéseit aszerint, hogy abban a depresszió milyen (bio-medikális, pszichológiai vagy társadalmi) keretezését adja a felhasználó.

Az annotátorokat a téma és a módszer iránt érdeklődő társadalomtudományi szakos hallgatók közül választottuk ki. A teljes adatbázisból, ami 70 000 posztból állt, egyszerű véletlen mintavétellel 4500-at választottunk ki felcímkézendő tanulólálmazként. Több tréninget tartottunk az annotátoroknak, és sok valós példát felvonultató, részletes annotálási irányelveket készítettünk, amelyet a pilot szakasz után is, és a fő annotálási szakaszban is folyamatosan frissítettünk. Öt címkét használtunk, mivel a három keretezés-típus mellé a „besorolhatatlan” (depresszióról van szó, de a keretezés nem azonosítható) és az „irreleváns” (nem depresszióról van szó) is hozzáadódott. Az annotálási feladat a legkevésbé sem volt triviális, ezért (1) az annotátorok szükség esetén két címkét rendelhettek a szövegekhez, egy elsődleges és egy (opcionális) másodlagos címkét. A bejegyzések 34%-a kapott második címkét



TEMATIKUS TANULMÁNYOK – Digitális szociológia

legalább az egyik annotátortól. Továbbá (2) minden szöveghez két független annotátorunk volt, akik együttesen legfeljebb négy címkét adtak. A végső, konszenzusos címke „többségi szavazáson” alapult, tehát a négy címke közül a leggyakoribbat választottuk. A (nagyon kevés, 12,3%-os) kétértelmű esetek feloldására egy harmadik annotátort (egy vezető kutatót) kértünk fel. Másodlagos konszenzusos címkét is kiosztottunk, ha a „szavazásnak” volt egy értelmű második győztese.

Az annotátorok közötti egyezés mutatójának meghatározására a Cohen-féle kappát használtuk, amely azt mutatja meg, hogy az annotátorok mennyivel jobban teljesítenek a véletlenszerűen besoroló annotátorokhoz képest. Tökéletesen egyező besorolásoknál az értéke 1, ezzel szemben, ha az annotátorok véletlenszerűen választják ki a címkéket, akkor a kappá egyenlő 0-val.

Annotációink másik sajátossága az volt, hogy az annotátorok másodlagos címkéket rendelhettek a bejegyzésekhez. Ha az egyezést az elsődleges címkék egyezéseként határozzuk meg, egyszerűen elvetve a választható másodlagos címkéket, akkor egy túlságosan konzervatív mérőszámot kapunk. Ezért a kappá „liberális” változatát használtuk, az egyezést úgy definiálva, hogy az egyik elsődleges címke megegyezik a másik annotátor által adott elsődleges vagy másodlagos címkével. A liberális kappá előnye az eredeti, konzervatív változatával szemben az, hogy figyelembe veszi a szövegek másodlagos jelentését is. Míg a liberális kappá túl optimista képet mutathat, addig konzervatív megfelelője túl szigorú értékeléshez vezet. Az „igazság” valahol a kettő között van, ezért mindkettőt bemutattuk. Konzervatív módon mérve az annotátoraink közötti egyezés 58,3%, a liberális mérőszám pedig 69,7% volt, ami elfogadható mértékű egyezést mutat.

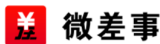
Crowdsourcing annotálás

A tanulás sikerét közvetlenül befolyásolja a tanuló-halmaz mérete – hasonlóan ahhoz, ahogyan survey-ek esetén a mintanagyság a megbízhatóság megfontosabb faktora. Ezért nem ritkák a kifejezetten nagy (több tíz- vagy százezres) elemszámú annotálandó adatbázisok. Ezekben az esetekben a crowdsourcing platformokon bérelhető bedolgozó annotátorok jelentenek megoldást, például az Amazon által működtetett Mechanical Turk, a Figure Eight,⁷ Lighttag vagy a kínai Weichaihi (1. ábra).

⁷ Korábbi elnevezése CrowdFlower.



TEMATIKUS TANULMÁNYOK – Digitális szociológia



THE PIONEER OF CROWDSOURCING IN CHINA

中文 | English

Home About Us

I'm A User

About Us

Home > About Us

About Us

Contact Us

Launched in Jan 2013, WeiChaiShi (WCS, 微差事 - Chinese word for 'Micro-Task') is the biggest crowd-tasking B2C platform in china. WCS is a mobile APP that instantly connects businesses to an on-demand mobile workforce who are incentivized to collect, capture, and report real-time data for brand clients. As of today, 3 million smart-phone registered users have completed more than 7.5 million simple tasks, acting as a go-to team to effectively connect consumers, stores, and brands.

The type of tasks includes commercial inspection, data collection, research & census and experiential marketing, etc. Through this new business model, survey companies significantly improve sampling coverage and efficiency, brands are able to create followers while lowering operations and marketing cost. In 2014, WCS won the award of 'Best business Model in China' judged jointly by '21st Century Business Review' and '21st Century Business Herald'.

WCS APP, It's EASY, It's FUN, and It PAYS!

1. ábra. A WeiChaiShi kínai crowdsourcing cég beköszönő weboldala. A bedolgozók mobiltelefonján keresztül adatgyűjtést, survey-eket és kísérleteket is kínálnak

Forrás: <http://www.weichaiishi.com/>

A crowdsourcing platformon annotáltatni kívánó kutatók egyszerűen felhívást tesznek közzé a platform bedolgozói között, melyben az annotáció darabjára kívülről az annotátorok elvárt képességeit (minimális képzettségét, anyanyelvét, korábbi munkáikkal kapcsolatos elégedettségi arányt stb.) adják meg (2. ábra). A viszonylag képzetlen „tömegmunkások” alkalmazása látszólag ellentétben áll a replikálhatóságra vonatkozó elvárásokkal, azonban számos, a platformot használó kutatás számol be megbízható annotációkról egyszerűbb feladatok kapcsán (Snow et al. 2008).

A crowdsourcing egy tágabb terület, az ember-alapú számítás (human-based computing, lásd még: elosztott gondolkodás, distributed thinking) részét képezi. Ez olyan számítástechnikai megoldásokat foglal magába, amelyekben egy számítógép úgy látja el a funkcióját, hogy bizonyos lépéseket embereknek szervez ki, egy szimbiózisszerű interakcióban (Mühlhoff 2019). Megfordulnak a szerepek: a gép kér fel embereket egy probléma megoldására, majd integrálja a megoldásokat. Az olyan számításigényes feladatok esetében, mint a képfelismerés, a humán annotáció fontos részét adja a mélytanuló algoritmusok tanításának.

A crowdsourcinggal kapcsolatban komoly etikai kérdések merülnek fel. Fort és Cohen (2011) már tíz éve arra figyelmeztetett, hogy az Amazon Mechanical Turk többszázézes bedolgozói tömeget működtet online, szerte a világban, jó részük él Indiában és Törökországban. A jellemző órabérek 2 dollár alatt vannak, és az általános elképzeléssel szemben a felhasználók jellemzően nem kismamák vagy diákok, aki hobbiként dolgoznak itt, hanem olyan munkavállalók, akik a megélhetéshez szükséges forrásként tekintenek munkájukra. Az alacsony bérrel kívül további ko-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

The screenshot shows the Amazon Mechanical Turk worker interface. At the top, there's a navigation bar with 'amazonmturk Worker', 'HITS', 'Dashboard', and 'Qualifications'. A search bar for 'All HITS' is present. Below this, a table lists 'HIT Groups (1-20 of 2106)'. The table has columns for Requester, Title, HITS, Reward, Created, and Actions. A tooltip is visible over the second row, indicating a language requirement.

Requester	Title	HITS	Reward	Created	Actions
Amazon Requester Inc. - C	[French language proficiency requir...	61,046	\$0.01	17h ago	Preview Accept & Work
Amazon Requester Inc. - C	[日本語能力が必 Questionnaire sur la relativité des produits aux intérêts (répondre par oui ou non)	28,379	\$0.01	7h ago	Preview Accept & Work
Amazon Requester Inc. - C	Product to Interest Audit (single yes/...	27,670	\$0.01	1h ago	Preview Accept & Work
Amazon Requester Inc. - C	[dominio del idioma español requeri...	19,719	\$0.01	21h ago	Preview Accept & Work
Amazon Requester Inc. - C	[Proficiência no idioma português br...	17,485	\$0.05	3m ago	Preview Qualify
Crowdsurf Support	Transcribe up to 35 Seconds of Med...	13,896	\$0.12	5d ago	Preview Accept & Work
TC Research	Find the Email for These Mental He...	12,180	\$0.50	1m ago	Preview Accept & Work
UnSpun Opinions	Opinion Survey	11,846	\$0.03	2/23/2018	Preview Qualify
KronoPin	Find the Website Address for a Con...	10,577	\$0.15	3d ago	Preview Accept & Work
Assistive Technology Rese	1 minute survey: Smart speakers at ...	10,557	\$0.01	1d ago	Preview Qualify
Armin Hamzic	Tell us if a picture shows a specific f...	8,217	\$0.05	10d ago	Preview Accept & Work
nttkAN	Image Annotation (WARNING: This ...				

2. ábra. Az Amazon Mechanical Turk egy oldala a felkínált munkákkal, a munkaadók (requesters) által fix darabbérért kínált un. HIT-ekre (Human Intelligence Tasks), elemi munkaegységekre osztott feladataival

Forrás: blog.mturk.com, <https://blog.mturk.com/quick-update-another-improvement-to-the-mturk-worker-experience-9cfd0b1963e7>

moly problémát jelent, hogy a munkavállalók (azaz „turkers”) híján vannak az olyan alapvető munkahelyi jogoknak, mint a kollektív alku, a szakszervezet alakításának lehetősége, a munkáltatói jogsérelmek orvoslásának lehetősége, ezért rendkívül kiszolgáltatottak.

Egy friss munka (Gray – Suri 2019) gazdaságantropológiai megközelítésben vizsgálja ezeknek a láthatatlanul dolgozó „szellemmunkásoknak” a körülményeit. A szerzők szerint az átláthatóság hiánya azt a benyomást kelti a közvéleményben, hogy a mesterséges intelligencia egyedül működteti a kortárs kényelmi szolgáltatásokat, miközben számos ponton elengedhetetlen az emberi beavatkozás: a szellemmunkások az annotáláson túl számos támogató munkát (személyazonosítás, feliratozás



TEMATIKUS TANULMÁNYOK – Digitális szociológia

stb.) végeznek olyan óriások számára, mint az Amazon, a Google, az Uber és a Microsoft. A munkával járó anonimitás és rugalmasság kétségtelenül sokuknak előnyös, pl. az egyébként gyakran diszkriminált munkavállalói csoportoknak (nők, fogyatékkal élők). Ugyanakkor Fort és Cohen (2011) korábbi megfigyelései még mindig érvényesek, azzal a különbséggel, hogy ma már dolgozók millióira vonatkoznak: a szellemmunkások rendkívül alulfizetettek, kiszolgáltatottak, munkavállalói jogaik erősen sérülnek. A szerzők javaslatainak (munkavállalói juttatások bevezetése, szakszervezet alakítása, a munkavállaló és munkáltató közötti emberi kommunikációt lehetővé tevő új központok létrehozása) tépje nagy: a méltányos emberi munka jövőjét biztosítanák.

A humán annotálás kihívásai szociológiai alkalmazásokban

Az ipari/üzleti alkalmazásokban már sokszorosan bizonyított felügyelt gépi tanulás szociológiai alkalmazásai sajátos kérdéseket vetnek fel. A sajátosság oka, hogy ezekben az alkalmazásokban komplex fogalmak megtanulása az algoritmus feladata (lásd: gyűlöletbeszédet tartalmaz-e egy tweet), szemben az olyan könnyebben annotálható feladattal, hogy negatív vagy pozitív-e egy szolgáltatással kapcsolatos megjegyzés.

Kutatásunk (Németh – Sik – Máté – 2020) már idézett esete jól példázza ezt: a depresszió keretezésének eldöntése nem bizonyult egyszerű feladatnak. A pilot során, az annotálási irányelvekben felsorolt elveken alapuló értelmezést több egyéni és csoportos fordulóban gyakorolták az annotátorok. Mégis, a pilot szakasz többszöri meghosszabbítása után is elégtelen maradt az annotátorok közötti egyetértés (az elsődleges címkék százalékos egyezése 60% alatt volt). Ezen a ponton vált időszerűvé saját módszertani háttérfeltevéseink felülvizsgálata. Rá kellett jönnünk, hogy az általános kategóriák (keretezés típusok) és a konkrét hozzászólások ilyen jellegű társítása nem egyértelmű feladat. Gadamer hermeneutikai elméletét (2004) követve innét kezdve a jelentések kialakulását nyelvileg közvetített, interszubjektív, iteratív folyamatnak tekintettük, amelyben a jelentések ténylegesen egy folyamat során konstruálódnak. Módszertanilag ezt az interszubjektív folyamat az annotálási irányelvek folyamatos, iteratív frissítésében jelent meg. Ez a folyamat lényegesen különbözik azoktól a hermeneutikailag egyszerűbb üzleti alkalmazásoktól, amelyek explicit és egyértelmű kategóriák (lásd pozitív/negatív/neutrális szentiment) előre meghatározott készletét alkalmazzák. A mi eljárásunk inkább egyfajta kvalitatív kódolásként határozható meg, mivel a kategóriáinkat egy előzetes absztrakt elméletből származtattuk, és induktív módon alakítottuk ki őket a kutatás során.

Az interszubjektivitás elismerésének másik megnyilvánulása a kettős annotációra való áttérés volt. Ahelyett, hogy azt feltételeztük volna, hogy minden egyes hozzászólás egy vagy két kategóriába tartozik, amelyeket egy megfelelően képzett



TEMATIKUS TANULMÁNYOK – Digitális szociológia

annotátor azonosítani tud, úgy közelítettük meg a fórumbejegyzéseket, mint amelyek többféleképpen értelmezhetők. Annak érdekében, hogy minimalizáljuk az értelmezés esetlegességét, két független annotátorral kódoltattunk minden posztot. A végső, konszenzusos címke a két annotátor kódjának egyesítésén alapult a fent leírt módon.

Tapasztalataink szépen visszatükrözik más kutatók benyomásait. Aroyo és Welty (2015) egyenesen a humán annotálás hét mítoszaként utal azokra a naiv/pozitivisták elképzelésekre, melyek egyértelműen kódolható szövegeket tételeznek fel. Szerintük a humán annotálás egy elavult szemantikai eszményen alapul, ami az egyetlen helyes igazság meglétét tételezi fel, ebből további mítoszokat levezetve, mint például az annotátorok közötti eltérés hátrányos voltát, azt a reményt, hogy a megfelelően részletezett annotálási irányelvek megoldják a problémát, annak hitét, hogy a szakértők besorolása mindig helyesebb, mint a laikusoké, vagy annak elvárását, hogy egyetlen kategóriába besorolható legyen a szöveg. Új szemantikai elméletet javasolnak, a „crowd truth”-ra alapozva, melynek lényege, hogy az emberi értelmezés szubjektív, és hogy az annotátorok különböző interpretációi jó reprezentációját adják ennek a szubjektívitásnak, s az ésszerű interpretációk tartományának.

A Mesterséges Intelligencia torzítás és az annotálás

A Mesterséges Intelligencia torzítás (MI-torzítás, angolul AI-bias, az Artificial Intelligence rövidítéseként) lényege, hogy a nyelvtechnológia outputja maga is társadalmi torzításokat (pl. kisebbségekkel, időssekkel vagy nőkkel szembeni hátrányos megkülönböztetést) mutat, amellyel mintegy felerősíti azok társadalmi hatását (Ntoutsis et al. 2020). Az egyik legismertebb példa az Amazon kísérleti rekrutációs algoritmusának esete (Dastin 2018). Az algoritmus létrehozásának célja az volt, hogy az 1-től 5-ig automatikusan osztályozza önéletrajzok alapján a jelentkezőket. 2015-re azonban kiderült, az algoritmus gender-alapon torzított a szoftverfejlesztő és más technikai jellegű pozíciók esetén. Ennek oka az volt, hogy a modellt a céghez a megelőző 10 évben benyújtott jelentkezések alapján tanították, de a technológiai iparág férfi-domináns, így a jelentkezések is férfiakról jöttek elsősorban. Emiatt az MI rosszabbul minősítette azokat a jelentkezőket, amelyekben a „női” szó szerepelt, legyen az akár csak a „női sakkcsapat kapitány” kifejezés. A fejlesztők megkísérelték oly módon javítani, hogy az ilyen fajta nemi információtól való függést kiiktatták a tanulóból, de ez nyilván nem feltétlenül jelenti a probléma megoldását, hiszen egy ilyen rendszer más, nem vizsgált dimenziók mentén is tartalmazhat torzítást. A fejlesztők végül felhagytak ennek az algoritmusnak a használatával.

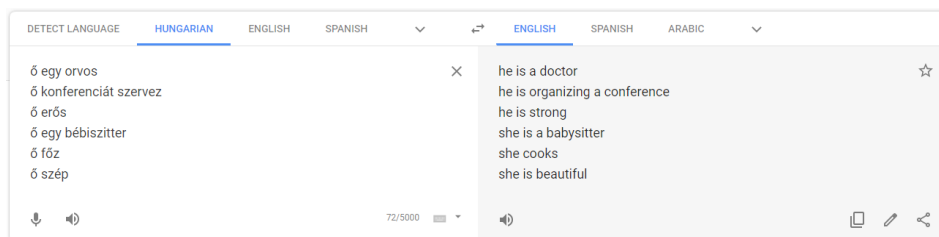
Az MI-torzítás általánosan megfogalmazva egyfajta anomália a gépi tanuló algoritmus outputjában; vagy az algoritmus fejlesztési folyamatában alkalmazott felte-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

vésekre vagy a tanuló-halmazban rejlő inherens társadalmi torzításokra vezethető vissza. Az annotálás kapcsán a továbbiakban az utóbbira koncentrálnak.

A tanuló halmaz, ember alkotta élő szöveg lévén mindazokat a viszonyokat tükrözi vissza, amik a társadalomban is megtalálhatók. Az Amazon példáján: a tanuló-halmaz kiegyensúlyozatlan volt nemi összetételét tekintve. Hasonló ismert példa a Google Translate esete is. Farkas Annával (Farkas – Németh 2022) végzett kutatásunkban azt vizsgáltuk, hogy ha magyarról angolra fordítunk foglalkozásneveket tartalmazó mondatokat („Ő egy orvos”), akkor a fordító hímnemű vagy nőnemű névmást használ, s hogy ez a döntés korrelál-e a foglalkozások tényleges nemi megoszlásával illetve a magyarok survey-jel mért attitűdjével (azaz azzal, hogy inkább férfias vagy nőies foglalkozásnak vélnék valamit a magyarok). Eredményünk szerint a fordító erősen torzít a nőkkel szemben, és működése közelebb áll az attitűdökhöz, mint a tényleges foglalkozásszerkezethez (3. ábra).



3. ábra. A Google Translate nemi torzítása
Forrás: Farkas – Németh 2022: 2

A fordító torzításának oka itt is az, hogy kétnyelvű szöveghalmazokon tanul, s ha a társadalmi attitűd vagy a tényleges foglalkozásszerkezet inkább férfiasnak mutat egy foglalkozást, akkor az a szövegekben is megmutatkozik majd, így a fordító is visszatükrözi ezt az egyenlőtlenséget. (A Google mai működésében már nem detekálható ez az anomália, a visszajelzések nyomán úgy módosították az algoritmust, hogy mindkét névmást megadja a fordítás.)

Az annotáción alapuló torzítás a tanuló-halmaz szintjén vihet anomáliát a rendszerbe, kétfajta úton. Vagy az annotátorok, maguk is társadalmi normákat képviselve, tükrözik vissza a címkézésben a társadalmi viszonyokat, vagy azzal okoznak torzítást, hogy nem reprezentálják kellőképpen a társadalom egészét. Mindkét probléma egyszerre volt jelen abban a gyűlöletbeszéd-kutatásban (Sap et al. 2019), ahol a torzítást az okozta, hogy a jobbára fehér bőrű annotátorok a címkézendő tweetek szerzői által használt afroamerikai dialektust hajlamosabbak voltak offenzívként megjelölni. E dialektus ugyanis több káromkodást tartalmaz, és még ha az nem is irányul más felhasználók ellen, akkor is sértőnek, offenzívnek érezték a más nyelvi normákat képviselő kódolók.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Eltávolítható-e a torzítás az algoritmusokból? Ahogy láttuk, az MI algoritmusok épp annyira lehetnek kiegyensúlyozottak bármilyen szempontból, mint amennyire a tanuló-halmaz az. „Reprezentatív” annotátor-csapat és „reprezentatív” tanuló minta kialakítására érdemes törekedni ugyan, de tökéletes reprezentativitás még elvileg is nehezen elképzelhető. Felmerül annak kérdése is, hogy valóban a nagy átlagot kell-e az algoritmusnak hoznia, nem lehet-e egy-egy kisebbségi csoport véleménye relevánsabb – gondoljunk itt akár egy tényellenőrzési feladatra, ahol a „tények” detektálása nem feltétlenül reprezentatív szavazással dönthető el.

Megkísérrelhetjük továbbá nem az input, hanem az output oldalon is a javítást: a torzítás eltávolítását egy-egy dimenzió mentén (ahogy az Amazon fejlesztői tették), de számtalan látens dimenzió létezhet, ezért tökéletesen torzítatlan MI algoritmus nehezen elképzelhető – ugyanakkor törekedni kell tesztelésükre és javításukra.

Összegzés

A hatalmas mennyiségű digitális szöveges adat elérhetősége és új elemzési potenciálja széles perspektívát jelent a szociológia számára. A számítógépes társadalomtudomány, jelen példákön a szövegbányászat várhatóan akkor lesz beépíthető a mindennapi kutatásba, ha az interdiszciplináris együttműködések széles körben elterjednek, ha a szükséges tudás és kompetencia beépül az egyetemi képzésbe. A fejlett programozási ismereteket nem igénylő, alacsony küszöbű szövegelemző platformok ugyanakkor már most megjelentek (mint például a Google Trends, Google Ngram Viewer, vagy az európai fejlesztésű Sketch Engine), támogatva az átalakulást – ezek további fejlődése valószínűsíthető a közeljövőben.

Láttuk, hogy a szövegeken alapuló gépi tanulás is új lehetőségeket kínál a szociológia számára. A gépi tanulási megoldások egyre nagyobb része humán annotáción alapul, és az ilyen ember-gép együttműködésen alapuló számítások még erősebb felfutása várható nem csak az iparban, de a tudományban is. Az előre annotált és nyilvánosan elérhető adatbázisok már most támogatják a felügyelt tanulás saját alkalmazását. Ugyanakkor, mivel a szociológiai fogalmak általában komplexebbek, e humán annotáció során nem kerülhető meg az „igazság” egyértelműségének elvetése és az emberi értelmezés szubjektivitásának elfogadása.

A szociológus ugyanakkor nem csak felhasználóként, hanem a kritikai nézőpont képviselőjeként, a felügyelt gépi tanulás társadalmi hatásaira, etikai problémáira rámutató aktorként is jelen kell, hogy legyen. A crowdsourcing annotálás komoly munkaerőpiaci problémákat generál, az MI-torzítás pedig társadalmi hátrányokat erősíthet fel. Ezen túl a cikk keretein túlmutató, de zárásként mindenképp megemlítenő kérdés az internetes privacy és szabadság problémája. Kérdés, milyen hatása lesz ezekre a nagytömegű szövegek gyors feldolgozására képes technológia, nem válik-e a szövegbányászati technológia a cenzúra és megfigyelés mindenható eszközévé?



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Irodalom

- Aroyo, L. – Welty, C. (2015): Truth is a lie: Crowd truth and the seven myths of human annotation. *AI Magazine* 36(1): 15-24. <https://doi.org/10.1609/aimag.v36i1.2564>
- Dastin, J. (2018): Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> (utolsó elérés: 2021. április 7.)
- Eisenstein, J. (2019): *Introduction to Natural Language Processing*, MIT Press, Cambridge, MA
- Farkas A. – Németh R. (2022): How to measure gender bias in machine translation: Real-world oriented machine translators, multiple reference points. *Social Sciences & Humanities Open*, 5(1): 100239.
- Fort, K. – Adda, G. – Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics* 37(2):413–420. doi: 10.1162/COLI_a_00057
- Gadamer, H. G. (2004): *Truth and Method*. 2. kiadás. Crossroad, New York
- Gray, M. L. – Suri, S. (2019): *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston
- Hovy, E. – Lavid J. (2010). Towards a 'science' of corpus annotation: a new methodological challenge for corpus linguistics. *International Journal of Translation*, 22(1): 13–36.
- Ignatow, G. – Mihalcea, R. F. (2017): *An Introduction to Text Mining: Research Design, Data Collection, and Analysis* (1st edition.). SAGE Publications, Inc., Los Angeles
- Jelveh, Z. – Kogut, B. – Naidu, S. (2014): Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1804–1809). Doha, Qatar: Association for Computational Linguistics. <http://www.aclweb.org/anthology/D14-1191>. (utolsó elérés: 2021. 11. 20.)
- Nahili, W. – Rezeg, K. – Kazar, O (2020): Big Data Analytics using Supervised Learning: A Comprehensive Review of Recent Techniques. *International Journal for Research in Applied Science and Engineering Technology*, 8(1): 305-312.
- Mühlhoff, R. (2019): Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society*, 22(10): 1868-1884. <https://doi.org/10.1177/1461444819885334>
- Németh R. – Koltai J. (2021): Sociological knowledge discovery through text analytics. In: Rudas T. – Péli G. (szerk.): *Pathways between Social Science and Computational Social Science – Theories, Methods and Interpretations*. Springer, New York



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Németh R. – Katona E. – Kmetty Z. (2020): Az automatizált szöveganalitika perspektívája a társadalomtudományokban. *Szociológiai Szemle*, 30(1):44-62.
- Németh R. (2021): *Az okság alternatív fogalmi és módszertani megközelítései a szociológiában*. Savaria University Press, Szombathely
- Németh R. – Sik D. – Máté F. (2020): Machine Learning of Concepts Hard Even for Humans: The Case of Online Depression Forums. *International Journal of Qualitative Methods* 19(1): 1-8. <https://doi.org/10.1177/1609406920949338>
- Ntoutsis, E. et al. (2020): Bias in data-driven artificial intelligence systems – An introductory survey. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery* 10(6): 1-14. <https://doi.org/10.1002/widm.1356>
- Poletto, F. – Stranisci, M. – Sanguinetti, M. – Patti, V. – Bosco, C. (2017): Hate Speech Annotation: Analysis of an Italian Twitter Corpus In: Basili, R. – Nissim, M. – Satta, G. (eds.): *Proceedings of the Fourth Italian Conference on Computational Linguistics CLiC-it 2017: 11-12 December 2017, Rome*. Accademia University Press, Torino <https://doi.org/10.4000/books.aaccademia.2448>.
- Ryan, L. – McKie, L. (eds.) (2015). *An end to the crisis of empirical sociology? Trends and challenges in social research*. Routledge, London
- Sap, M. – Card, D. – Gabriel, S. – Choi, Y. – Smith, N. A. (2019): The Risk of Racial Bias in Hate Speech Detection. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy.
- Savage, M. – Burrows, R. (2007): The Coming Crisis of Empirical Sociology. *Sociology: A Journal of the British Sociological Association*, 41(5):885–899. <https://doi.org/10.1177/0038038507080443>
- Snow, R. – O'Connor B. – Jurafsky D. – Ng A. Y. (2008): Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Honolulu, Hawaii
- Vapnik, V. N. (2000): *The Nature of Statistical Learning Theory*. 2. kiadás, Springer-Verlag, New York, Inc.



Magyar videóblogger hálózatok online

HEGEDŰS DÁNIEL¹

ABSZTRAKT

A web 2.0 és digitális környezet, illetve az ezekben a terekben létrejövő közösségi média vitathatatlanul átformálta életünket: nemcsak a mindennapi tapasztalatainkat befolyásolta, de egy teljesen új társas terepet is teremtett, elősegítve korábban soha nem látott társadalmi gyakorlatok és szereplők kialakulását, elterjedését. A digitális technológiák demotizáló (Turner 2010) hatása teljesen új terepeket hozott létre a hírességek (újra)termelődéséhez, aminek egyik legnépszerűbb formája a videóblogolás. A vlogerek különböző mikro-celeb gyakorlatokon keresztül meglehetősen nagy hírnévre tettek szert rövid időn belül (Marwick 2015, Senft 2012), így társadalmi szerepük kulcsfontosságúvá vált a különböző tudáshalmazok, ötletek, értékek terjesztése, nyilvános artikulációja kapcsán. Ezek a kognitív mintázatok különböző diszkurzív gyakorlatokon keresztül jutnak el a rajongókhoz, melyek konkrétan – a közösségi médiás affordanciáknak megfelelően – a posztok, videók, kommentek, tweet-ek stb. sokaságából állnak össze és amely aktivitási elemekkel a rajongók különböző interakciókat is végezhetnek (komment, like stb.). Ez a fajta diszkurzív folyamat a hírességeket (jelen esetben videóbloggereket) úgynevezett szakértői intézmény (Giddens 1990) pozícióba helyezi, aminek segítségével különböző attitűdbeli, mentális, viselkedési sémákat nyújthatnak rajongóiknak. Mindeközben a digitális környezet a társadalomtudományos megismerés számára is teljesen új perspektívákat nyit, mivel a kutatók éppen ezekhez a diszkurzív panelekhez férhetnek hozzá nagy volumenben, tudományos elemzés céljából. Jelen kutatás is éppen a digitális adatokban rejlő megismerési potenciált aknázza ki, ugyanis a SentiOne nevű social listening alkalmazás segítségével rekonstruálja a hírességek alkotta online hálózatokat a különböző közös említések alapján (törtéjnen az egy posztban, vagy rajongói kommentben). A hálózat felrajzolásán felül a SentiOne lehetőséget nyújt különböző, az említésekhez kapcsolódó metrikák létrehozására, amellyel még részletesebb kép nyerhető a hazai online térben hálózatosodó videóblogger hírességekről és az általuk létrehozott diskurzusról.

KULCSSZAVAK: digitális szociológia, szakértői intézmények, hálózatiság, diskurzuselemzés

ABSTRACT

Hungarian Videoblogger Networks Online

The web 2.0 phenomenon and social media – without question – not only reshaped our everyday experiences, but they have established an environment for new types of social practices and social actors. The demotization (Turner 2010) effect of such technologies has created entirely new fields where celebrities might emerge from: one of them is videoblogging. Many video bloggers

¹ ELTE Szociológia Doktori Iskola, Interdiszciplináris Társadalomkutatások Doktori Program, e-mail: shrilby@windowslive.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

gained great reputation through peculiar micro-celebrity practices (Marwick 2015, Senft 2012), and, as a result, became key figures in distributing ideas, values and knowledge in today's society. These cognitive patterns are disseminated with a discursive apparatus that is largely based on social media activity, including posts, tweets, self-imagery and the videos themselves, which are tied to a certain logic according to environmental affordances, creating the possibility for fans to interact (share, comment, like, retweet etc.) with artifacts of the celebrity. This mechanism puts the celebrity in a so-called expert system (Giddens 1990) position as they provide adequate schemas of attitude, mentality or behavior. Most importantly, all of these public interactions are accessible for scholars to conduct scientific research. With the help of the SentiOne application this research attempts to reconstruct online networks of video bloggers based on mentions, which either occurred in an artifact (post, video description etc.) or in a fan comment. Apart from the network itself, SentiOne enables us to get insights regarding each individual connection established in it with different types of aggregated data.

KEYWORDS: *digital sociology, expert systems, networks, discourse analysis*

Bevezetés

Tanulmányomban a hazai YouTube-szféra bizonyos struktúráinak feltárására teszek kísérletet, a digitális szociológia, a big data paradigma és egyéb hálózatkutató megközelítések segítségével. Ennek során a videóbloggerek közötti kapcsolatokat és az ezek alkotta hálózatokat igyekszem rekonstruálni, illetve társadalomtudományos szempontból releváns módon értelmezni, ami egyszerre mind a digitális szociológia eszköztárában rejlő lehetőségeket is jól illusztrálja. Témám kiindulópontja, hogy a hírességeket – így a nagy követői bázissal rendelkező videóbloggereket is – giddens-i értelemben vett szakértői intézményeknek tekintem, akik ebben a minőségükben online konstruált diskurzusaikkal értékeket, tudásokat, viselkedési és gondolkodási sémákat közvetítenek rajongóik felé. A vizsgálat ezeknek a szakértői szereplőknek és diskurzusaiknak a hálózatosodását képes megragadni, ahol a hálózati kapcsolat meglétének és erősségének az indikátorai, az ezekben a szakértői diskurzusokban történő említések lesznek. A celebek nevének is az egy kontextusban történő megjelenése a tudományos szerzői hálózatokhoz hasonlóan, bizonyos fajta kapcsolatra enged következtetni, mely kapcsolat említéseket aggregálva, hálózati ábrákban leképezhetővé válik. Mivel a hazai videóblogger celebszféra feltehetőleg nem elszigetelt szereplőkből áll, az általuk generált diskurzus is dinamikus ökoszisztémát alkot, melynek hálózati felrajzolása segítheti az ebben a miliőben zajló érték közvetítések mélyebb megértését.

A digitális tér nemcsak ezt a teljesen új generációját termelte ki a hírességeknek, nemcsak az itt folyó diszkurzív folyamatoknak adott helyet, hanem a társadalomtudományos megismerés számára is korábban elképzelhetetlen perspektívákat teremtett. Számtalan digitális eszköz áll a kutatók rendelkezésére, amellyel feltárhatják az online világban zajló társas folyamatokat. Kutatásomban is éppen egy ilyen eszköz-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

re, a SentiOne elnevezésű, *social listening* szoftverre támaszkodom. Egy említés alapú keresőmotorról van szó, amely számtalan online platformon képes kereséseket végezni, illetve a találatokról adatbázisokat és elemzéseket készíteni. A szoftver adta lehetőségeknek és a kutatás céljainak megfelelően kutatási kérdésem a következő: *milyen módon hálózatosodik a hazai top videóblogger miliő a diskurzusaikban megjelenő közös említések alapján?*

A tanulmányban először a mikro-celeb jelenséget és annak szociológiai értelmezéséhez szükséges szakirodalmat járom körbe (különös tekintettel a lehetséges szakértői státuszra), illetve ezen szereplők hálózatosodását vizsgálom, majd bemutatom az adatgyűjtéshez használt *social listening* szoftvert a SentiOne-t. Ezután részletezem a kutatás módszertani megfontolásait, kitérve mind az adatgyűjtésre, mind az elemzésre, végül bemutatom az eredményeket, az idevágó lehetséges értelmezésekkel, módszertani korlátokkal és lehetséges további irányokkal egyetemben.

Mikro-celebek, videóbloggerek, szakértőiség

A téma megfelelő szakirodalmi kontextualizálásához három altémakört szükséges áttekinteni: a digitális teret és annak legfontosabb affordanciáit,² a hírességeket, és egy speciális alcsoportjukat a mikro-celebeket, valamint ezen csoport lehetséges szakértői pozíciójának pontos mibenlétét.

Az internet és a digitális tér soha nem látott mértékben alakította át az egyének mindennapi tapasztalatait: a társas cselekvések számtalan aspektusa reprezentálódik valamilyen formában ezeken az online felületeken, sőt, ahogy átlépett a web 2.0-ás szakaszába, az internet egyre több és több teljesen új cselekvési formát is kitermelt.³ A számtalan ilyen új forma közül kerülnek ki a kutatás alanyai is: a videóbloggerek (vlogerek).

A digitális érának, illetve a web 2.0-ás internethasználati módnak az interaktivitáson és a felhasználói tartalomelőállításra felül még egy további tulajdonsága meghatározó: a gyakorlatilag végtelen tárolókapacitása. Ezeknek az affordanciáknak az összessége vezetett az úgynevezett long-tail jelenséghez (Anderson 2004). Ez alapvetően az elérhető tartalmak alapján felállított skálafüggetlen görbét jelent, ahol a görbe vége a végtelenbe nyúlik és ahol az x tengelyen az egyes tartalmak, míg az y tengelyen a megtermelt profit áll. Ez a piac számára elvileg azt jelenti, hogy a végtelen kis bevételt hozó tartalom profit szinten jóval nagyobb nyereséget termel ki a kis-

² Azon jellemzők és tulajdonságok összessége, amelyek kijelölik az adott objektummal végezhető cselekvések körét (melyek lehetnek az alkotó szerint szándékoltak vagy nem szándékoltak), illetve ezen cselekvések pontos paramétereit is.

³ Nyilvánvalóan az internethasználat már a sokkal passzívabb web 1.0 korában is létrehozott több, azelőtt nem létező interakciós közeget, viszont a felhasználói tartalomelőállítást és az ezekhez a tartalmakhoz való hozzáférést és minden korábrinál sokkalta interaktívabb böngészést jelentő web 2.0 ezt tovább fokozta.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

számú „slágertermékhez” képest⁴ (ami alapján Anderson egy teljesen új piaci logikát vizionál) (Anderson 2004). Bár a tárgyalt digitális közeg elvileg valóban lehetővé teszi a kvázi végtelen tartalom elérését, fontos látni, hogy a modell elsősorban a kultúraipar termékeire és szereplőire (így a videóbloggerekre is) igaz. Ezeknek az esetében azonban az offline világra korábban is jellemző long-neck modell marad jelentős a fogyasztói oldalon, azaz a tartalmak terén elérhető bőség mellett is valójában csak igen kevés szereplő fogja birtokolni a figyelem és profit legnagyobb részét, így a mindennapi tapasztalatok, az identitáskonstruálás és a szocializáció szempontjából mégiscsak ezek lesznek a meghatározóak (a long-tail így a kínálati oldal, míg a long-neck elsősorban a keresleti oldal jellemzője lesz) (Dessewffy – Láng 2015b).

Ez a vloggerek esetében konkrétan azt jelenti, hogy bár a megfelelő platformokon elvileg végtelen mennyiségű szereplő jelenhet meg, csak kisebbségük fog rendelkezni a rajongói figyelem, interakciók és az ezen alapuló profit túlnyomó hányadával, vagyis a közeg szükségszerűen hat a celebrifikáció (Turner) irányába. Ezt a folyamatot erősítik tovább a közeg affordanciái is. A web 2.0 éppen az interaktivitás és fokozódó tartalomelállítást (melyek volumene nyilván többszörösen meghaladja az egyéni befogadóképességet) következményeként egyfajta kurátorszerepbe kerül, amely révén tárolja, rendezi és újraelosztja a felhasználói tartalmakat (Hogan 2010), ezzel alapvetően befolyásolva a digitális térben való tájékozódást, navigációt. A digitális tér (amit mindezen jellegzetességei miatt már web 3.0-ként is aposztrofálnak) ebben az értelemben nem pusztán a vizsgált interakciók színtereként jelenik meg, hanem kvázi-cselekvőként, amely az emberi interakciók során saját maga is döntéseket hoz, alapjaiban befolyásolva az interakciók kimenetelét, vagy egyáltalán létrejöttének esélyét (Latour 1996, Sik 2013, Bucher 2012, Birkbak – Carlsen 2016).

Ezek a folyamatok együttesen egy teljesen új hírességformát is kitermeltek: az internet celebrityt (Abidin 2018), illetve ezen belül a tanulmány fókuszában álló videóbloggereket. Ezek a szereplők a tér affordanciáinak tekintetében jelentős mértékben különböznek a hagyományos média (*broadcast media*) hírességeitől,⁵ mivel kihasználhatják a közösségi platformok adta lehetőségeket, így az önpromóciót, a stratégiai intimitást, a közösségi médiás elérés extenzivitását,⁶ vagy az *én* márkává

⁴ Fontos kiemelni, hogy ez a modell az olyan információk, kulturális javak és szolgáltatások esetében lehet igaz, melyek függetlenek a hordozóktól és hozzáadott értékük nem, vagy csak kevésbé változik a digitális közvetítő közeg által (egy zenemű vagy egy film nyilvánvalóan lényegesen kevésbé függ a hordozójától, mint egy híres festmény).

⁵ A híresség mint társadalmilag széleskörben ismert személy, elsősorban kulturális közvetítőcsatornához (cultural intermediaries) kötött (Rojek 2001). Videóbloggerek esetében ez a digitális tér, amely nyilvánvalóan teljesen újszerű gyakorlatokat tesz lehetővé a celeb-rajongó relációban – bár a tradicionális hírességekre is jellemző irreciprok kapcsolat továbbra is fennállni látszik (Hegedűs 2020). További fontos különbség, hogy ezek a szereplők többnyire a professzionális világon kívül tevékenykednek (van Dijk 2009).

⁶ Bár eredeti tevékenységük a YouTube videómegosztóhoz kötődik, a celeb perszóna fenntartása érdekében minden fontosabb közösségi platformon aktívan reprezentálják magukat.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

alakítását (Senft 2012, Khamis – Ang – Welling 2016, Marwick 2015, Marwick – Boyd 2011). Napjainkban ezek a határok a „hagyományos” hírességek és mikro-celebek között elmosódtak: a fent említett gyakorlatokat csaknem minden ismert személy⁷ aktívan gyakorolja egyszerre több digitális felületen.

Az internet celebrityket (ezen belül is a mikro-celeb videóbloggereket) és a hagyományos médiában kitermelt hírességeket lehetséges társadalmi funkcióik is igen hasonlóvá teszik a szociológiai értelmezés számára. Az ilyen közismert szereplők bizonyos értelemben a társas tanulást (szocializációt, integrációt) elősegítő kulturális narratívákat közvetítenek (Marshall 2010, Cocker – Cronin 2017, De Backer et al. 2007), azaz véleményeket, attitűdöket, viselkedéseket, tudáshalmazokat jelenítenek meg sokak számára⁸. Mindezt a kulturális közvetítőcsatornáknak köszönhető szinoptikonos pozícióból tehetik meg⁹ (Matthiesen 1997). Ezeknek a jellegzetességeknek a fényében a hírességeket (így a nagy rajongói bázisra szert tevő videóbloggereket is) szociológiai értelemben szakértői intézményeknek (Giddens 1990) lehet tekinteni (Hegedűs 2020), melyek az egyént szocializálják, integrálják, fegyelmezik (Giddens 1990, Lash 1993, Foucault 1971).

Fontos kiemelni, hogy a fegyelmezés ebben a konstellációban semmiképp nem egy durva, vagy akár erőszakot is alkalmazó kényszerítőerőt jelent. A szakértői intézmények diskurzusaikkal mindazonáltal újratermelik és meghatározzák a különböző társadalmi konvenciókat.¹⁰ Ez a gyakorlatban azt jelenti, hogy a celebek, nyilvánosan prezentált tevékenységeik során, valójában különböző értékalakzatokat, vélekedéseket, viselkedéseket jelenítenek meg, amelyek iránymutató narratívává is válhatnak az adott szereplőt követő rajongók számára. Eszerint ezeknek a hírességeknek a szakértőisége az úgynevezett mindennapi (esetleges, nem szisztematikus, homályos, rendszerezetlen) tudásra terjed ki elsősorban (Foucault 1976), amely a híresség-rajongó viszony mentén termel újfajta „hatalmi” viszonyokat – a tudományos tudásban megfigyelhető dinamikákhoz hasonlatosan. Jelen tanulmány szempontjából a legfontosabb e gondolatmenetben, hogy a szakértői intézmények bizonyos hozzáférési pontokat jelölnek ki klienseik (itt: a rajongók) számára, ame-

⁷ A mikro-celebritás fogalmához fontos hozzátenni, hogy nem feltétlen bizonyos tulajdonságokkal rendelkező csoportok megjelölésére, hanem általánosan bárki számára hozzáférhető gyakorlategyüttes-re utal (Senft 2012, VanDijk 2009).

⁸ A (poszt-)modernitásban – de már a hagyományos társas közösségek felszámolásakor és az egyre nagyobb volumenű média hatására – az egyén eleve sokkal inkább a különböző intézményekre és ipárgakra van utalva énjének és identitásának megkonstruálása során.

⁹ Ez a kevesek sokak általi megfigyelését jelenti (egyfajta kifordított panoptikumként) (Matthiesen 1997), ahol a megfigyelő tekintet nem feltétlen bír olyan fegyelmező erővel, mint Foucault panoptikumjában (Foucault 1990) – éppen ellenkezőleg: a megfigyelt viselkedésformák, vagy megjelenített narratívák válhatnak a fegyelmezés (szocializáció, integráció) eszközévé.

¹⁰ Giddens szerint mindez, a modernitás előtti időkben még elsősorban szemtől szembeni helyzetekben ment végbe, ám később már kifejezetten a szakértőiségen keresztül szűrve történik a társas konvenciók diszkurzív újratermelése (Giddens 1990, Lash 1993).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

lyen keresztül azok hozzájuthatnak a megfelelő diszkurzív tartalmakhoz (Giddens 1990). Esetünkben ezek a hozzáférési pontok a hírességek online aktivitási elemei (posztok, kommentek) lesznek. A tanulmány lényegében az ezek alapján konstruált gráfokkal térképezi fel, hogy a szakértői intézménynek tekintett celebek között létezik-e valamiféle kapcsolat és ha igen, az milyen mértékű és milyen mintázatot mutat.

Szakértői hálózatok

A bemutatott szakértői intézmény alapvetően egy társadalomtudományos kutatói absztrakcióként fogható fel, amellyel a hírességek és rajongók interakcióit értelmezem. Jelen tanulmányomban éppen ezt – az alapvetően absztrahált – megközelítést szeretném empirikusan is konkretizálni, illetve bizonyos értelemben visszaigazolni. A szakértői intézményt (minthogy intézmény) feltehetőleg a benne navigáló cselekvők bizonyos mértékű hálózati összekapcsoltsága jellemezi, így valószínűleg a honi videóblogger szféra szereplői sem teljesen elszigeteltek, hanem saját diskurzusukkal egy nagyobb volumenű milió diszkurzív közegébe ágyazódnak be. Ezt a logikát követve vizsgálom meg, hogy a hírességek fentebb bemutatott rétege, milyen jellegű hálózatos szerkezetet mutat a hazai online térben, azaz milyen mértékben kapcsolódnak egymáshoz az egyes szereplők, mennyire vannak elszigetelve, vagy éppen mennyire alkotnak egy interakciókkal, hivatkozásokkal, együttműködésekkel, vitákkal gazdagon átszőtt dinamikus ökoszisztémát.

A hálózati kapcsolódásokat az online térben nyilvánosan elérhető digitális lábnyomokon (Dessewffy – Láng 2015a, Kitchin 2014) alapuló rekonstrukción keresztül mutatom ki, ahol a konkrét kapcsolatokat egymás említései jelentik. Az említés itt minden szövegszerűen megjelenő hivatkozás, utalás egy celeb online felületein egy másik celebre.¹¹ Az említések (a hálózat élei) ezen a ponton két részre oszlanak. Egyrészt létrejöhetnek a celebek által felrajzolt élek (vagyis, hogy adott híresség a saját posztjaiban, szövegeiben stb. mennyire utal más celebekre), illetve a rajongók által konstruált kapcsolatok (a rajongók egy adott celeb tartalmainál mennyire említenek más celebeket, mennyire hivatkoznak másra stb.). A végeredmény tehát egy olyan hálózatos szerkezet lenne, ahol a celebrity alkotja a hálózat pontjait és az egymásra való hivatkozások a hálózat éleit (az előző gondolat értelmében minden pont (*node*))

¹¹ A tudományos hálózatokkal szemben a kapcsolat itt így nem a közös szerzőiségen (Mali – Ferligoj – Kronegger 2017, Newman 2004), hanem az online térben megjelenő, spontán viselkedésen alapuló digitális lábnyomokon alapul. Az így kapott hálózat háttérében álló adatok nyilvánvalóan jóval masztosabbak egy tudományos hálózathoz képest, mindazonáltal minden egyes ilyen említés mögött feltehetően valamiféle valós kapcsolat, együttműködés feltételezhető (különösen a posztokban megjelenő említések esetében). A kapott hálózat így nem műfajok, vagy más előfeltevéseken alapuló kategorizáció eredményei, hanem spontán aktivitások, azaz valós viselkedések által hátrahagyott digitális lábnyomok eredményeképp jön létre.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

között potenciálisan két él (*edge*) – egy rajongói és egy celeb – is húzódnak). Ez a kettős „élezés” a következő értelemmel bírhat a későbbi elemzés szempontjából: ha markánsan csak a celebek által konstruált élekkel átszótt a háló, akkor egy egyértelmű, valós, szorosan kapcsolódó cselekvőkkel rendelkező „szakértői intézmény” kontúrja rajzolódik ki, ha pedig markánsan csak a rajongói említések kötik össze a celebeket az azt jelentheti, hogy bár a celebek az önnön aktivitásuk szerint elszigeteltek, a rajongóik tapasztalati horizontján mégiscsak többé-kevésbé egységes tömbként jelennek meg. A két példa természetesen csak a feltárni kívánt szerkezetben rejlőd értelmezési lehetőséget kívánja demonstrálni, a ténylegesen megvalósuló hálózati konstelláció a fentiek bármilyen kombinációjaként, vagy teljesen váratlan alakzatban is megjelenhet.

A SentiOne

A digitális tér affordanciáinak, illetve a kutatás fókuszában álló társadalmi aktoroknak a szociológiai értelmezése mellett elengedhetetlen magának az adatgyűjtéshez és az elemzéshez felhasznált eszköz működésének és affordanciáinak a részletes bemutatása, hiszen mindez alapvetően határozza majd meg a feltehető és megválaszolható kérdések halmazát.

A hálózathoz felrajzolásához szükséges említések begyűjtését a SentiOne nevű *social listening* szoftverrel végeztem el.¹² A program legalapvetőbb működési elve, hogy megadott kulcsszavakra keres rá az online térben és a találatokat a beállításoknak megfelelően listázza. A kulcsszavak esetében a ragozások „kikerülésére” is jól bevált technikák állnak rendelkezésre, illetve a megfelelő keresést a program széleskörű nyelvtudása is biztosítja (az európai nyelvek többségét ismeri a program). Fontos adalék, hogy kulcsszavak kizárására is lehetőségünk van, ilyen módon kikerülhetjük a számunkra irreleváns szóösszetételeket és kontextusokat. A találatok listázásakor, a találatok az eredeti kontextusok megjelölésével kerülnek bemutatásra (például egy Facebook komment az eredeti poszttal együtt), illetve fontos tudni, hogy a SentiOne a teljes online teret vizsgálja, így egyszerre gyűjt adatokat a Facebook-ról a, a Twitter-ről, Instagram-ról és más nem közösségi média jellegű webhelyekről (blogok stb.). Ez a kulcsszavas logika kitűnően alkalmas a fentiekben vázolt módon felállítani kívánt hálózat empirikus feltérképezésére, amellet, hogy számos további érdekes tulajdonsággal is rendelkezik.

A program algoritmusai alapján a SentiOne becslést tesz a kulcsszavas szöveget hagyó cselekvő nemére, illetve a szöveg affektív tartalmát is képes megbecsülni (egy pozitív-semleges-negatív skála mentén). Ezek a becslések természetesen nem százszázalékos pontosságúak, mindazonáltal a beépített algoritmusok alapján a prog-

¹² <https://sentione.com>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ram tanul. Ha a listázott találatainkon mi magunk tévesnek ítélünk meg egy becslést (például egy negatív ironizáló szöveget az alkalmazás rosszul ítelt meg) azt módosíthatjuk és megfelelő számú ilyen „javítás” után a program elsajátítja az általunk alkalmazott szempontrendszert. A kulcsszavas keresést követő találatokról a program számtalan, különböző típusú – ábrákkal szemléltetett – statisztikát készít. Ezek magukba foglalnak különböző időbeli eloszlásokat, geolokációt és más paraméterek szerinti eloszlásokat (nem, affektív típusok stb.). Minden ábra dinamikus, sokszorosítható (így az azonos ábrák közül egyet-egyet átállítva könnyen készíthetünk szemléletes összehasonlításokat) és könnyedén letölthető többféle fájlformátumba.

Természetesen a SentiOne bemutatására nem a „reklám” és nem is az adatok „fetiszálása” miatt van szükség, hanem egy fontos szemléltetés miatt. Mivel hálózatunkban az egyes éleket az említések adják ki, a programon keresztül az egyes kapcsolódások gazdag leírására lehetünk képesek – még azok aggregált formájában is (a statisztikák, eloszlások, ábrák által). Mindez jól szemlélteti a digitális eszközökben rejlő megismerési potenciált, mindazonáltal legalább ilyen fontos figyelmet fordítani a korlátaikra is. A digitális platformokat vezérlő algoritmusokhoz hasonlóan az szoftveres analitikai eszközök is gyakran feketedoboz-szerűek (Bucher 2012), ugyanis kutatóként sohasem lehetünk teljesen biztosak abban, hogy egy-egy lekérdezés során pontosan milyen hozzáféréseket adnak, illetve tagadnak meg a szoftvertől az egyes platform tulajdonosai. Továbbá éppen a hatalmas adatmennyiség miatt meg lehetőségen körülményessé válik az adatbázis helyességének „kézi” visszaellenőrzése – ez gyakorlatilag csak szűrőpróbaszerűen tud megtörténni.

Hálózati módszertan

Tanulmányom alapvetően az online aktivitások begyűjtésével, rendezésével kívánja a hazai top videobloggerek hálózatosodásának kérdését vizsgálni. Ennek érdekében a Big Data paradigma eszköztárához nyúl – ami esetünkben elméletileg garantálja a teljes hazai online tér itt releváns aktivitásainak¹³ lefedését. Természetesen a téma által kijelölt társadalmi szféra még így is hatalmas ezért különböző megfontolások mentén szűkítéseket kell alkalmazni. A következőkben ezeket a módszertani megfontolásokat részletezem.

Mivel a hálózatot a celeb nevének említései alapján rajzoló fel, ezért a SentiOne keresőmotorját használva értelemszerűen a celeb névre történt keresés. A kétféle hálózatra (híresség és rajongói) felrajzolásához ez kétféle megközelítéssel történt. Elsőként a hírességek hálózatának módszertani hátterét mutatom be.

¹³ A továbbiakban az aktivitást a posztban vagy kommentben előforduló említésekre értem.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A videóbloggerek egy online ranglista alapján lettek összegyűjtve, amely a top 100 magyar YouTube csatornát rangsorolja.¹⁴ Ebből a listából történt a kutatás számára releváns szereplők, összesen hatvanöt videóblogger csatornájának kiválasztása¹⁵ (1. ábra).

1 Videómánia	33 Magyarósi Csaba
2 PamKutya	34 Nasz Dániel
3 Radics Peti	35 Sajt32
4 TheVR	36 pingvinharcos
5 luckeY	37 Paul Street
6 UNFIELD	38 Inez Hilda Papp
7 Barni.	39 Maris
8 JustVidman	40 I Have No Idea Hungary
9 Nessaj	41 FollowAnna
10 Cresser	42 Gergo Janosik
11 Dezső Bence	43 Isti Plays
12 BENIPOWA	44 royalszabi
13 zsDay	45 Otherworld
14 UborCraft	46 Vicc Elek
15 HollywoodNewsAgency	47 Inspirációk Csorba Anitától
16 zsozeatya	48 Már megint Kitty
17 Csecse Attila	49 Baluka
18 BTURBO	50 Videójáték Zsoltival
19 OwnMcKendry	51 The Nosika Show
20 Viszkok Fruzsi	52 Chabinho
21 Polla Channel	53 DareDewil
22 Anett - Ancsa	54 Herby
23 Isti Szalay	55 Zsomac
24 GoodLike	56 iDark
25 Gery	57 Varazslo
26 James	58 Tanulom Magam
27 DoggyAndi - Gameplay	59 Csizmadia Gabi
28 Benedek Nagy	60 Henry Kettner
29 IceBlueBird	61 Zozo Kempf
30 KD csapat	62 SZBYST
31 AlexGaming	63 TheBigO
32 CsChannel	64 Pinghorizon
	65 Dazis

1. ábra. Hazai top 65 videóblogger csatorna, 2021

Forrás: Saját szerkesztés

¹⁴ <https://starnetwork.hu/youtube-statisztikak/youtube-top100-hu/>

A lista kapcsán érdemes ismét hangsúlyozni az algoritmusoknak és más affordanciáknak való kutatói „kiszolgáltatottságot”: a lista feltehetően megtekintés, feliratkozósám és engagement alapon lett kalkulálva, bár erről semmilyen dokumentáció nem elérhető.

¹⁵ A kutatás a lista 2019 szeptember 30-i állapotát veszi alapul. Fontos megjegyezni, hogy a hatvanöt csatornába bekerült néhány csatorna, amelyek valójában egy személy „holding”-jába tartoznak (BENIPOWA és Benedek Nagy, Isti Szalay és Isti Plays, Polla és Anett és Ancsa).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A SentiOne-ban említéseiket az adott vlogger nevével gyűjtöttem be egy 2017. 09. 30-tól 2019. 09. 30-ig tartó, azaz egy kétéves időintervallumban, a következő platformokról: Facebook, Instagram, YouTube és Twitter. Az összeállított adatbázisokból ezután kiszűrtem a celebekhez tartozó aktivitásokat (posztok, kommentek, videók), majd az így kapott tisztított táblákon belül kerestem minden egyes celeb esetében a többi hatvannégy videóblogger nevére. Ezeket a találatokat regisztrálva alakult ki az a tábla, amely a későbbiekben a Gephi-vel felrajzolt hálózat alapját szolgáltatta.¹⁶

A rajongók esetében eredetileg egy adott celeb aktivitása alatt, rajongói kommentben megjelenő más celebek előfordulásait akartam regisztrálni, ám ezt mind a technikai nehézség (a kutatáshoz használt eszköz affordanciái nem tették lehetővé), mind az adatok számossága miatt elvettem. A rajongói tapasztalati horizontot jelentő hálózat alapja végül az egy rajongói kommentben egyszerre szereplő celebek említései lettek.¹⁷ Az adatgyűjtés ezen fázisát kifejezetten a SentiOne keretei között végeztem¹⁸, ahol az eredményekre ilyen módon különféle metrikákat lehet lekérni, további információkat nyerve a celebeket követő rajongói horizontról. Ezen a ponton fontos megjegyezni, hogy a rajongók esetében választott módszer mindössze csak egy a rajongói tapasztalati horizont lehetséges indikátorai közül, mindazonáltal technikailag csak ez bizonyult kivitelezhetőnek.

Eredmények

A gráfok a következők mentén lett felrajzolva: az egyes celebekhez tartozó pontok a bejövő és kimenő kapcsolatok fokszámainak megfelelően nőnek, vagy lesznek kisebbek (egyenesen arányosan). Az élek mérete pedig a hivatkozás mennyiségét jelöli: sok említés vastagabb élet eredményez. Már az ábrák értelmezése előtt kitűnik, hogy a vizsgált hálózatnak nincsenek elszigetelt elemei, minden celebhez tartozik legalább egy él – amennyiben minden típusú aktivitásukat figyelembe vesszük (2. ábra). A rajongói, illetve a celebek aktivitásaiból leképzett gráfokhoz tartozó általános statisztikákat a 3. és 4. ábra részletezi.

¹⁶ Ezek tartalmazták, hogy az adott celeb mely más celebekre hivatkozik (így a kapott gráfok irányítottak lesznek) és azt is, hogy mindezt hányszor teszi meg a vizsgált időintervallumban. Ez utóbbi adat a hálózat éleinek súlyát adta meg a későbbiekben (a gráfokon az él vastagsága jelöli ezt a paramétert).

¹⁷ A keresésben ugyanúgy az adott celeb nevére kerestem, úgy hogy kizártam azokat a találatokat, ahol éppen az adott híresség a tartalom szerzője.

¹⁸ A celebek együttes említéseit excell-ben regisztráltam, ami a celebek hálózatához hasonlóan a Gephi-s network alapja lett.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

	A	B	C	D	E	F	G	H	I	J
			Post Mention	Comment	Channel/Fanpage connection	Common Fanbase	Sum Fan Mention	Aktivitás	Sum Kimenő említések	Sum Beérkező (Celeb) Említések
1	SUM		477	3359	0	260	31339			0
2	Videómánia						2719	1	1	24
3		PamKutya	1				2			
4		Videómánia					2			
5		Radics Peti		2						
6		TheVR					1			
7		UNFIELD					1			
8		Barni.		3						
9		NessaJ		1						
10		BENIPOWA / Benedek Nagy		1						
11		zsDav		1						
12		HollywoodNewsAgency		1						
13		Cecse Attila		1						
14		Viztkok Fruzi	2	4						
15		Gery		1			1			
16		CsChannel		1						
17		Magyarósi Csaba		3						
18		Paul Street		1						
19		Henry Kettner		2						
20		Videómánia		2						
21		TheVR	2							
22		luckeY		1						
23		Barni.		1						
24		NessaJ		2						
25		OwMkandry		1						
26		GoodLike		2						
27		Nasz Dániel		1						
28		I Have No Idea Hungary		1						
29		ELZYST		1						
30		PamKutya					1			
31		luckeY		1						
32		NessaJ	4	1			1			
33		Dresó Bence		1						
34		inocentya					2			
35		Maris								
36		Otherworld		1						
37		Videójáték Zsoltival					1			
38		Videómánia		2						
39		Radics Peti		1						
40		TheVR		1						
41		UNFIELD		3						
42		Barni.		8						
43		NessaJ		2			1			
44		Cresser		5			1			
45		Dresó Bence		4						
46		BENIPOWA	1	1			2			
47		zsDav		1						
48		LiborCraft		1			2			
49		Cecse Attila		1						
50		OwMkandry		6			1			
51		Ametti - Alcsa					1			
52		GoodLike		3						
53		Gery		1			3			
54		James	1							
55		DoggyAndi - Gameplay		3						
56		Benedek Nagy		1			1			
57		IceBlueBird		2						
58		KD csapat		1			1			
		luckeY					904	88	36	24

2. ábra. Celeb aktivitások összesítése, 2021
Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Network Overview		
Average Degree	39.508	Run
Avg. Weighted Degree	40.079	Run
Network Diameter	3	Run
Graph Density	0.637	Run
HITS		Run
Modularity	0.061	Run
PageRank		Run
Connected Components		Run
Node Overview		
Avg. Clustering Coefficient		Run
Eigenvector Centrality		Run
Edge Overview		
Avg. Path Length	1.363	Run

3. ábra. Rajongói gráf statisztikák, 2021

Forrás: Saját szerkesztés

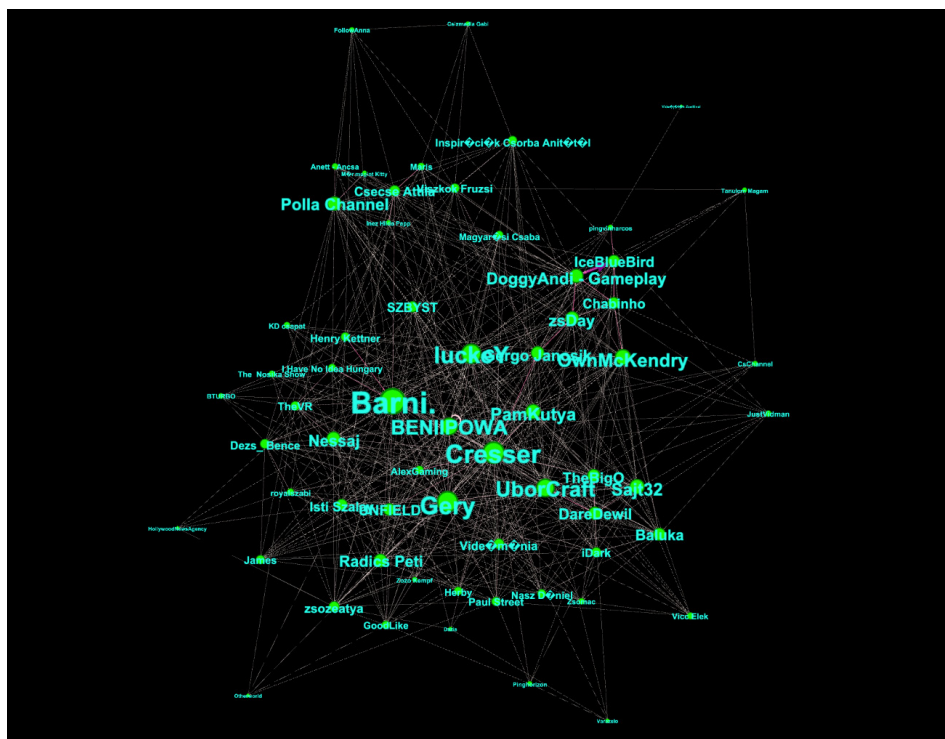
Network Overview		
Average Degree	11.476	Run
Avg. Weighted Degree	11.476	Run
Network Diameter	5	Run
Graph Density	0.185	Run
HITS		Run
Modularity	0.144	Run
PageRank		Run
Connected Components		Run
Node Overview		
Avg. Clustering Coefficient		Run
Eigenvector Centrality		Run
Edge Overview		
Avg. Path Length	2.108	Run

4. ábra. Celeb gráf statisztikák, 2021

Forrás: Saját szerkesztés

TEMATIKUS TANULMÁNYOK – Digitális szociológia

A celebek aktivitásain alapulól táblákból összesen három gráf készült, egy, amely kizárólag a celebek által írt kommentek alapján rajzolható fel (5. ábra), egy a posztokban előforduló említésekből (6. ábra), illetve egy általános összesítő hálózat, amely tartalmazza a celebek összes vizsgált aktivitását, mind posztokat, mind kommenteket¹⁹ (7. ábra).



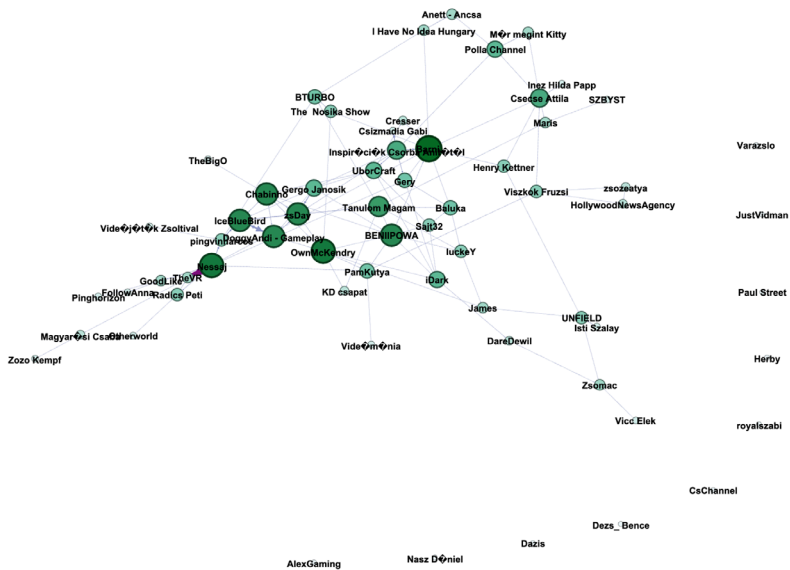
5. ábra. Celeb kommentek gráfja, 2021

Forrás: Saját szerkesztés

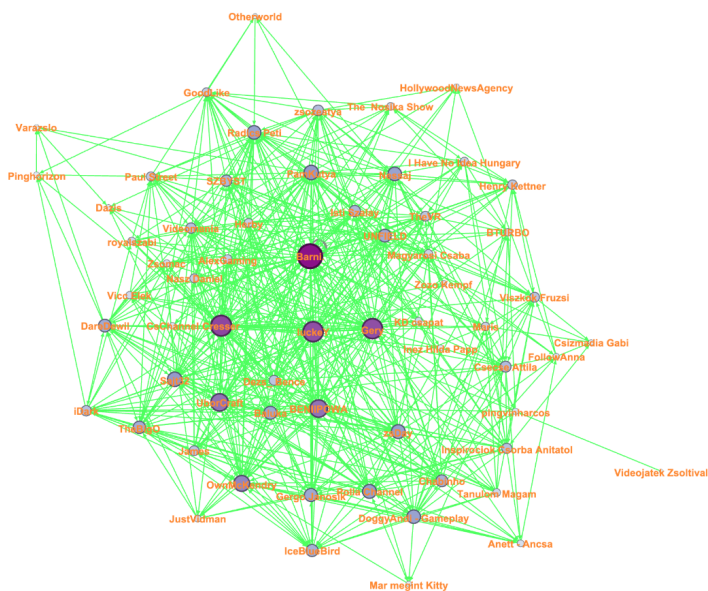
¹⁹ Ezeknek a gráfoknak (5, 6. és 7. ábra) a formázása során a ForceAtlas, Expansion és Noverlap parancsokat használtam (ez utóbbi kettőt csupán az áttekinthetőség és szemléletesség érdekében). A ForceAtlas parancs a skálafüggetlen eloszlások hálózati ábrázolásához ideális (<https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>), mivel a hírességek miliője az aktivitások alapján is éppen ezt a logikát követi.



TEMATIKUS TANULMÁNYOK – Digitális szociológia



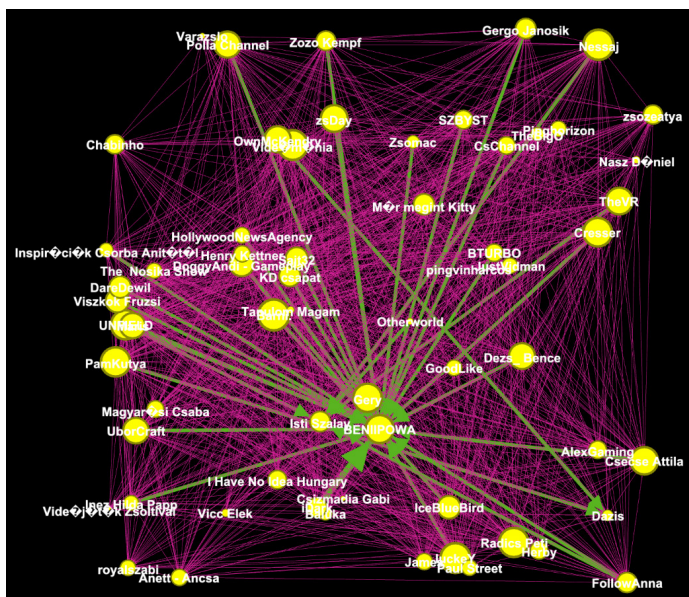
6. ábra. Celeb poszt említések gráf, 2021
Forrás: Saját szerkesztés



7. ábra. Posztok és kommentek összesített gráf, 2021
Forrás: Saját szerkesztés

TEMATIKUS TANULMÁNYOK – Digitális szociológia

Bár összességében minden vloggerhez tartozik legalább egy él, jól látszik, hogy csak a posztokat vizsgálva a hálózat jóval töredezettebb képet mutat: tíz szereplő kétéves posztolási aktivitása önmagában nem utal semmiféle együttműködésre, vagy kapcsolatra a hálózat többi részével. Ezekre a szereplőkre tehát csak kommentben hivatkozik a többi celeb (ami talán egy lazább kapcsolat meglétét jelentheti²⁰) (6. ábra). A vizsgált millió alapvető hálózatos szerkezetét egyfajta „tömbösség” jellemzi, vagyis láthatólag (force atlas paranccsal) nem alakulnak ki jól azonosítható kiscsoportok, azaz a különböző együttműködések, említések nem a műfaji, tematikai, vagy egyéb határok mentén képződnek. Emellett a hálózat igen sűrűnek mutatkozik: a pontok átlagos távolsága 1,36 lépés a rajongói kommentek esetében és 2,1 a hírességek esetében (3. és 4. ábra). Mindazonáltal az említésalapú kapcsolatok mennyisége mégiscsak kirajzol egyfajta centrum-periféria mintázatot (9. és 10. ábra). Ezt a mintázatot lényegileg még akkor is fenntartja a hálózat, amikor az azt leképező gráfokat klaszterekre próbáljuk szétszedni²¹ (11. és 12. ábra), vagy a rajongói tapasztalati horizont irányából, tehát a rajongói kommentekben előforduló közös említések felől vizsgáljuk (8. ábra).



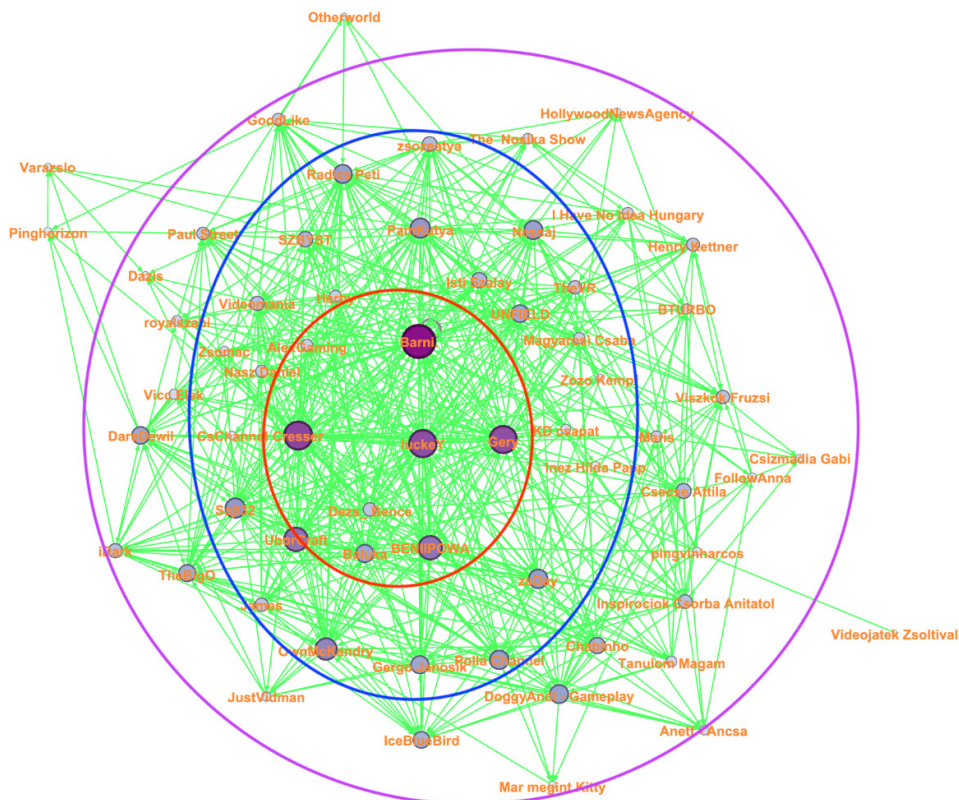
8. ábra. Rajongói kommentek gráfja, 2021

Forrás: Saját szerkesztés

²⁰ A valódi együttműködések, közös megjelenéseket többnyire posztokban és nem kommentek formájában artikulálják a vizsgált szereplők.

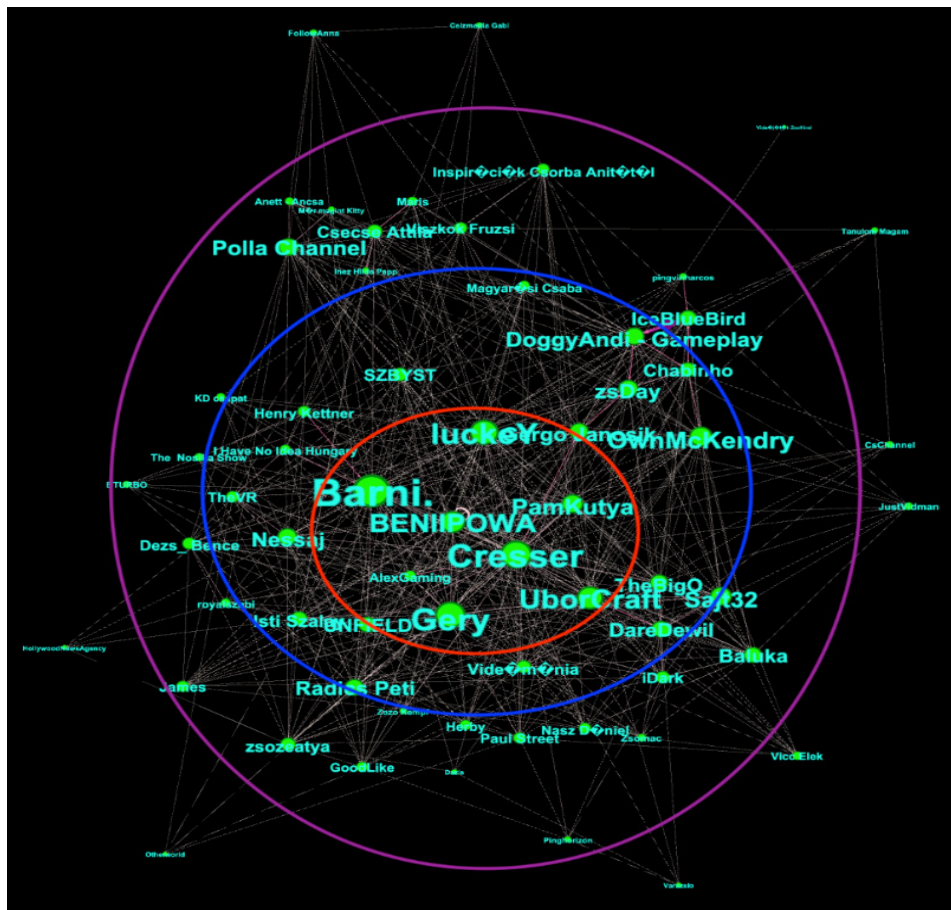
²¹ Ezt a célt a Gephi OpenOrd paranca szolgálja (<https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>).

TEMATIKUS TANULMÁNYOK – Digitális szociológia



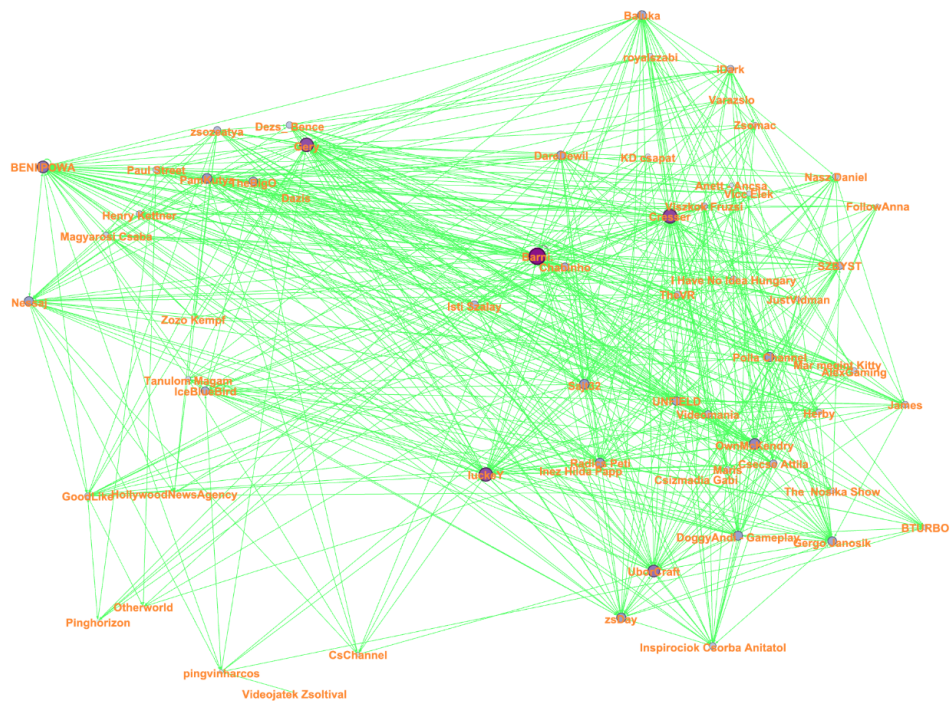
9. ábra. Centrum-periféria a celeb posztok esetében, 2021
Forrás: Saját szerkesztés

TEMATIKUS TANULMÁNYOK – Digitális szociológia



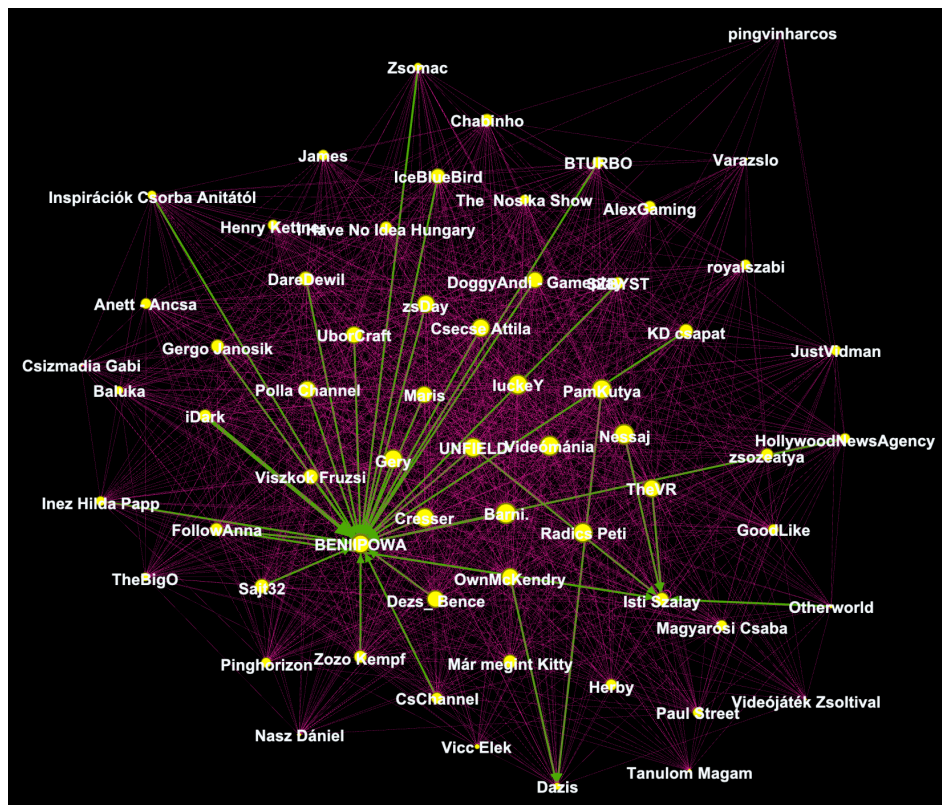
10. ábra. Centrum-periféria a celeb kommentek esetében
Forrás: Saját szerkesztés

TEMATIKUS TANULMÁNYOK – Digitális szociológia



11. ábra. Celeb aktivitás „klaszterek”, 2021
Forrás: Saját szerkesztés

TEMATIKUS TANULMÁNYOK – Digitális szociológia



12. ábra. Rajongói kommentek „klaszter”, 2021
Forrás: Saját szerkesztés

Fontos kiemelni, hogy a rekonstruált minta, vagyis, hogy ki foglal helyet a gráf „centrumában” és ki a „perifériáján” csak minimális összefüggést mutat a hírességrangsorban elfoglalt hellyel: a legtöbb kapcsolattal rendelkező szereplők túlnyomó többségükben a videóbloggerek legnépszerűbb egyharmadából kerülnek ki (15. ábra). Ehhez hasonlóan a hírességekre kalkulált aktivitás (posztok és kommentek összege) sem követi a csatorna rangsorát (megtekintések, feliratkozások és engagement aggregátuma) (13. ábra és 17. ábra). Ez azt jelenti, hogy a top négy csatornát bizonyos fokú „elszigeteltség” jellemzi, közvetlen alattuk azonban igen gazdag a hivatkozás és együttműködési háló, aminek sűrűsége a rangsoron egyre lejjebb haladva – néhány kiugrással,²² de – csökken. A rangsorhoz ebből a szem-

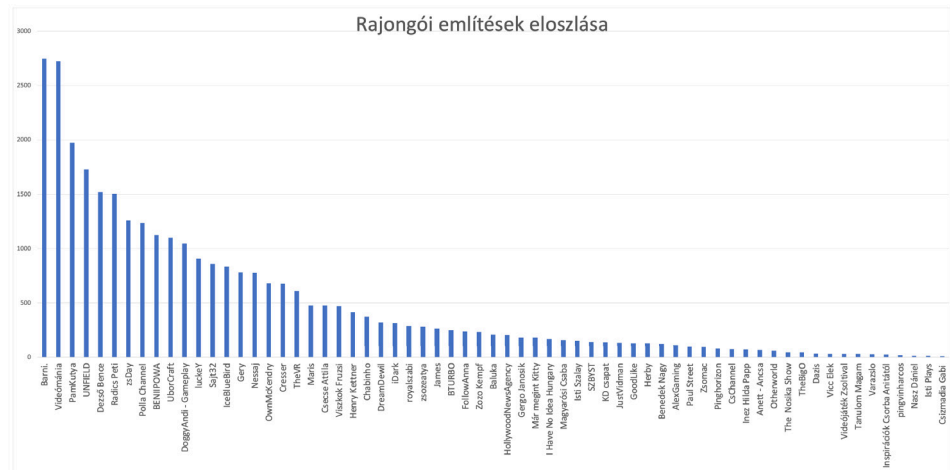
²² A kiugrások feltehetően barátságokkal magyarázhatóak, ahogy a baráti kapcsolat mentén szerveződő együttműködések egy rangsorban lentebb helyet foglaló szereplő mellé is nagy aktivitást generálnak.

TEMATIKUS TANULMÁNYOK – Digitális szociológia

pontból sokkal közelebb áll a rajongói horizont alapján felrajzolt eloszlás (14. ábra és 16. ábra).



13. ábra. Hírességek aktivitási eloszlása, 2021
Forrás: Saját szerkesztés



14. ábra. Rajongói említések eloszlása, 2021
Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

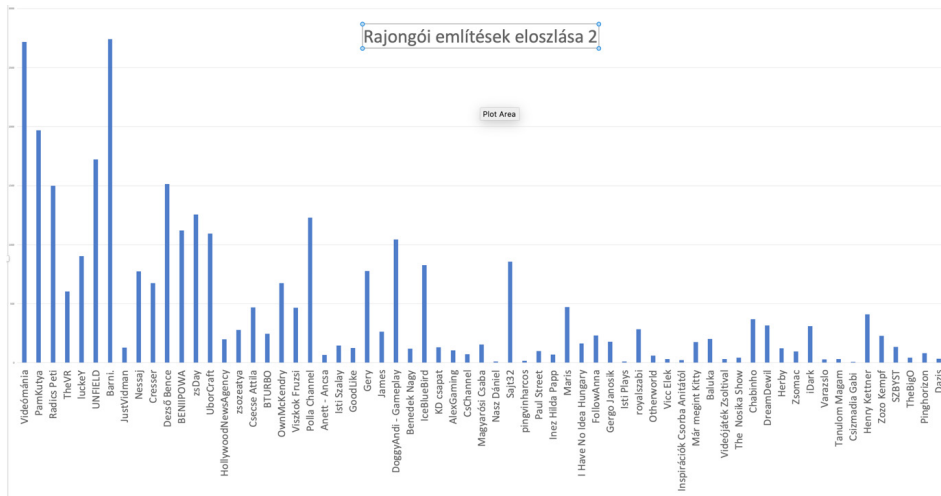
1 Videómánia	33 Magyarósi Csaba
2 PamKutya	34 Nasz Dániel
3 Radics Peti	35 Sajt32
4 TheVR	36 pingvinharcos
5 luckeY	37 Paul Street
6 UNFIELD	38 Inez Hilda Papp
7 Barni.	39 Maris
8 JustVidman	40 I Have No Idea Hungary
9 Nessaj	41 FollowAnna
10 Cresser	42 Gergo Janosik
11 Dezső Bence	43 Isti Plays
12 BENIPOWA	44 royalszabi
13 zsDay	45 Otherworld
14 UborCraft	46 Vicc Elek
15 HollywoodNewsAgency	47 Inspirációk Csorba Anitától
16 zsozeatya	48 Már megint Kitty
17 Csecse Attila	49 Baluka
18 BTURBO	50 Videójáték Zsoltival
19 OwnMcKendry	51 The Nosika Show
20 Viszkok Fruzsi	52 Chabinho
21 Polla Channel	53 DareDewil
22 Anett - Ancsa	54 Herby
23 Isti Szalay	55 Zsomac
24 GoodLike	56 iDark
25 Gery	57 Varazslo
26 James	58 Tanulom Magam
27 DoggyAndi - Gameplay	59 Csizmadia Gabi
28 Benedek Nagy	60 Henry Kettner
29 IceBlueBird	61 Zozo Kempf
30 KD csapat	62 SZBYST
31 AlexGaming	63 TheBigO
32 CsChannel	64 Pinghorizon
	65 Dazis

15. ábra. Videóbloggerek rangsor és aktivitás, 2021

Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia



16. ábra. Rajongói említések eloszlása, 2021

Forrás: Saját szerkesztés



17. ábra. Hírességek aktivitási eloszlása, 2021

Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Összegzés

A hírességek említés alapon felrajzolni kívánt hálózataról (amely az együttműködések, így az összekapcsoltság mértékének indikátora) összességében megállapítható, hogy igen sűrűn átszótt (ezt igazolja többek között a kevés lépésből álló átlagos távolság és a markáns klaszterek hiánya is). Az említések mennyisége, vagyis a hírességek aktivitása a – megtekintések, feliratkozások és engagement alapon – legnépszerűbb videóblogger esetében relatíve alacsony, a felső középmezőnyben megnő, majd néhány kiugrással ugyan, de viszonylag egyenletesen csökken (17. ábra). Mindez feltehetőleg azt jelenti, hogy a szakértőiség (így a szocializáló, integráló, fegyelmező diskurzus) nem individuális, elszigetelt szereplőkön keresztül jelenik meg, hanem egy sűrűn átszótt és átjárható ökoszisztémában, miliőben jön létre. A vlogger hírességek így feltehetően együtt alkotnak egy szakértői intézményt (annak ellenére is, hogy a csatornák témái igen nagy változatosságot mutatnak, még a rajongói tapasztalati horizonton is feltehetőleg egy tömböt képeznek²³). A rajongói közös említések (a rajongói tapasztalati horizont) is teljes mértékben ezt a logikát követi: bár eloszlását tekintve sokkal jobban leképezi a csatornák közötti rangsort, átjárhatóságát tekintve még a hírességek aktivitási hálózatánál is összekapcsoltabb képet mutat (minden ponthoz tartozik él, majdnem feleakkora az átlagos lépéstávolság a pontok között). A jövőbeni diszkurzív kutatások szempontjából mindez azt is jelentheti, hogy a hazai videóblogger szférában konstruált diskurzusokat egészében és nem individuális szereplőkként érdemes vizsgálni, hiszen mind a rajongók mentális térképén, mind maguknak a hírességeknek a mindennapi spontán gyakorlataiban is így jelennek meg.

Ezen a ponton fontos utalni arra, hogy a kutatás és eredményei, bár a Big Data bizonyos jegyeit magukon hordozzák, mégiscsak időben, földrajzilag és egyéb faktorok mentén korlátozott mintavételen alapulnak (Boyd – Crawford 2012, Kitchin 2014). Mindez óvatosságra int a tanulmány eredményeinek reflektálatlan túláltalánosítása kapcsán. Már csak azért is, mert a hálózatban szereplő említés alapú kapcsolatok az említések pusztá létezése alapján lettek felrajzolva, és az ebből kirajzolódó mintázatok alapján következtettek együttműködésre, közelségre, emberi kapcsolatokra és a szakértőiség hálózati leképződésére. Nem szabad továbbá megfelekedni a digitális platformok és social listening eszközök feketedoboz jellegéről sem: még a leggondosabb mintavételezés és módszertani design mellett sem lehetünk teljesen biztosak ezek működésének mikéntjében, tipikus hibáiban, esetleges torzításaiban, így az ezeken keresztül megalkotott és értelmezett adatbázisok érvényességét is bizonyos óvatossággal kell kezelni. Mindazonáltal jelen tanulmány is jó illusztrációja lehet annak, hogy a digitális terep, az ott megjelenő társas gyakorlatok, illetve az ezek felde-

²³ A celeb aktivitások hálózatából is igen kevés szereplő esik ki – ami a tanulmány esetében akár a mintavétel időbeli paramétereinek következménye is lehet.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

rítését szolgáló piaci és tudományos eszközök – kellő reflexió mellett – milyen merőben új kérdések feltevését és megválaszolását teszik lehetővé a szociológia számára.

Irodalom

- Abidin, C. (2018): *Internet Celebrity. Understanding Fame Online*. Bingley, Emerald Publishing Limited
- Anderson, C. (2004): *The Long Tail*. <https://www.wired.com/2004/10/tail/> (utolsó letöltés: 2020. 05. 14.)
- Birkbak, A. – Carlsen, H. (2016): *The Public and its Algorithms: Comparing and Experimenting with Calculated Publics*. In: Amooore, L. – Piotukh, V. (eds.): *Algorithmic Life. Calculative Devices in the Age of Big Data*. Routledge, Oxford, 21–34.
- Boyd, D. – Crawford, K. (2012): *Critical Questions for Big Data. Provocations for a Technological, and Scholarly Phenomenon*. *Information, Communication and Society* 15(5) 662–679. DOI: 10.1080/1369118X.2012.678878
- Bucher, T. (2012): *Want to be on the Top? Algorithmic power and the Threat of Invisibility on Facebook*. *New Media & Society*, 14(7): 1164–1180. DOI: 10.1177/1461444812440159
- De Backer, C. – Nelissen, M. – Vyncke, P. – Braeckman, J. – Mcandrew, F. (2007): *Celebrities: from teachers to friends*. *Human Nature*, 18(4):334–354.
- Dessewffy T. – Láng L. (2015a): *Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon*. *Replika* 2015/3-4. szám, 92–93.
- Dessewffy T. – Láng L. (2015b): *Digitális Pareto – az online közösségeken belüli megoszlás mintázatairól*. *Információs Társadalom*, 15(3): 7–17.
- Foucault, M. (1971): *Orders of discourse*. *Social Science Information*, 10(2): 7–30.
- Foucault, M. (1976): *Society Must be Defended. Lectures at the College de France, 1975–1976*. New York, Picador
- Foucault, M. (1990 [1975]): *Felügyelet és büntetés. A börtön története*. Budapest, Gondolat
- Giddens, A. (1990): *The Consequences of Modernity*. Cambridge, Polity Press
- Hegedűs D. (2020): *Magyar vlogger-diskurzusok a Facebook-on*. Jel-Kép Kommunikáció, közvélemény, média, 2020/1 1-19. DOI: 10.20520/JEL-KEP.2020.1.1
- Hogan, B. (2010): *The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online*. *Bulletin of Science, Technology & Society*, 20 (10) 1–10. DOI: 10.1177/0270467610385893
- Khamis, S. – Ang, L. – Welling, R. (2016): *Self-branding, ‘micro-celebrity’ and the Rise of Social Media Influencers*. *Celebrity Studies*, 8(2): 191–208. DOI: 10.1080/19392397.2016.1218292



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Kitchin, R. (2014): *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*. London, SAGE Publications
- Lash, S. (1993): *Reflexive Modernisation: The Aesthetic Dimension*. Theory, Culture, Society, 10(1): 1–23.
- Latour, B. (1996): *On Actor-network Theory. A Few Clarifications Plus More Than a Few Complications*. Soziale Welt, vol. 47: 369–381.
- Mali, F. – Ferligoj, A. – Kronegger L. (2017): *Scientific Co-Authorship Networks*. DOI: 10.1007/978-3-642-23068-4_6
- Marshall, P. D. (2010): *The Promotion and Presentation of the Self: Celebrity as a Marker of Presentational Media*. Celebrity Studies Vol. 1, No. 1, March 2010, 35–48. DOI: 10.1080/19392390903519057
- Marwick, E. A. (2015): *You May Know Me From YouTube: (Micro)-Celebrity in Social Media*. In: Marshall, P. D. – Redmond, S. (eds.): *A Companion to Celebrity*. Hoboken, John Wiley & Sons Inc., 333–350.
- Marwick, A. – Boyd, d. (2011): *To See and Be Seen: Celebrity Practice on Twitter*. *Convergence* 17(2): 139–58.
- Mathiesen, T. (1997): *The Viewer Society: Michel Foucault's 'Panopticon' Revisited*. Theoretical criminology: an international journal, 1(2): 15–232.
- Newman, M. (2004): *Who is the Best Connected Scientist? A Study of Scientific Co-Authorship Networks*. In: Ben-Naim, E. – Frauenfelder, H. – Toroczkai, Z. (eds.): *Complex Networks*. Springer, 337–370.
- Rojek, C. (2001): *Celebrity*. London, Reaktion Books
- Senft, T. (2012): *Micro-celebrity and the Branded Self*. In: Hartley, J. – Burgess, J. – Bruns, A. (eds.): *A Companion to New Media Dynamics*. Oxford, Wiley-Blackwell, 346–354.
- Sik D. (2013): *Lash modernitáselmélete*. Jel\Kép, 2013/3–4. Link: http://communicatio.hu/jelkep/2013/3_4/sik_domonkos.htm (utolsó letöltés: 2020. 05. 14.)
- van Dijk, J. (2009): *Users Like You? Theorizing Agency in User-generated Content*. Media, Culture and Society, 31(1) 41–58. DOI: 10.1177/0163443708098245

Linkek

<https://sentione.com>

<https://starnetwork.hu/youtube-statisztikak/youtube-top100-hu/>

<https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>



A feltétel nélküli alapjövedelem megítélésének empirikus elemzése YouTube hozzászólásokon keresztül¹

OLÁH ESZTER²

ABSZTRAKT

Az online térben nagy népszerűségnek örvend a világ egyik legnagyobb videómegosztó platformja, a YouTube, ahol a nézők véleményezhetik a videókat és azok témáit. A tanulmány célja annak vizsgálata, hogy a napjaink gazdasági-társadalmi változásaival párhuzamosan egyre nagyobb figyelmet kapó feltétel nélküli alapjövedelemmel kapcsolatban hazánkban milyen értékek és vélemények rajzolódnak ki a YouTube ilyen témájú videóinak kommentszekciójában. Kutatásunk vegyes módszerrel dolgozik, az adatok gyűjtése, tárolása, szentimentanalízise és a szószák módszere informatikai eljárásokkal valósult meg, míg a kategorizálás kézi kódolás révén. A szentimentanalízis eredménye mutatja, hogy a pozitív érvek kisebb arányban merülnek fel a megjegyzésekben. Olyan értékjellemzőkkel bírnak, mint a befogadás, a létjogán elv, az igazságosság és a szabadság hangsúlyozása. A pozitív érvek közül a legmagasabb támogatottságot a megvalósíthatóság képezi. A negatív kategóriájú értékek gyakrabban merülnek fel, így a másodrangú lét, igazságtalanság, kirekesztés, megfizethetlenség és a teljesítményelvűség értékeinek hangsúlyozása domináns az elemzett megjegyzésekben.

KULCSSZAVAK: szövegbányászat, számítógépes szövegelemzés, szentimentanalízis, feltétel nélküli alapjövedelem, munkaalapú társadalom

ABSTRACT

Empirical analysis of the judgment of unconditional basic income through YouTube comments

One of the world's largest video-sharing platforms is YouTube, where viewers can comment on the videos and their topics. The aim of this study is to examine the values and opinions about unconditional basic income according to the comment sections of several Youtube's videos which topic is the previously mentioned UBI which is receiving increasing attention in parallel with today's economic and social changes. Our research works with a mixed method, data collection, storage, sentiment analysis and the bag of words method which were implemented using IT procedures, while categorization was done through manual coding. The results of the sentiment

¹ A publikáció elkészítését az EFOP-3.6.1-16-2016-00022 számú projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

² Debreceni Egyetem Humán Tudományok Doktori Iskola Szociológia és Társadalompolitika Doktori Program, olah.eszter@arts.unideb.hu



TEMATIKUS TANULMÁNYOK – Digitális szociológia

analysis show that positive arguments appear to a lesser extent in the comments. Positive arguments have value characteristics such as inclusion, the principle of the right to exist, justice and freedom. Among the positive arguments feasibility enjoys the highest support. Negative category values arise more frequently, so the emphasis on the values of injustice, exclusion, unaffordability, and performance-orientation is dominant in the analyzed comments.

KEYWORDS: *text mining, computer text analysis, sentiment analysis, unconditional basic income, work-based society*

Bevezetés

Napjainkban egyre nagyobb jelentőséget kapnak a társadalmi folyamatok megértésében a digitális adatok. A világon előállított információ mennyisége gyors ütemben növekszik és egyre nagyobb mértékben digitálisan kerülnek tárolásra, ezen digitális adatokat tekinthetjük Big Data-nak (Salganick 2017). A nagy adathalmazok számítógépes analízisének során mintázatok és összefüggéseket lehet feltárni, különösen az emberi viselkedés és interakciók terén. Az ilyen típusú információk a társadalomtudományokat olyan vélemények és attitűdök feltárásához vezetheti, amelyek nem érhetőek el a hagyományos módszerek segítségével, vagy azért, mert a téma érzékeny elemeket érint, vagy azért, mert nehéz a megfelelő operacionalizálás (Németh et al. 2020).

Jelen tanulmány nyilvános YouTube videókhoz íródott kommenteken végzett elemzést, a feltétel nélküli alapjövedelem témakörére fókuszálva. Ezen adathalmaz nagysága mellékes abból a szempontból, hogy Big Data-alapúnak tekinthető-e, ennek a kérdésnek az eldöntésében a szövegekre jellemző strukturátlanságot és a standardizátlanságot vettük meghatározónak (Németh et al. 2020).

A közösségi média platformjai és a videómegosztó oldalak jelentőségének felértékelődése is hozzájárul a digitális adatok növekedéséhez. Ezekon a felületeken a magatartásbéli elköteleződést általában szimbolikusan fejezi ki az egyén, a kedveléseken, a megjegyzések írásán vagy a megosztásokon keresztül (Dubovi – Tabak 2021). Ezekon a felületeken a felhasználó dönthet úgy, hogy csendes megfigyelőjeként az eseményeknek passzív tartalomfogyasztó szerepben marad. Illetve választhatja azt is, hogy aktív szereplőként bekapcsolódik a különböző interakciókba (Khan 2017). Egy adott oldal vagy poszt követése és kedvelése a közösségi oldalakon magas fokú érdeklődést és kötődést mutat az adott téma iránt (Salganick 2017). A passzív megfigyelői szereptől jóval aktívabb és magasabb szintű érdeklődést takar a megjegyzések közzététele (kommentelés). Az, hogy az egyén képes bekapcsolódnival az érvelés folyamatába – válaszol más személyek megjegyzéseire, kijelentéseket tesz, állításokat cáfol meg vagy támaszt alá – egyfajta kognitív elköteleződést jelent, mely során a kommentelés aszinkron jellege miatt a hozzászólónak lehetősége van arra, hogy átgondolja véleményét és információt gyűjtsön az adott témára vonatkozóan, mielőtt közzéteszi gondolatait (Lucas et al. 2014).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A koronavírus-járvány tovább növelte a digitalizáció jelentőségét a mindennapjainkban, a távoktatás és a home office által. Továbbá gazdasági és társadalmi változásokat (munkahelyek bezárása, szolgáltatások leállása) tapasztalhattunk meg, melyek gyors és rugalmas alkalmazkodást igényeltek. A pandémia hatására beszűkült pénzkereseti lehetőségek gazdasági és szociálpolitikai intézkedéseket sürgettek, mivel a munka nélkül maradt egyének ellátásának biztosítása világszerte kulcskérdéssé vált. Ezen folyamatok aktualizálták a feltétel nélküli alapjövedelem 19. század óta felmerülő gondolatát. A feltétel nélküli alapjövedelem egy rendszeres időközönként fizetett készpénzjuttatást takar, mely univerzalizációból fakadóan mindenkire kiterjed, elhagyva a rászorultsági vizsgálatot és a munkakötelezettség megszabását (Van Parijs 2010). Az alapjövedelem köztudatban tartása annak köszönhető, hogy támogatói szerint a 21. század kihívásaira megoldási javaslatként szolgálhat, hiszen egyszerre képes megteremteni a szabadabb döntéshozatal lehetőségét a munkavégzés terén és a diszkriminációmentes szociális ellátórendszert (Beck 2010).

Jelen kutatás fókuszpontja egy nehezen kutatható kérdéskörnek tekinthető, hiszen az emberek nem rendelkeznek az alapjövedelemmel kapcsolatosan személyes tapasztalattal, mivel hazánkban nem került bevezetésre egy ilyen univerzális juttatás. Ezért a YouTube platformján a témához kapcsolódó videókhoz íródott megjegyzések elemzésével szeretnénk jobban megérteni a hozzászólók attitűdjeit ezen gazdaságpolitikai eszközzel kapcsolatban. Az alapjövedelem köznyelvi értelmezése sokszínűbb, mint ahogyan a szakirodalom alkalmazza Van Parijs fogalmi lehatárolását. A média résztvevői, a politikai pártok kampányai és az egyes országok kísérleti programjai igen sokféle tartalommal használják a fogalmat. Az általunk elemzett kommentek esetében sem lehetséges pontosan meghatározni, hogy milyen definíció szerint alkotnak véleményt az alapjövedelemmel kapcsolatban a kommentelők. Mindezt figyelembe véve jelen kutatásunk arra helyezi a hangsúlyt, hogy megvizsgálja hozzászólásaik során milyen tartalommal töltik meg a vitában résztvevők az alapjövedelem fogalmát és ehhez az általuk használt fogalomhoz milyen reflexiókat és félelmeket társítanak.

A tanulmány első részében összefoglaljuk a feltétel nélküli alapjövedelem értelmezési lehetőségeit, majd az empirikus kutatásunk adataira építve elemezzük a videómegosztó platformon található megjegyzéseket, hogy kiderítsük milyen jellegzetes értékminták mutathatók ki ezen gazdaságpolitikai eszközzel kapcsolatban az online felület felhasználói között.

A feltétel nélküli alapjövedelem értelmezési lehetőségei

A Van Parijs által használt fogalom szerint működő alapjövedelem típusú juttatást Magyarországon még nem vezettek be. Mégis visszatérő eleme a média hírközléseinek, a közgazdasági és szociológiai diskurzusoknak, ahol vita folyik arról, hogy meg-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

valósítható-e egy ilyen típusú gazdaságpolitikai eszköz. A feltétel nélküli alapjövedelem megítélése nem tekinthető egységesnek. A viták négy fő kérdés körül bontakoznak ki (Gilbert et al. 2018):

Mennyiben megfizethető egy alapjövedelem típusú univerzális juttatás (finanszírozás):

Egy jóléti juttatás esetén kulcskérdés, hogy hosszú távon mennyiben fenntartható. A jóléti programok és az azokkal járó adóterhek nagysága meghatározzák a jóléti államokkal kapcsolatos vitákat (Tomka 2008). Az univerzális jóléti rendszerek működési rendjét élesen bírálják, mivel olyanok is részesednek az ilyen típusú juttatásokban, akik egyébként nem szorulnak rá. Így az ilyen elven működő rendszerek legnagyobb hibájának tekinthető a célzást figyelembe véve az egyéni szükségletek figyelmen kívül hagyása. A finanszírozhatóság és a finanszírozhatatlanság álláspontja mellett közel azonos arányban hozhatók fel érvek és ellenérvek, legyen szó bármilyen típusú jóléti juttatásról. Így a kérdés az, hogy a jóléti programoknak a társadalmi következményei inkább pozitívan, vagy negatívan ítéltetők-e meg, mivel ez megmutatja, hogy mennyiben éri meg a finanszírozással járó esetleges gazdasági következmények terhét felvállalni.

Mennyiben értelmezhető alapvető jogként a társadalom tagjai számára (filozófia):

Magyarországra – számos közép-európai országhoz hasonlóan – egy olyan érték jellemző, amelynek során erőteljesen jelen van a szociális dilemma problémája (Sachs 2015). A dilemma lényege tulajdonképpen az, hogy az egyes egyének mennyiben hajlandóak költségeket vállalni a közösség egészének érdekében, tudva, hogy ezzel a gazdaság működésének javulása is elérhető, ami az egyén életére is pozitív hatást gyakorolhat. A védelemhez való jogot azért lehet követelni, mert kollektív érvényű és törvényileg intézményesített garanciákat jelent, amik függetlenek az egyén személyes tulajdonságaitól és érdemeitől (Castel 2005b). Kérdés, hogy alapjövedelem esetén az egyének alapvető jogként ismernek-e el bármilyen szintű támogatást a magasabb életszínvonal érdekében vagy úgy vélik, hogy mindenki maga felelős saját boldogulásáért.

Mennyiben megvalósítható a hozzá kapcsolódó érdekek metszetében (megvalósíthatóság):

Az a morális gondolat, miszerint a boldog és produktív lakosság növekvő szabadideje jelentős veszélyt jelenthet a fennálló rendre nézve, a mai napig fenntartja társadalmunkban azt a nézetet, hogy a munka az önmegvalósítás eszköze (Graeber 2018). A munkavégzés felemészti az ébren töltött idő nagyobb részét és már csak arra marad ideje az egyénnek, hogy a fogyasztói örömek kielégítésére szánjon időt. Graeber nézete szerint a munkavégzés önmagában erkölcsi értéket jelent, aki pedig nem hajlandó magát munkafegyelem alá vetni az nem is érdemel támogatást. Ennek a gondolkodásmódnak a fenntartása érdekében igyekeznek lekötni a dolgozókat, akár olyan munkákkal is, amelyek a gazdaság számára nem szükségesek. Graeber



TEMATIKUS TANULMÁNYOK – Digitális szociológia

igazgatási feudalizmusnak nevezi a fennálló rendszert, melyet a gazdasági erők ösztönöznek. Amiben azért dolgozunk többet, mert az uralja gondolkodásunkat, hogy a fogyasztói örömök utáni vágyakozást fáradtságos munkával lehet kielégíteni, ahogyan ez a weberi gondolatokban is megjelenik. Eszerint a piaci munkát a jövedelmi hasznon túl, elvi kötelességgént végzi az ember, így a produktív és hasznos munka hozzátartozik a jelentőségteljesnek vélt emberi élethez (Weber 1982). A feltétel nélküli alapjövedelemmel kapcsolatban az egyik leggyakoribb félelem, hogy csökkenjen a munkavállalási motivációt, azáltal, hogy a jövedelmet elválasztja a munkaerőpiaci részvételtől (Bergmann 2004). A feltétel nélküli jelleg veszélyeztetné azt a hozzáállást, hogy a produktív és hasznos munka hozzátartozik a jelentőségteljesnek vélt emberi élethez. A kevesebb munkavállaló hatással lenne a fogyasztás csökkenésére is, ami mérsékelné a gazdaság növekedési ütemét (Baksay et al. 2017). Ez azt eredményezné, hogy a tőke a kedvezőbb befektetési terület felé vándorolna. A tőke elvándorlásának igazi veszélyét a döntéshozói és politikai körök érzékelik, hiszen az alapjövedelem veszélyeztetné a tőke hatékonyságát, a profitra irányuló termelést (Artner 2014).

Mennyiben gyakorol ellenősztonzó hatást a munkavállalásra (motiváció):

A munka jelentőségének fennmaradása mellett szól az ember belső indíttatása, a kényszerítő erő a cselekvésre, a semmittevés helyett, azért, hogy az egyén életében az örökös hiányt kielégíthesse (Keynes 1965, Fromm 2002). Mivel nem az alapvető fiziológiai szükségleteink képezik a cselekvésre való indíttatás alapját, hanem egy mélyebb késztetés, az emberi akarát (Schopenhauer 1992), így a feltétel nélküli alapjövedelem a munkavégzés kényszer jellegének feloldásán túl és az alapvető fiziológiai szükségletek kielégítésén túl a belső faktorokra nehezen tudna hatást gyakorolni. Mivel a felhalmozási vágyat nem tudná kielégíteni és a munka pszichológiai jelentőségét is csak nehezen tudná pótolni. Hiszen az alapjövedelem egyrészt az alapvető szükségletek kielégítését tenné lehetővé és nem az addigi megszokott életszínvonalat (Beck 2010). Másrészt az alapjövedelem a munka kényszere alól szabadít fel, nem pedig az értelmes munkavégzés alól. Annak ellenére, hogy a kapitalizmus alapját a munka adja, mely identitást biztosít a munkavállaló számára, Szelényi (2014) úgy tartja, hogy mindig lesznek olyanok, akiket nem ösztönözne munkavállalásra egy feltétel nélküli juttatás és garantált munkánélküli ellátásként funkcionálna számukra.

A kutatás elméleti háttéréből látható, hogy a feltétel nélküli alapjövedelem nem értelmezhető egyoldalúan. Ezen gondolatok mentén a következő kutatási kérdések fogalmazódnak meg, melyekkel jobban megérthető a társadalom viszonyulása a feltétel nélküli alapjövedelemhez:

Milyen központi fogalmak alkotják a kutatás YouTube kommentár korpuszát, amelyek megalapozzák a társadalom értelmezését a feltétel nélküli alapjövedelmet illetően?



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Milyen megítélés alá esik a YouTube megjegyzésekben a feltétel nélküli alapjövedelem, mely az elemzett online felület felhasználóinak viszonyulását fejezi ezen eszközzel kapcsolatban?

Milyen értékminták jellemzőek az alapjövedelemmel kapcsolatos YouTube megjegyzések íróira?

Módszertan

ADATGYŰJTÉS

A YouTube az egyik legnagyobb online platform, melynek világszerte havonta több mint 2 milliárd bejelentkezett felhasználója van³. Ilyen magas nézettség mellett a társadalmi kérdésekkel foglalkozó videók jelentős hatással lehetnek a nézők véleményformálására. Tang és munkatársai megállapították, hogy a YouTube kevésbé kitett ennek a befolyásnak, hiszen ezen a platformon a felhasználók konkrétan keresnek rá arra (célorientált keresés), ami őket érdekli, ez gátolja a félretájékoztató valószínűségét (Tang et al. 2021).

Mivel a feltétel nélküli alapjövedelem széles körben nem került bevezetésre Magyarországon, így a társadalom tagjai személyes tapasztalatokkal nem rendelkeznek ezen gazdaságpolitikai eszközzel kapcsolatban. Azt feltételezzük azonban, hogy a YouTube videók megtekintése pontosabb betekintést enged a kommentelők számára az alapjövedelem alkalmazásának feltételrendszerébe, jóllehet bizonyítékunk arra vonatkozóan nincs, hogy a vélemény írója a videó megtekintését követően, vagy attól függetlenül fogalmazta-e meg véleményét. A videókhoz fűzött kommentek elemzése azonban mindkét esetben (megtekintette, illetve nem tekintette meg a videót teljes hosszában) hasznos információval szolgálhat arra vonatkozóan, hogy a kommentelők társadalma hogyan értelmezi ezt a jóléti juttatást.

A YouTube videók keresése 2021. február 28-án zajlott keresztmetszeti kutatásként, a „feltétel nélküli alapjövedelem” és „alapjövedelem” kifejezésekre történő kereséssel. A keresést inkognitó módban végeztük, elkerülve a felhasználói fiókhoz kapcsolódó preferenciákat, melyeket a YouTube figyelembe vesz a gépi-tanulás alapú ajánlások révén, melyek a platform algoritmusai szerint működnek. Többirányú keresést folytattunk célorientált keresés alkalmazásával, egyrészt relevancia szerinti rendezés, másrészt megtekintések száma szerinti rendezés szerint, a minél szélesebb lefedettség érdekében. Az esetlegesen megjelenő duplikált videók törlésre kerültek. A videónál fontos kritérium volt, hogy három percnél hosszabb tartalom legyen, hogy képes legyen annyi információt szolgáltatni, ami diskurzust alakít ki a komment szekcióban. Azon videók nem kerültek kiválasztásra, melyeknél a meg-

³ <https://www.youtube.com/intl/hu/about/press/>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

jegyzések írásának lehetőségét kikapcsolta a videót megosztó személy. Továbbá nem kerültek a mintába azon videók kommentjei sem, melyek tíznél kevesebb hozzászólással rendelkeznek. A videókat a kutatók nem tekintették meg, hiszen a kutatás fókuszában nem a videók tartalma áll. Objektív minta kialakítására törekedtünk annak érdekében, hogy meghatározhassuk, hogy az átlag felhasználó milyen képet alakít ki az alapjövedelem témakörével kapcsolatban, politikai hovatartozástól függetlenül. Annak ellenére, hogy a közgondolkodásban a téma átpolitizált, törekedtünk arra, hogy mind támogatói, mind elutasítói oldalról széles körben jelenítsünk meg véleményeket.

A végső elemzésbe 22 videó került, melyek hosszú időszakot ölelnek fel, hiszen a videók 2013 és 2021 között készültek. Az empiria alapjául szolgáló megjegyzések kigyűjtésére a YouTube API⁴ segítségével került sor.⁵ A videókhoz készített adatbázis többek között a következő metaadatokat tartalmazta: videóazonosító (videóhoz hozzárendelt ID és URL), a videókhoz kapcsolódó megjegyzések, megjegyzések kedvelésének száma, közzétételük időpontja. Az adatbázis tartalmazta egyrészt az „eredeti megjegyzéseket”, melyek közvetlenül a videóhoz kapcsolódtak, másrészt a „megjegyzéshez fűzött megjegyzéseket” is (úgynevezett „reply”), ami egy másik kommenthez érkezett válaszként.

Egy videó esetén maximum 100 komment gyűjthető a YouTube API segítségével, melyek a mások által legkedveltebb gondolatoknak minősülnek. A legtöbb kedvelést kapó megjegyzések tartalma jelzi a legjellemzőbb megközelítést az adott témának. Ahogyan az a bevezetőben elhangzott egy adott poszt kedvelése a közösségi oldalon magas fokú érdeklődést és kötődést mutat az adott téma iránt (Salganick 2017). Mivel sokan lehetnek olyanok, akik úgy fejezik ki véleményüket, hogy nem írnak megjegyzést az adott videó alá, de a kedvelés lehetőségét használva kifejezik egyetértésüket. Azonban fontos megjegyezni, hogy a YouTube felhasználók véleménynyilvánítása nem tekinthető reprezentatívnak a magyar társadalom egészére nézve.

A videókat sorszámmal különböztettük meg az elemzés során, címeiket adatvédelmi okokból nem publikáljuk, hogy harmadik személy számára ne legyen visszakereshető az egyes megjegyzések forrása. Hazánkban a közösségi média adatainak elemzésekor a Facebook kínál elsődleges terepet a szöveges adatok kinyerésére, mivel a Twitter használata még kevésbé elterjedt. A 2018-as adatvédelmi botrány után a Facebook szigorúan korlátozta a Facebook API-n keresztül elérhető adatok körét a független kutatók számára (Ben-David 2020).

Ezzel szemben YouTube API rendelkezésre bocsájta a kigyűjtött kommenteket, így nyilvánosnak tekinthetők az ott elérhető adatok, ami azt jelenti, hogy megtekinthetők az adatforrás előfizetése nélkül és a megjegyzés írójának kifejezett engedélye

⁴ Application Programming Interfaces – alkalmazásprogramozási felület

⁵ Ezúton szeretnék köszönetet mondani Tóth Ákosnak a módszertan során fellépő programozási feladatok megvalósításáért.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

nélkül. A legetikusabb eljárás mód szerint eljárva, az elemzés során az adatbázisunkból eltávolításra kerültek a felhasználónevek és elkerüljük a kommentek szözszerinti idézését és a szabad átfogalmazásukat is. Mivel a kutatás etikája azt kívánja meg, – adatvédelem szempontból – hogy ha lehetséges mellőzni kell a pontos idézéseket, mivel így biztosítható annak elkerülése, hogy a felhasználók beazonosításra kerüljenek (Reilly 2014). A hozzáférhető adatok nem tartalmaznak személyes információkat, mint név, kor, lakhely, vagy iskolai végzettség, így az adatbázis csak összesített információkat használ, és nem teszi lehetővé szocio-demográfiai háttérre irányuló következtetések levonását⁶.

ADATTISZTÍTÁS

Az API segítségével leszűrt megjegyzésekből előállt a szöveges adatbázisunk, ami több szűrésen esett át az előfeldolgozási fázis előtt. Így törlésre kerültek a következők:

1. A duplikált kommentek.

2. A magyarázat nélküli kommentek: hordoztok ugyan pozitív vagy negatív értékítéletet, de magyarázatot nem fűztek hozzájuk (hangulatjelek, trágár kifejezések). Ezen hozzászólások az elemzés során nehezen értelmezhetőek, mivel érvelés hiányában nem derül ki, hogy magára a videóra vonatkoznak vagy az alapjövedelem témakörére.

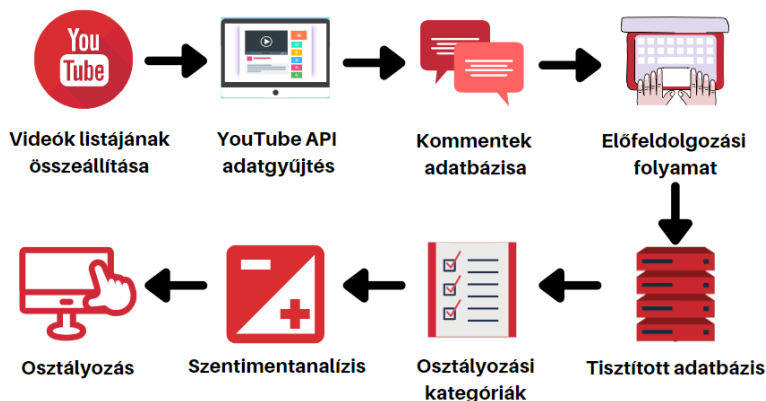
3. A videó készítőjére vagy a csatorna munkásságára vonatkozó kommentek.

Így a több ezer megjegyzésből 489 kommentet sikerült végleges adatbázisba foglalni. Alapvetően kvalitatív információforrás állt rendelkezésünkre, amit kvantitatív módon is elemezhetővé szükséges tenni. Ehhez nyújt segítséget az előfeldolgozási fázis, melynek során a szöveg normalizálására törekszünk. Az adatelemzés megkönnyítése érdekében a következő lépésben a korpusz tisztítása következett, mivel egy elemzésre alkalmas numerikus adatbázis előállítására van szükség (Németh et al. 2020). Az előfeldolgozás során három fontos eljárást szükséges megvalósítani: tiltólistás szavak eltávolítását, a tokenizálást és a szótövesítést (Mészáros – Sebők 2018, Çoban et al. 2021). A magyar számítógépes nyelvészet egyik meghatározó műhelyéhez kapcsolódik a Magyarlanc nevű program fejlesztése, melyet az MTA-SZTE Mesterséges Intelligencia Kutatócsoportja és annak Nyelvtechnológiai Csoportja dolgozott ki (Zsibrita et al. 2013). A Magyarlanc segítségével végeztük el az előfeldolgozás lépéseit. Ennek első lépésében törlésre kerültek az úgynevezett tiltólistás szavak, melyek nincsenek hatással a mondat jelentésére, mint a kötőszavak, névelők

⁶ A kutatás által alkalmazott módszertan innovativitásából adódik, hogy nem a megszokott standard eljárások szerint ad lehetőséget a valóság jelenségeinek elemzésére. Azonban a reflexiók, tudattartalmak és viszonyulások, melyek e kutatás alapját adják feltárhatóak ezzel a módszerrel.

TEMATIKUS TANULMÁNYOK – Digitális szociológia

vagy névmások, így nem hordoznak releváns információt az elemzés szempontjából. A szöveg előkészítésének fontos lépése a tokenizálás folyamata is, aminek során a rendelkezésünkre álló megjegyzéseket tokenekre, azaz szavakra bontjuk a szóközök mentén, eltávolítva az írásjeleket. A karaktersorozat (token) a későbbiekben jelentéstani elemzési egységként szolgál. Az előfeldolgozás harmadik lépése Mészáros és Sebők értelmezésében a ragozások eltávolítása, a szótövesítés (stemming). Ezen eljárás során a szó csonkolása történik, ami így előállíthat nem értelmes szavakat is, a magyar nyelv agglutináló nyelvi jellege okán (Mészáros – Sebők 2018). Az előfeldolgozási folyamat elvégzésével rendelkezésünkre állt a tisztított adatbázis.



1. ábra. Az elemzési folyamat

Forrás: Saját szerkesztés

Ahogy az 1. ábra szemlélteti a tisztított adatbázis segítségével következett az elemzés alapjául szolgáló osztályozási kategóriáknak a kidolgozása a szózsák módszerrel, a széntimentanalízis és végül a kigyűjtött megjegyzéseknek a kategorizálása az előzetesen előállított osztályozási kategóriákba.

ADATELEMZÉS

A kutatás adatelemzése során a deduktív módszertant használtuk. Mivel előzetesen kialakítottunk egy előre definiált kategória-rendszert (Mészáros – Sebők 2018). Így a szövegbányászati feladat célja a korpusz elemeinek besorolása volt ezen kategóriákba, az emberi nyelvi tudást ötvözve a számítógépes tudással. Szabadon hozzáférhető szoftverrel végeztük az elemzést, R programnyelvben.

Az adott dokumentumok jellemzőinek feltárása a szavak és kifejezések gyakoriságával a leghatékonyabb (Mészáros – Sebők 2018). Ennek egyik alapja a szózsák



TEMATIKUS TANULMÁNYOK – Digitális szociológia

módszere. Ezen eljárás során az egyes szavak gyakoriságát elemeztük a rendelkezésünkre álló korpuszon belül. Ennek során meghatároztuk a leggyakoribb kifejezéseket, szókapcsolatokat, melyeket 5-5 csoportba soroltunk be negatív és pozitív értékelés alapján. A leggyakoribb kifejezések grafikus ábrázolásához a szófelhő eljárását alkalmaztuk. A YouTube kommentár korpuszában a központi fogalmak megállapításával igyekeztünk választ adni arra a kérdésre, hogy mely kifejezések segítségével történik a feltétel nélküli alapjövedelem hozzászólók általi értelmezése.

A feltétel nélküli alapjövedelemmel kapcsolatban érdekes információkat hozdozhat az, hogy a megjegyzések írói milyen álláspontot képviselnek ezen eszközzel kapcsolatban. Ennek eszköze a szentimentelemzés (másnéven szentimentanalízis), ami a szöveg szerzőjének attitűdjeit és véleményét tárja fel (Németh et al. 2020). Ez az elemzés képes megmutatni a vélemények polaritását, így az egyes szövegrészek csoportosíthatóvá válnak pozitív, negatív és semleges kategóriákba. A szentimentanalízis fontos a kommunikáció során, mivel segít megérteni, hogy az adott üzenetre milyen a közvélemény reakciója (Thelwall 2017). Mivel meghatározott számú kategóriába szeretnénk besorolni a szövegrészeket, így a szentimentelemzés is egyfajta osztályozási problémának tekinthető. Ezen természetesnyelv-feldolgozási módszer (NLP módszer) alapját a magyar nyelvre érvényes szentimentszótár képezte (Németh et al. 2020). Az NLP-be minden olyan módszer beletartozik, amelyek nagy mennyiségű, természetes nyelven született szövegek elemzésére alkalmasak lehetnek (Hirschberg – Manning 2015). A véleményelemzés is ehhez járul hozzá. Az elemzésünket a Magyar Szentiment Lexikon segítségével végeztük el, melyben a szavak polaritását meghatározták (Szabó 2014). Ezen szólista 7688 kifejezést tartalmaz, mely biztosította, hogy közel az összes YouTube megjegyzés kategorizálhatóvá váljon a pozitív vagy negatív csoport valamelyikébe. Félíg felügyelt módszerként alkalmaztuk ezt az eljárást, hiszen, ahol automatikusan a gép nem tudta kategorizálni az adott szövegrészletet – a köznyelvi megfogalmazás miatt – ott kézi kódolásra tértünk át. A szentimentanalízis során nem alkalmaztunk neutrális csoportot, mivel azt feltételeztük, hogy azon személyek, akik kommentelnek egy előzetesen kiforrott nézettel rendelkeznek, melyet megszeretnének osztani a közvéleménnyel. A szöveg hangulatelemzése hozzájárul ahhoz, hogy megismerjük, hogy a kommentelők körében a feltétel nélküli alapjövedelem kapcsán az elfogadás vagy az elutasítás megítélése a dominánsabb, mivel ez mutatja az elemzett online felület felhasználóinak viszonyulását ezen gazdaságpolitikai eszköz felé.

A szózsák módszerének segítségével kialakított 5-5 csoport alapján és a szentimentanalízis eredményeit alapul véve elvégeztük a kommentek kategóriákba való besorolását, kézi kódolással. Ennek egyik legfőbb előnye, hogy a szöveg kutató általi értő olvasása miatt érvényesebb eredményeket ad, mint egy gépi kódolás esetén (Mészáros – Sebők 2018). Mivel köznyelvben íródott megjegyzések elemzése



TEMATIKUS TANULMÁNYOK – Digitális szociológia

képezi a kutatás alapját, így fontos, hogy a kézi kódolás segíti az ironia és a szleng felismerését vagy a közvetett utalásokat is. Annak érdekében, hogy a kutatói szubjektivitás hatását csökkentjük a kézi kódolás esetén kettős vak kódolást alkalmaztunk. A két kódoló, egymástól függetlenül bekezelte a letisztított kommenteket az előzetesen meghatározott lista alapján. Majd közösen határozta meg a végső besorolást, megvitatta az önálló kódolás eredményét. Ezzel azonosítva a leggyakrabban megjelenő értékmintákat a feltétel nélküli alapjövedelemmel kapcsolatban, a YouTube platformján.

Eredmények

KÖZPONTI FOGALMAK ÉS A FELTÉTEL NÉLKÜLI ALAPJÖVEDELEM TÁRSADALMI MEGÍTÉLÉSE A KUTATÁS YOUTUBE KOMMENTÁR KORPUSZÁBAN

Elsőként arra kerestük a választ, hogy melyek azok a fogalmak, amelyek meghatározzák a feltétel nélküli alapjövedelem értelmezését a YouTube általunk vizsgált megjegyzéseiben. Ennek alapja a leggyakoribb kifejezések meghatározása a szószák módszerével. A 2. ábrán az áttekinthetőség érdekében a 20-nál többször előforduló szavak megjelenítése látható. A szófelhőben a szavak mérete mutatja az előfordulásuk gyakoriságát. A kutatás témája miatt nem meglepő, hogy az „alapjövedelem” kifejezés a legmeghatározóbb az elemzett korpuszban. Ezen túlmenően a szófelhőben három kulcsdimenziót azonosíthatunk. Egyrészt a munkát középpontba állító dimenziót, ahol a „dolgozik”, a „pénz”, az „ingyen” és a „munka” képezi a kulcsfogalmakat. Ezen kifejezés élesen utalnak a jelenlegi társadalmi berendezkedésünkre, melyben a munka központi helyet foglal el az egyének életében. A következő dimenzió a morális kategória, melybe olyan szavak sorolhatók be, mint „ember”, „jog”, „szükség”, „segély”, „társadalmi” és „szociális”, melyek az előző csoporttól eltérően a feltétel nélküli alapjövedelem morális és jóléti területére utalhatnak. A harmadik dimenzió a nagy rendszerek kategóriája, melybe a „kormány”, „állam” és „kapitalizmus” szavak tartoznak, szimbolizálva, hogy a feltétel nélküli alapjövedelem témaköre nehezen elválasztható a jelenlegi gazdasági és társadalmi viszonyok értelmezésétől.



TEMATIKUS TANULMÁNYOK – Digitális szociológia



2. ábra. Leggyakoribb szavak grafikus ábrázolása
Forrás: Saját szerkesztés

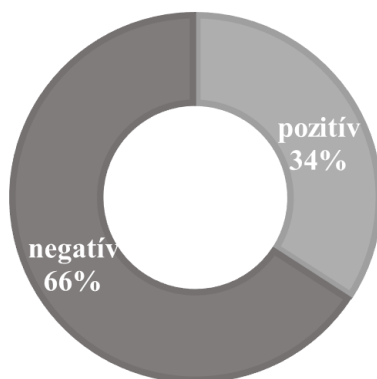
A szavak és kifejezések megjelenésének gyakorisága szemlélteti a központi fogalmakat a kutatás YouTube kommentár korpuszában. Mivel nem állt rendelkezésre előzetesen a feltétel nélküli alapjövedelem témaköréhez kötődő szólista, ezért egy ilyen típusú eszköz validálása is ezen kutatás során történik. Mivel egy ilyen szólista segíti az egyes megjegyzések osztályozását, és a közöttük lévő logika feltérképezését. A szólista kiindulópontjának a tanulmány elején említett Gilbert és munkatársai által kidolgozott megközelítést tekintjük (Gilbert et al. 2018). Eszerint a feltétel nélküli alapjövedelemhez kapcsolódó viták négy fő kérdés körül bontakoznak ki: a finanszírozhatósági, a morálfilozófiai, a megvalósíthatósági, és motivációbeli kérdések. Így a szólistánk egyik kategóriáját az alapjövedelem fiskális jellemzőihez kötődő megjegyzések képezik. Két dimenzióját különböztetjük meg: megvalósítható, mint pozitív vagy megfizethetetlen, mint negatív dimenzió. A második kategóriát a morálfilozófia szempontok alkotják, melyek elsődlegesen a juttatás feltétel nélküliségéhez kapcsolódnak (két dimenziója: igazságosság és igazságtalanság). Gilbert és munkatársainak harmadik pillérét – megvalósíthatóság – jelen kutatás során két részre bontjuk, mert úgy ítéljük meg, hogy így jobban lefedhető a kérdéskör. Egyrészt az univerzalitás (befogadás és kirekesztés dimenzió) kapcsán felmerülő érdekekre bontható, amely a megvalósítás előzményének nagy kérdése. Másrészt a szubjektív jóléthez (szabadság és másodrangú lét dimenzió) kötődő érvekre, melyek a megvalósítás pszichológiai és társadalmi következményeire vonatkoznak. Az ötödik pillér



TEMATIKUS TANULMÁNYOK – Digitális szociológia

esetén a motivációbeli kérdések állnak a középpontban, melyek a normarendszer átalakulásához kötődnek (léti jogosultság és teljesítményelvűség dimenzió). Mind az öt kategóriánál megkülönböztethető a pozitív és negatív dimenzió is, ahogyan az elméleti keretknél is érzékelhető volt a feltétel nélküli alapjövedelem értékelésében meghúzódó kettősség. A csoportok kulcsszavait a 4. és 5. számú ábrák tartalmazzák, amiket a szavak és kifejezések gyakoriságának vizsgálatával állítottunk elő és a kutatás későbbi fázisában ezen csoportokba kategorizáltuk be az egyes megjegyzéseket.

Ezen csoportok polaritását a szentimentanalízis segítségével határoztuk meg, a rendelkezésünkre álló szótár segítségével (Szabó 2014). A 3. ábra szemlélteti, hogy a megjegyzések 66%-a negatívnak minősül.



3. ábra. A megjegyzések szentimentanalízisének eredménye (%)
Forrás: Saját szerkesztés

Az alapjövedelemre jellemző pozitív tartalmú kifejezéseket öt csoportba tudtuk besorolni: megvalósíthatóság, befogadás, igazságosság, léti jogosultság és szabadság elnevezésekkel illettük ezeket a csoportokat. A 4. ábrán látható, hogy a megvalósíthatóság körébe kerültek azok a szavak, melyek tisztán az alapjövedelem fiskális lehetőségeire utalnak, mint pénznyomtatás, jövedelemadó vagy átcsoportosítás. A befogadás csoportjába az univerzalitás teljes elfogadására jellemző szókapcsolatok kerültek, melyek többek között a következőkkel írhatók le: mindenki, nem segély vagy segítség. Az igazságosság csoportja a feltétel nélküliséghez kötődik, mint alanyi jog vagy keresztényi kötelesség értelmezése. A léti jogosultság csoportját, – mely a normarendszer átalakulásának pozitív értelmezését fedi le – olyan kifejezések határozzák meg, mint az örömforrás munkán kívüli keresése vagy az átalakuló munkafogalom. A szubjektív jólét pozitív oldalát a szabadság kérdésköre határozza meg, ahol a méltó élet, a bizalom és a pozitív gondolkodásmód kerül hangsúlyozásra.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

MEGVALÓSÍTHATÓSÁG

megvalósítható; megtakarítás;
fokozatosság; lehetséges;
átcsoportosítás; gazdaságnövelő;
adó; AFA; pénznyomatás; unió;
progresszív jövedelemadó

BEOFADÁS

mindenki; rászoruló védelme;
elkerülhetetlen; nem segély;
segítség; társadalmi hierarchia;
minőségi megélhetés

IGAZSÁGOSÁG

nem megbélyegző; alapjog;
éhenhalás ellen; emberi jog;
alanyi jog; születési jog;
keresztényi kötelesség; nem
alamizsna

LÉTJOGOSULTSÁG

átalakult munkafogalom;
robotizáció; kreativitás; örömforrás/
értékteremtés munkán kívül;
kötelességérzet; önkéntesség;
láthatatlan munka felértékelődése

SZABADSÁG

szabadság; méltó élet; élet
értelme; bűnözés csökkenése;
szolidaritás; felszabadítás;
stresszmentesség; pozitív
gondolkodásmód; szabadidő;
bizalom

4. ábra. A feltétel nélküli alapjövedelem pozitív jegyeinek szólistája

Forrás: Saját szerkesztés

A feltétel nélküli alapjövedelem negatív jegyeit szintén öt csoportba tudtuk besorolni, melyek az 5. ábrán látható módon a pozitív jegyek ellentétjeit képezték: megfizethetetlen, kirekesztés, igazságtalanság, teljesítményelvűség és másodrangú lét. A megfizethetetlen csoportba minden olyan kifejezés bekerült, amik az alapjövedelem fiskális elutasítására utalnak: infláció, adósság vagy versenyhátrány. A kirekesztés az univerzalitás teljes elutasítását jelenti, az élőködés és a visszaélés kifejezések hangsúlyozásával. A feltétel nélküliség igazságtalanság dimenzióját a parazita viselkedés és potyautasság szófordulatai határozzák meg. A normarendszer negatív oldala a teljesítményelvűség elnevezést kapta, mely a munkaalapú társadalom és a munka általi értékteremtés szókapcsolataival írható le. A másodrangú lét elnevezésű csoport a szubjektív jólét értelmezéséhez kötődik, mely a polarizációval és az állami függőséggel értelmezhető.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

MEGFIZETHETETLEN

fiktív; reménytelen; infláció
növelő; drága; államadósság;
deficit; nincs fedezet;
eladósodottság; gazdasági
különbség; versenyhátrány

KIREKESZTÉS

kisebbség; bevándorló;
sokgyermekes; gazdag;
köz munkás; dolgozni nem akaró;
univerzalitás zsákutca; elherdál;
élősködés; visszaélés

IGAZSÁGTALANSÁG

parazita; potyautas; pénz nem
megoldás; társadalmi hozzájárulás
csökkenése; feltétel szükséges

TELJESÍTMÉNYELVŰSÉG

munkaalapú; elkényelmesedés; lusta;
szenvedés útján; munkavégzés
értékteremtés; munka kiváltsága;
hivatástudat; ingyen ebéd; vegetáció;
siker záloga; megváltoztathatatlan
alaptermészet; munka általi elismerés;
teljesítménycsökkenés

MÁSODRANGÚ LÉT

érdekellentét; állami szerepvállalás;
függőség; önállóság csökkenése;
pszichológiai korlát; polarizáció;
hedonizmus; gondolkodás
korlátozása

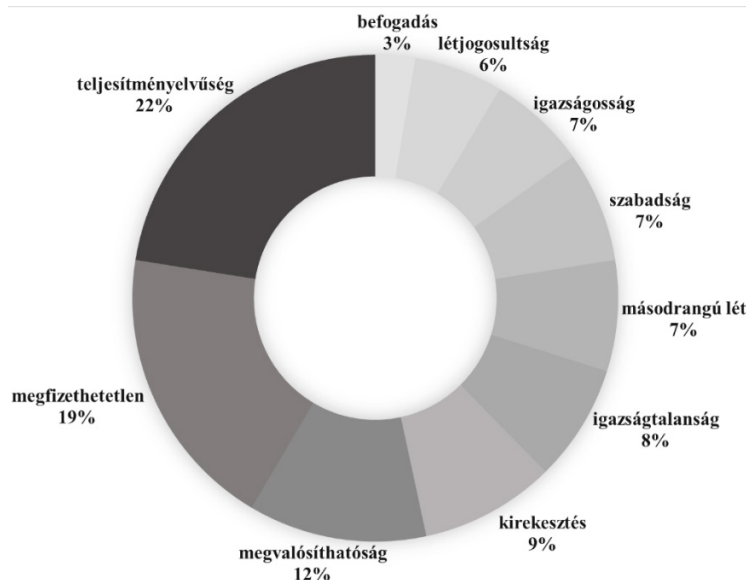
5. ábra. A feltétel nélküli alapjövedelem negatív jegyeinek szólistája

Forrás: Saját szerkesztés

A megjegyzésekre jellemző értékminták

Az előbbieken bemutatott szólisták képezik a következőkben bemutatandó osztályozás alapját, mely során kézi kódolás segítségével az egyes kommenteket kategorizáltuk az 5-5 csoport valamelyikébe. A szózsák módszerével előállított szólisták segítségével zajlott az osztályozás. Ennek során arra voltunk kíváncsiak, hogy kimutathatók-e domináns értékminták, melyek a feltétel nélküli alapjövedelemmel kapcsolatos megjegyzések íróit jellemzi. A 6. ábrán láthatóak az egyes kategóriák százalékos megoszlásai:

TEMATIKUS TANULMÁNYOK – Digitális szociológia



6. ábra. A megjegyzésekre jellemző értékminták megoszlása (%)

Forrás: Saját szerkesztés

A finanszírozáson belül a két szélsőséges álláspontot a megvalósíthatóság lehetősége és a megfizethetlenség alkotja. A második és a harmadik legjellemzőbb értékekről van szó az osztályozás eredménye szerint, tehát a finanszírozhatóság kérdésköre foglalkoztatja leginkább a kommentelőket. Mivel az elemzett megjegyzések 19%-a ítéli meg úgy, hogy a feltétel nélküli alapjövedelem megfizethetetlen. 12%-a azonban pozitívan tekint az alapjövedelem finanszírozhatóságára és megvalósíthatónak tartja azt.

A legnagyobb arányban a hozzászólókra a teljesítményelvűség hangsúlyozása jellemző, mely mellé úgy tűnik az alanyi jog nehezen párosulhat. A normarendszer megváltozásának két végpontja között ugyanis a legnagyobb a szakadék a többi vizsgált kategóriához képest. Hiszen a teljesítményelv 22%-a, kerül szembe a létjogán elv elfogadottságának 6%-os értékével.

Meglehetősen nagy különbség tapasztalható az univerzalitás két értéke között is, mivel a kirekesztés a negyedik leggyakoribb pontja az alapjövedelemmel szembeni negatív értékítéleteknek, a maga 9%-ával, míg ennek ellentéte, a befogadás a legkisebb értéket képviseli, elenyésző 3 %-kal.

A feltétel nélküliség esetén a kommentelők a két végpont értékeit közel azonos arányban támogatták: az igazságtalanság 8%, míg az igazságosság 7%-ban jelent meg az érvelések során.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A szubjektív jólét két értéke a másodrangú lét (7%) és a szabadság (7%) azonos arányban merültek fel a megjegyzések között.

A 6. ábrán az is szembetűnő, hogy a pozitív érvek merülnek fel kisebb arányban a megjegyzésekben, mivel a befogadás, a létjogán elv, az igazságosság és a szabadság a legkisebb arányban előforduló dimenziók. A pozitív érvek közül a legmagasabb támogatottságot a megvalósíthatóság képezi (12%), ami elszakadt a többi pozitív érvcsoporttól és bekerült a negatív szempontok közé, melyek jóval gyakrabban merülnek fel, mint a pozitív megközelítésű kategóriák. Így a másodrangú lét, igazságtalanság, kirekesztés, megfizethetlenség és a teljesítményelvűség dimenziója kerül az élmezőnybe.

Diskurzus

A FELTÉTEL NÉLKÜLI ALAPJÖVEDELEM MEGÍTÉLÉSE

Az elemzett válaszok 66%-a negatív értékelő véleményt tartalmazott, és csak 34%-a fejezte ki támogatását ezen eszköz valamely tulajdonságával kapcsolatban. Annak oka, hogy közel kétszerese a negatív értékítélet a pozitív hozzáállásnak az lehet az oka, hogy a médiában az újszerűnek és kiszámíthatatlannak tűnő témák a társadalom idealizált rendjét veszélyeztető kockázatként kerülnek feldolgozásra (Kitzinger 2000). A feltétel nélküli alapjövedelem kettős értelemben is bizalmatlanságot vált ki az egyénekből. Egyrészt bevezethetősége számtalan kérdést vet fel, az ezen eszközzel kapcsolatos diskurzusok tele vannak kételyekkel és megválaszolatlan kérdésekkel. Másrészt ezek a kérdések az emberi élet egyik legkényesebb és legmeghatározóbb területéhez, az egzisztenciához kapcsolódnak. Az emberek életében a fenyegettség érzése az egzisztenciális kiszolgáltatottság kérdéskörében a legerősebb, hiszen emberi természetünkben ered létfenntartásunkra való állandó törekvésünk (Policy Agenda 2018). Ennek értelmében a negatív értékítélet magas aránya logikus reakcióként mutatkozik a kommentelők válaszaiban.

A leggyakoribb szavak grafikus ábrázolása is alátámasztja, hogy a megjegyzések az egzisztenciális biztonságra fókuszálnak. A javak elosztásának terén kétféle elosztórendszerrel beszélhetünk, egyrészt a piaci rendszerről, másrészt a jóléti állam által megvalósított újraelosztásról (Örkény – Székelyi 2011). Ez a kettősség jelenik meg a szöveghő esetében is. Egyrészt bizonyos szavak a munkaerőpiac révén elérhető egzisztenciát szimbolizálják (dolgozik és munka kifejezések), másrészt megjelenik a megélhetés biztosításához szükséges jóléti hozzájárulás szükségessége is (jog, segély és szociális kifejezések).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

EGZISZTENCIÁLIS BIZTONSÁGRA HATÓ ERŐK: NORMARENDSZER ÉS FINANSZÍROZHATÓSÁG

Az osztályozás során láthatóvá vált, hogy a feltétel nélküli alapjövedelemmel kapcsolatban a legjellemzőbb értékminta a teljesítményelvűség hangsúlyozása (22%) és a megfizethetlenségére való hivatkozás (19%). Ez a két negatív értékminta is szorosan kötődik az egzisztenciális biztonsághoz. A teljesítményelvűség (normarendszer negatív dimenziója) kapcsán az a félelem fogalmazódik meg a kommentekben, hogy az alapjövedelem teljesítménycsökkenést és lustaságot eredményezne a társadalomban. A megfizethetlenség (finanszírozhatóság negatív dimenziója) kapcsán erőteljes ellenállás figyelhető meg az alapjövedelem fiskális lehetőségeivel kapcsolatban.

A jelen kutatás alapjául szolgáló megjegyzésekben a munkavégzést szorosan összekapcsolják az értékteremtéssel és a hivatástudattal. A magyar ember jellemzően munkacentrikus (Kapitány – Kapitány 2014), így a munka általi elismerés és siker elsődleges szerepet játszik életünkben, melyek a kommentekben is hangsúlyozásra kerültek. A társadalom azt ismeri el, aki dolgozik, mivel a társadalom középpontját képezi a munka. Ez a fajta munka-felfogás alapvetően jellemző a kommentekre. Ezért a megjegyzések íróinak nagy százaléka csak munkaalapú társadalomban képes gondolkodni, amivel a feltétel nélküli alapjövedelemmel járó kevesebb munkaórát és teljesítménycsökkenést nem tekintik összeegyeztethetőnek.

Kapitányék lehetséges megoldásnak látják a kilépés-kivonulás lehetőségét a munka túlcentralizált rendszeréből (Kapitány – Kapitány 2014). Ennek következtében erősödne a saját lábra állás jelentősége, mely során az egyén saját céljait követheti, szabadon választható tevékenységeivel, ami javítja teljesítményét és nagyobb kielégülést eredményez számára. A saját lábra állás a kistermelő vagy kisvállalkozó pozíciókkal és az értelmi munka végzésével lehetséges. A feltétel nélküli alapjövedelem esetén a létjogosultság elnevezésű értékminta csoportban (normarendszer pozitív dimenziója) a munkának ez a szabadabb formája fogalmazódik meg, az átalakult munkafogalomra való hivatkozással.

A szabadabb munka egyik alapfeltétele a belső motiváció az adott tevékenység elvégzésére (Kapitány – Kapitány 2014). Keynes megítélése szerint a belső készítés arra sarkallja az egyént, hogy jobban kedvelje a cselekvést, mint a tétlenkedést (Keynes 1965). A feltétel nélküli alapjövedelem létjogosultság kategóriájának értelmezésekor nem a munka értékeinek felszámolódása látható, hanem a munka elsődleges szerepének fellazulása. A létjogosultság típusú megközelítést meglehetősen alacsony arányban képviselték a kommentek írói (6%), mivel a jelenlegi gazdasági és társadalmi rendszerben a szabad munkafelfogás távolinak tűnik. Ennek oka, hogy a piaci viszonyok és a politikai rendszer működése nem kedvez a munka szabadságának (Kapitány – Kapitány 2014). Honneth (2009) ezzel összefüggésben arra hívja fel a figyelmet, hogy a munka világától való intellektuális elfordulás ten-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

denciájával a népesség általános hangulata nincs összhangban, mivel a népesség nagy része még mindig a munkafolyamatban betöltött szerepe alapján jelöli meg társadalmi identitását. Ez eredményezte a teljesítményelvűség értékének dominanciáját a létjogán elvvel szemben.

A szabadabb munkafelfogással ellentétben a finanszírozhatóság pozitív dimenziója – a megvalósíthatóság – jóval nagyobb támogatottságot kapott (12%). Átcsoportosításokkal vagy az adórendszer átalakításával tartják megvalósíthatónak a feltétel nélküli alapjövedelem összegét. Ennek alapja lehet, hogy amennyiben az alapjövedelem csökkenti a meglévő szociális védelmi rendszerek bürokratikus kiadásait és összetettségüket, akkor az pozitívan hathat a fiskális megvalósíthatóságára is (Hirsch 2015).

ÉRDEKEK METSZETÉBEN

A negyedik leggyakrabban felmerülő kategória a kirekesztés (univerzalitás negatív dimenziója). A kommentek 9%-ára erőteljes bünbakképzési mechanizmus jellemző. Ez egyfajta tömeglélektani jelenségnek tekinthető, melynek során pontosan azon egyéneket nevesítik meg a problémák okozóiként, akik azokat a társadalmi problémákat legnagyobb arányban elszenvedik (Csoba 2011). Ez a mechanizmus általában nem bizonyított tényeken nyugszik, sokkal inkább előítéleteken alapszik. A hozzászólók erőteljesen a kisebbség tagjait rekesztenék ki a feltétel nélküli alapjövedelem köreiből. E csoportba tartozónak tekintik a sokgyermekes családokat és a közmunkásokat is, akiket érdemtelennek bélyegeznek meg egy univerzális típusú juttatásra.

Mivel a munkaalapú társadalom fontos meghatározója az emberek gondolkodásának, így ezen csoportokkal a kommentelők elsődleges problémája a munkanélküliségük. A munkanélkülieket elsősorban azért érik támadások, mert a társadalom többi tagja úgy ítéli meg, hogy a szociális ellátásaikkal kivonják magukat a társadalmi kötelezettségeik alól és nem kapcsolódnak be a társadalmi értékek termelésébe (Csoba 2011).

Castel megítélése szerint a ressentiment érzése az irigység és megvetése keveréke (Castel 2005a). Az, hogy ez az érzés a társadalmi helyzetkülönbségből táplálkozik, oda vezet, hogy az egyén saját balszerencséjéért a társadalmi hierarchiában vagy a közvetlenül alatta, vagy a közvetlenül fölötte álló társadalmi csoportot okolja. A ressentiment érzése olyan erős frusztrációt vált ki, mely védekező attitűdöket alakít ki az egyénben, aminek következtében bármilyen újdonságot idegennek tekintenek és elzárkóznak a pluralizmus elől, a különbségeket ezek értelmében nem hajlandóak tolerálni. A kirekesztés kategóriáján belül a hozzászólók a bevándorlók – mint közvetlenül alattuk – és a jómódúak – mint közvetlenül felettük – társadalmi csoportját nevezik meg, akiknek semmilyen körülmények között nem adnának alapjövedelem



TEMATIKUS TANULMÁNYOK – Digitális szociológia

típusú juttatást. A magas jövedelműeket nem tartják rászorulóknak egy univerzális juttatásra (Tanner 2015, Piachaud 2016). A bevándorlók kirekesztésének igényét a bevándorlásellenes politikai retorika is erősíthette. 2010-től Magyarországon bevándorlásellenes retorika volt megfigyelhető (Dessewffy et al. 2018).

Az erőteljes kirekesztő hozzáállással szemben az univerzalitás pozitív dimenziójára – befogadás – elenyésző utalás történt, csupán 3%-a az elemzett mintának. A megjegyzésekben erőteljesebben van jelen az univerzalitástól való elzárkózás, a bizonyos típus emberek kirekesztése, mint a mindenkit befogadó attitűd.

A feltétel nélküliség kategóriájának két dimenziójában már nem tapasztalható ekkora megosztottság a hozzászólások között (igazságtalanság – 8%, igazságosság – 7%). Az univerzalitás kategória befogadás dimenziójának alacsony aránya miatt meglepő, hogy a feltétel nélküliség kategória igazságosság dimenziója kétszeres jelentőségre tett szert. Ennek keretében ugyanis a hozzászólók úgy ítélik meg, hogy a jóléti állam által megvalósított újraelosztásra szükség van. Ennek segítségével elkerülhető a társadalmi leszakadás, biztosítható az állampolgárok jóléte, a társadalmi egyenlőtlenségek csökkentése, vagy a méltó emberi élethez való feltételek megteremtése (Örkény – Székelyi 2011).

Örkény és Székelyi vizsgálatában rekonstruálta, hogy mit tartanak az emberek igazságosnak és igazságtalannak, és milyen elveket választanak akkor, amikor egy igazságosabb világról beszélnek (Örkény – Székelyi 2010). Jellemző, hogy a legelfogadottabb igazságossági elvnek az individualista irányultságú ideológia bizonyult (egyéni teljesítmény és érdem hangsúlyozása, rendszer a semlegesség és az esélyegyenlőség elvén nyugszik) és az egalitarianizmus is viszonylag magas arányban elfogadott (egyenlőtlenségek mérséklődnek, állam normatív szabályozójává válik a társadalmi elosztásnak, kollektív intézményi felelősség hangsúlyozása). A korábban hangsúlyozott munkaalapú társadalom magas támogatottságát igazolja Örkény és Székelyi vizsgálata is, miszerint a legigazságosabb elv az individualista irányultságú ideológia. Ennek keretében az egyéni teljesítmény és érdem kerül hangsúlyozásra a társadalom működésében. A feltétel nélküliség kategória két dimenziójának egymáshoz közeli értékét az individualista elv elsődlegessége és az egalitarianizmus elvének erős, de nem domináló pozíciója indokolja. Az igazságtalanság dimenzióin belül az individualista igazságossági elv érvényesülése látszik, mivel képviselői hangsúlyozzák, hogy a rendszer a semlegesség és az esélyegyenlőség elvén nyugszik, így mindenkinek egyenlő esélye van előbbre jutni. A jogokat pedig nem lehet ab ovo szavatolni, mert az csak a társadalmi hozzájárulás csökkenését eredményezné. Az igazságosság dimenzió kevéssé marad el az igazságtalanság dimenziója mellett, ennek oka az egalitarianizmus igazságossági elv erős, de nem domináló pozíciója lehet. Ennek keretében az egyenlőtlenségek mérséklődésére kerül a hangsúly és arra, hogy az állam normatív szabályozójává válik a társadalmi elosztásnak. Ennek értelmében a megjegyzések arra vonatkoznak, hogy a feltétel nélküliség igazságos mivel képes megakadályozni a szegénységet és a társadalmi leszakadást. A két elv



TEMATIKUS TANULMÁNYOK – Digitális szociológia

hangsúlyos volta a társadalmon belül, megnehezíti a feltétel nélküli alapjövedelem megítélését az igazságosság szempontjából.

A GONDOLKODÁSUNK BÉKLYÓI

A szubjektív jólét kategória két dimenziója – mind a szabadság, mind a másodrangú lét 7% – megegyező arányban fordult elő a megjegyzésekben. A szabadság körébe tartozó értékek – mint szolidaritás, bizalom vagy pozitív gondolkodásmód – a feltétel nélküli alapjövedelem szubjektív jólét terén bekövetkező pozitív változásait takarja. Ezzel szemben a másodrangú lét során a függőség és a polarizáció erősödő folyamatai kerültek hangsúlyozásra. A másodrangú lét a legalacsonyabb arányban előforduló negatív dimenzió lett. Ennek oka az lehet, hogy az egyén először arra törekszik, hogy az anyagi jellegű, létbiztonsággal kapcsolatos szükségleteit kielégítse, ezek kielégítése után következhet a magasabb rendű szükségletek felismerése (Inglehart – Flanagan 1987). Az élet anyagi dimenzióinak biztosítása jobban foglalkoztatja az embereket – ami a finanszírozhatósághoz és a munkavállalás normájának átalakulásához kötődik – mint a magasabb rendű igények megfogalmazása, amelyek nehezebben azonosíthatók.

Magyarország a zárt gondolkodás jegyeit mutatja (Keller 2009). Ezek megfigyelhetőek abban, hogy kevésbé tartjuk fontosnak a civil és politikai szabadságjogokat, kisebb mértékű a mindennapi aktív politikai szerepvállalásunk, kevésbé toleráljuk a másságot, így kevésbé is bízunk másokban, illetve értékeinkben kisebb szerepet szánunk az önmegvalósítási értékeknek. Ezek a jellemzők azt eredményezik, hogy a szabadság kérdésköre nem válik dominánssá a másodrangú lét szempontjaihoz képest.

A szubjektív jólét esetén szemben áll egymással, azaz érv, hogy az alapjövedelem nagyobb fokú szabadságot képes garantálni és az, hogy az állami szerepvállalástól való függőség növekedne. A jóléti rendszerek neoliberais kritikái szerint az állam növekvő szerepvállalása olyan rendszerhez vezet, amely korlátozza az egyén szabadságát (Tomka 2008). Egyrészt azért, mert a kötelező biztosítások ellentétesek az egyén választási szabadságával, másrészt a nagy adóelvonások és progresszív adók rendszere a tulajdonnal való rendelkezés jogát sértik. Hayek szerint a jóléti állam aláássa a piac hatékony működését, mivel a sikeresek vagyonát elveszi, míg ezzel párhuzamosan meghosszabbítja a rászorultak függőségét (Hayek 1978). Értelmezésében az állami beavatkozást szükséges minimálisra redukálni, annak érdekében, hogy az államtól való függést ne erősítse.

Nagy kérdés tehát, hogy ha az alapjövedelem bevezetésre kerülne, akkor az emberek képesek lennének-e élni az általa biztosított nagyobb szabadsággal és a magyar társadalom képes lenne-e erőteljesebben képviselni a posztmaterális értékeket? Vagy a másodrangú lét függőségi állapotába süllyednénk-e, a zárt gondolkodás, gondolkodásmód korlátaiba ütközve?



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Összegzés

A tanulmányban a feltétel nélküli alapjövedelemre fókuszáltunk, mely napjainkban a gazdasági és társadalmi folyamatok változásaival párhuzamosan egyre nagyobb figyelmet kap. Az ezen eszközzel kapcsolatos értékeket és véleményeket igyekeztünk feltárni a YouTube ilyen témájú videóinak kommentszekciójában. Kutatásunk vegyes módszerrel dolgozik, mivel az adatok gyűjtése, tárolása, szentimentanalízise és a szózsák módszere informatikai eljárásokkal valósult meg, míg a kategorizálás kézi kódolás révén, kutatói eljárás eredménye.

A leggyakoribb szavak grafikus ábrázolása azt mutatja, hogy a megjegyzések az egzisztenciális biztonságra fókuszálnak. Ez az a témakör, amely leginkább foglalkoztatja a kommentelőket az alapjövedelem témakörével kapcsolatban. Egyrészt bizonyos szavak a munkaerőpiac révén elérhető egzisztenciát szimbolizálják (dolgozik és munka kifejezések), másrészt megjelenik a jóléti hozzájárulás területe is (jog, segély és szociális kifejezések).

A YouTube megjegyzések kapcsán arra is kíváncsiak voltunk, hogy milyen megítélés alá esik a feltétel nélküli alapjövedelem, illetve jellegzetes értékmintákat igyekeztünk kirajzolni, melyek az elemzett online felület felhasználóinak viszonyulását fejezi ki ezen eszközzel kapcsolatban. A kutatás eredményének tekinthető, hogy az alapjövedelemre jellemző pozitív tartalmú kifejezéseket öt csoportba tudtuk besorolni: megvalósíthatóság, befogadás, igazságosság, létjogosultság és szabadság elnevezésekkel illettük ezeket a csoportokat. A feltétel nélküli alapjövedelem negatív jegyeit szintén öt csoportba tudtuk besorolni: megfizethetetlen, kirekesztés, igazságtalanság, teljesítményelvűség és másodrangú lét. A negatív csoport közel kétszerese volt a pozitív tartalmú megközelítéseknek. Ezek a csoportok a finanszírozhatóság, univerzalitás, feltétel nélküliség, normarendszer és szubjektív jólét egy-egy dimenzióját alkották.

A finanszírozás központi kérdéskörnek tekinthető a feltétel nélküli alapjövedelem kapcsán, mivel mind a pozitív, mind a negatív értékek jelentős arányban merültek fel a megjegyzésekben. Ennek a kategóriának a felértékelődése az egzisztencia fontosságát mutatja az emberek életében.

A feltétel nélküli alapjövedelem munkavállalási motivációra gyakorolt hatását nemcsak a szakirodalmak, de a kommentelők is központi helyre helyezik. A legnagyobb arányban a hozzászólókra a teljesítményelvűség hangsúlyozása jellemző. A normarendszer megváltozásának két végpontja között a legnagyobb a szakadék a többi vizsgált kategóriához képest. A teljesítményelvűség hangsúlyossága azzal magyarázható, hogy a munka általi elismerés elsődleges szerepet játszik életünkben.

Az alapjövedelem harmadik legsarkalatosabb pontja az univerzalitás elfogadottsága, mivel ezen a területen a támogatók aránya eltörpül az ellenzőkkel szemben. A megjegyzésekben erőteljesebben van jelen az univerzalitástól való elzárkózás, bizonyos emberek kirekesztése, mint a mindenkit befogadó attitűd.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A feltétel nélküliség és a szubjektív jólét két-két dimenziója között nem érzékelhetünk jelentős különbségeket, ennek az lehet az oka, hogy ezek a típusú érvek és ellenérvek másodlagos kérdésként merülnek fel az emberek gondolkodásában, amelyek gyökerei a normarendszer és az univerzalitás terén fogalmazódnak meg.

Jövőbeli kutatások kiindulópontját képezhetik a jelen kutatás során összeállított osztályozási kategóriák. Az eszköz használatával a Big Data-alapú elemzések, így a későbbiekben erőteljesebben támaszkodhatnak a gépi tanulás eszközeire.

Irodalom

- Artner A. (2014): A feltétel nélküli alapjövedelem relevanciája és kérdőjelei. *Eszmélet*, 102: 109–130.
- Baksay G. – Bókay M. – Palotai D. – Szalai Á. (2017): *A feltétel nélküli alapjövedelem árnyoldalai*. Magyar Nemzeti Bank Kiadványa
- Beck, U. (2010): *A munka szép új világa*. Belvedere Meridionale Kiadó, Szeged
- Ben-David, A. (2020): Counter-archiving Facebook. *European Journal of Communication*, 35(3): 249–264. oldal, <https://doi.org/10.1177/0267323120922069>
- Bergmann, B. R. (2004): A swedish-style welfare state or Basic Income: Which should have priority? *Politics & Society* 32: 107–118. oldal, <https://doi.org/10.1177/0032329203261101>
- Castel, R. (2005a): A társadalmi biztonság elvesztése. Mit jelent védetten élni? *Esély*, 4: 42-65.
- Castel, R. (2005b): A társadalmi biztonság elvesztése. Mit jelent védetten élni? *Esély*, 6: 3-22.
- Çoban, Ö. – Özel, S. A. – İnan, A. (2021): Deep Learning-based Sentiment Analysis of Facebook Data: The Case of Turkish Users. *The Computer Journal*, 64(3): 473-499., <https://doi.org/10.1093/comjnl/bxaa172>
- Csoba J. (2011): *Munkaerő-piaci változások, leszakadó társadalmi csoportok*. Debreceni Egyetem Szociológia és Szociálpolitika Tanszék, Debrecen
- Dessewffy T. – Mezei M. – Naszályi N. (2018): Harry Potter, avagy a politikai bölcsek köve? Populáris kultúra és politikai aktivizmus. *Politikatudományi Szemle*, XXVII/4: 105–130. oldal
- Dubovi, I. – Tabak, I. (2021): Interactions between emotional and cognitive engagement with science on YouTube. *Public Understanding of Science*, 1-18., <https://doi.org/10.1177/0963662521990848>
- Fromm, E. (2002): *Menekülés a szabadság elől*. Napvilág Kiadó, Budapest
- Gilbert, R. – Murphy, N. A. – Stepka, A. – Barrett, M. – Worku, D. (2018): Would a basic income guarantee reduce the motivation to work? An analysis of labor respon-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- ses in 16 trial programs. *Basic Income Studies*, 13(2): 1–12. oldal, <https://doi.org/10.1515/bis-2018-0011>
- Graeber, D. (2018): *Bullshit Jobs: A Theory*. Simon & Schuster, New York
- Hayek, F. (1978): *Law, Legislation and Liberty. Volume 2: The Mirage of Social Justice*. University of Chicago Press, Chicago
- Sachs, J. (2015): Investing in Social Capital. In: Helliwell, J. – Layard, R. – Sachs, J. (szerk): *World Happiness Report 2015*. Sustainable Development Solutions Network, New York, 152–167.
- Hirsch, D. (2015): *Could a 'citizen's income' work?* Joseph Rowntree Foundation Programme Paper, York
- Hirschberg, J. – Manning, C. D. (2015): Advances in natural language processing. *Science*, 349(6245): 261–266. oldal, DOI: 10.1126/science.aaa8685
- Honneth, A. (2009): Munka és elismerés. Kísérlet egy újrafogalmazásra. *Replika*, 68: 125–140.
- Hradil, S. (1995): Régi fogalmak és új struktúrák. Milió-, szubkultúra- és életstílus-kutatás a 80-as években. In: Andorka R. – Hradil, S. – Peschar, J. L. (szerk.): *Társadalmi rétegződés*. Aula, Budapest, 347–385.
- Inglehart, R. – Flanagan, S. C. (1987): Value Change in Industrial Societies. *The American Political Science Review*, 81: 1289–1319. oldal, <https://doi.org/10.2307/1962590>
- Kapitány Á. – Kapitány G. (2014): Alternatív életstratégiák kulcselemei. *Magyar Lettre Internationale*, 98: 64–70.
- Keller T. (2009): *Magyarország helye a világ értéktérképén*. Társ, Budapest
- Keynes, J. M. (1965): *A foglalkoztatás, a kamat és a pénz általános elmélete*. KJK, Budapest
- Khan, M. L. (2017) Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66: 236–247. <https://doi.org/10.1016/j.chb.2016.09.024>
- Kitzinger D. (2000): A morális pánik elmélete. *Replika*, 40: 23–48. oldal
- Lucas, M. – Gunawardena, C. – Moreira, A. (2014): Assessing social construction of knowledge online: A critique of the interaction analysis model. *Computers in Human Behavior*, 30: 574–582. oldal, <https://doi.org/10.1016/j.chb.2013.07.050>
- Mészáros E. – Sebők M. (2018): A szövegbányászati módszerek alkalmazásának lehetőségei a joggyakorlat-elemzésben. *Forum Sententiarum Curiae*, 2: 6-12.
- Németh R. – Katona E. R. – Kmetty Z. (2020): Az automatizált szövegelemzés perspektívája a társadalomtudományokban. *Szociológia Szemle*, 30(1): 44–62.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Örkény A. – Székelyi M. (2010): Az igazságosság labirintusaiban. *Szociológia Szemle*, 20(2): 4–41.
- Örkény A. – Székelyi M. (2011): A nagy elosztási rendszerekről vallott nézetek generációs vonatkozásai. In: Örkény A. – Székelyi M. (szerk.): *Az igazságosság labirintusaiban. Társadalmi méltányosság és generációs igazságosság a 21. század Magyarországon*. Sik Kiadó, Budapest, 91–105.
- Piachaud, D. (2016): *Citizen's Income: rights and wrongs*. Centre for Analysis of Social Exclusion. London School of Economics, London
- Policy Agenda (2018): Félelmek a munka világában.
- Reilly, P. (2014): The 'Battle of Stokes Croft' on YouTube: The development of an ethical stance for the study of online comments. *Sage Research Methods Cases Part 1*. <https://dx.doi.org/10.4135/978144627305013509209>
- Salganick, M. J. (2017): *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, Princeton
- Schopenhauer, A. (1992): *A nemi szerelem metafizikája. Az élethez való akarat igencéléről. Az élet semmiségéről es győtrelméről*. Reprint: Hatágú Síp Alapítvány, Budapest
- Szabó M. K. (2014): *Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai*. In: Nyelv, kultúra, társadalom konferencia, Budapest
- Szelényi I. (2014): Variációk a szegénység témára. In: Kolosi T. – Tóth I. Gy. (szerk.): *Társadalmi Riport 2014*. Táarki, Budapest, 629–635.
- Tang, L. – Fujimoto, K. – Amith, M. T. – Cunningham, R. – Costantini, R. A. – York, F. – Tao, C. (2021): "Down the Rabbit Hole" of Vaccine Misinformation on YouTube: Network Exposure Study. *Journal of Medical Internet Research*, 23(1): 1–9., doi:10.2196/23262
- Tanner, M. (2015): The pros and cons of a guaranteed national income. *Cato Institute Policy Analysis*, 773: 1–35.
- Thelwall, M. (2017): The heart and soul of the Web? Sentiment strength detection in the social Web with sentistrength. In: Holyst, J. (szerk.): *Cyberemotions: Collective Emotions in Cyberspace. Understanding Complex Systems*. Springer, Berlin, 119–134.
- Tomka B. (2008): *A jóléti állam Európában és Magyarországon*. Corvina Kiadó, Budapest
- Van Parijs, P. (2010): Alapjövedelem: egy egyszerű és erőteljes gondolat a huszonegyedik század számára. *Esély*, 5: 9–41.
- Weber, M. (1982): *A protestáns etika és a kapitalizmus szelleme*. Gondolat Kiadó, Budapest



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Zsibrita, J. – Vincze V. – Farkas R. (2013): magyarul: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: Proceedings of RANLP 2013, 763–771.



Mi a téma? Politikai témák és felhasználói reakciók szövegbányászati vizsgálata a 2018-as országgyűlési kampány politikusi Facebook-oldalain¹

BENE MÁRTON²

ABSZTRAKT

A tanulmány szövegbányászati eszközökkel vizsgálja, hogy a felhasználók hogyan reagálnak a 2018-as magyarországi országgyűlési választás kampányának vezetői témáira a jelöltek Facebook-oldalain. Azt feltételezi, hogy a kiemelt (migráció, korrupció) és a kampánnyal kapcsolatos posztok több reakciót váltanak ki, míg a közpolitikai témák és a mobilizációs tartalmak kevésbé lesznek népszerűek. Mindezek túlmenően a témagazda-hatás elméletet is teszteli a felhasználói reaktivitás tekintetében. E feltevéseket egy olyan adatbázison teszteli a kutatás, amely a választás minden mérhető támogatottságú jelöltjének a kampány alatt közzétett összes posztját tartalmazza (511 jelölt 38030 posztja), a témákat pedig szövegbányászati eszközökkel azonosítja, ily módon használva ki a közösségi médiában rejlő „big data”-elemzés lehetőségét. Az eredmények azt mutatják, hogy a korrupció, a fejlesztéspolitika és a kampány a legnépszerűbb témák a Facebookon, míg a migráció csak az ellenzéki politikusok oldalán volt vonzó, a kormánypárti politikusok követői kerültek a migrációs témájú posztokra való reagálását. A legérdekesebb eredmény a fordított témagazda-hatáshoz kapcsolódik, miszerint a politikusok saját témáikkal általában kevésbé tudnak reakciókat kiváltani, mint ellenfeleik ugyanezekben a témákban.

KULCSSZAVAK: közösségi média, témagazda-hatás, migráció, felhasználói reakciók, szövegbányászat

ABSTRACT

What's the matter? A text mining analysis of political topics and user engagement on politicians' Facebook pages during the 2018 Hungarian general election campaign

The research investigates the way users interact with leading topics of the 2018 Hungarian general election campaign on candidates' Facebook pages. It expects that the prominent (immigration, corruption) and campaign-related topics generate more user engagement, while

¹ A tanulmány az Innovációs és Technológiai Minisztérium ÚNKP-21-5 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült (ÚNKP-21-5-ELTE-1094). A tanulmány angol nyelvű változata a Journal of Information Technology & Politics folyóiratban jelent meg.

² Társadalomtudományi Kutatóközpont, ELTE ÁJK, e-mail: Bene.Marton@tk.hu



TEMATIKUS TANULMÁNYOK – Digitális szociológia

policy topics and mobilization content are less interacted. It also tests the theory of issue ownership in relation with user engagement. These expectations are tested on a dataset that includes all posts (38030 posts) posted by all candidates during the campaign (511 candidates). Topics are identified by text mining methods. The study demonstrates that corruption, development policy and campaign are highly engaged topics, while immigration was more interacted only on opposition politicians' pages since the followers of pro-government candidates engage less with immigration-related content. The most surprising result is that a reversed issue ownership effect can be detected since politicians are generally less successful with their own topics.

KEYWORDS: social media, issue ownership, immigration, user engagement, text mining

Bevezetés

Az elmúlt néhány év fontos fejleménye, hogy az állampolgárok online kommunikációs aktivitása a politikai kommunikációs ökoszisztéma ütőerévé vált (Blumler 2016). Míg korábban a politikai szereplők kommunikációjukat az állampolgárok kognitív és affektív szükségleteire szabták, ma már az emberek egymás közötti politikai társalgásait is célba veszik. Mivel az állampolgárok közösségi média aktivitása képes a politikai üzeneteket széles körben láthatóvá tenni, ezért a politikusok kommunikációja mindinkább alkalmazkodik a felhasználókéhoz, e folyamat pedig a politika 'viralizációjának' fogalmával ragadható meg (Bene 2020). Emiatt fontos megértenünk az állampolgárok politikai tartalmakkal folytatott interakcióit a közösségi médiában, mivel e mintázatok jelentős hatással vannak a politikai szereplők kommunikációjára (Ennser-Jedenastik et al. 2021).

Egyre kiterjedtebb szakirodalom foglalkozik azzal, hogy a közösségi média felhasználók a politikai szereplők milyen típusú tartalmaira reagálnak (Bene 2020, Heiss et al 2019, Keller – Kleinen-von Königsłow 2018, Štětka et al 2019, Xenos et al. 2017). E kutatás azonban a szakirodalom két még mindig meglévő hiányosságát próbálja pótolni a 2018-as magyarországi országgyűlési kampány vizsgálatán keresztül. Először is, még mindig nagyon korlátozott a tudásunk arról, hogy milyen típusú politikai tématerületek váltanak ki felhasználói reakciókat. Tekintve, hogy a politikai kommunikáció-kutatás milyen jelentőséget tulajdonított korábban a politikai témák nyilvánosságban való láthatóságának (lásd, Baumgartner – Jones 1993; Iyengar et al 1982, McCombs – Shaw 1972), ez a vonatkozó irodalom egy meglepő hiányossága. A kutatás elsődleges fókusza ezért arra irányul, hogy a kampány legfontosabb témái a jelöltek Facebook-oldalain milyen mértékben tudtak felhasználói reakciókat kiváltani. Azt feltételezi, hogy a leginkább kiemelt (migráció, korrupció), illetve a kampányhoz szorosan kapcsolódó témák a legsikeresebbek ebben a tekintetben, miközben a hagyományos közpolitikai ügyek, illetve a mobilizációs kísérletek kevésbé népszerűek. Mivel az egyes témák általában erősen kötődnek adott politikai szereplőkhöz, ezért a kutatás azt is megvizsgálja, hogy a témagazda-elmélet (issue



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ownership theory) (Petrocik 1996) mennyiben képes magyarázni a felhasználói reaktivitás mintázatait.

A második hiányosság abból fakad, hogy a létező kutatások nem használják ki a közösségi oldalak „big data”-potenciálját. A vonatkozó vizsgálatok rendszerint manuálisan gyűjtött adatokra támaszkodnak, ez pedig mintavételi döntéseket igényel, ami gyengíti az eredmények külső érvényességét, általánosíthatóságát. E kutatás szótár- és felügyelt gépi tanulási módszerekre támaszkodó szövegbányászati eszközökkel dolgozza fel az adatokat, így a Facebook tartalmak teljes populációja figyelembe vehető. Ezért a hipotéziseket egy egyedülállóan kiterjedt adatbázison teszteli, amely 511 jelölt a kampány hét hete alatt közzétett több mint 38 030 ezer posztját tartalmazza.

Elméleti keret

A POLITIKAI KOMMUNIKÁCIÓ „VIRALIZÁCIÓJA”

A politikai szereplők kommunikációs erőfeszítéseinek célja a szavazók politikai attitűdjeire és magatartására való hatásgyakorlás. A hatás eléréséhez azonban a kommunikációnak csatornákra van szüksége. A 20. század végéig a politikai szereplők könnyen hozzáfértek a kommunikációs csatornákhöz, egyrészt amiatt, mert a tömegmédiá a politikát a nyilvános élet egy kiemelt, önjogán tudósításra érdemes szférájaként kezelte („sacerdotal approaches to politics”, lásd, Blumler 1965:98-105), másrészt pedig azért mert a pártok maguk is nagyszámú csatornát birtokoltak vagy kontrolláltak (újságok, aktivista hálózatok stb.). A politikai kommunikáció a 20. század végén kezdődő, úgynevezett „harmadik korszakától” (Blumler – Kavanagh 1999) kezdve azonban a csatornákhöz való hozzáférés egyre nagyobb kihívássá vált a politikai szereplők számára, ugyanis a sokcsatornás médiatérben a közönség egyre inkább fragmentálódott, és a mindinkább kereskedelmi logika által irányított tömegmédiában a politika elvesztette kiemelt szerepét.

Az információbőség e kontextusában a politikusoknak meg kell küzdeniük a korábban magától értetődő figyelemért egymással, illetve minden egyéb, akár teljesen más területen tevékenykedő tartalomtermelővel (Coleman – Blumler 2009). A politikai szereplők nézőpontjából ez azt jelenti, hogy politikai üzeneteikkel már nem csak a választókat kell megcélzniuk, hanem az üzenetet közvetítendő csatornák figyelmét is meg kell, hogy ragadják kommunikációjukkal.

A célpont megkettőződése a politikai kommunikáció működésére is hatással van. A szakirodalom már kiterjedten dokumentálta, hogy a politikai kommunikáció hogyan változott meg amiatt, hogy a politikai szereplők üzeneteiket egyre inkább a professzionális média kereskedelmi logikájához igazítják annak érdekében, hogy



TEMATIKUS TANULMÁNYOK – Digitális szociológia

láthatóságukat növeljék (Lundby 2009). A kutatók e folyamatot a „politika mediatiszációjának” nevezték el (Strömbäck 2008).

A közösségi oldalak, különösen az utóbbi években kiemelt politikai információforrássá váló Facebook, a politikai szereplők számára alternatív csatornákat kínálnak a választók elérésére. A Facebookon a tartalomterjedés domináns logikája a viralitás (Klinger – Svensson 2015), hiszen a felhasználók nyilvánosan látható interakciói a posztokkal (Reakciók,³ kommentek és megosztások) képesek az eredeti üzenetet széles körben elterjeszteni. Ezek az interakciók egyfelől megjelenhetnek az ismerősök hírvonalán, így azok is láthatják az adott politikai szereplő posztját, akik őt egyébként nem követik. Ráadásul ily módon e tartalmakat saját ismerőseik közvetítésén és ajánlásán keresztül érzékelik, ami növelheti azok hitelességének (Turcotte et al. 2015) vagy relevanciájának érzetét (Anspach 2017). Másfelől, a hírvonalakon megjelenő tartalmakat algoritmusok szűrik, és bár a Facebook titokban tartja az algoritmus működésének pontos szempontrendszerét, az közismert, hogy a több interaktivitást kiváltó tartalmak nagyobb valószínűséggel jelennek meg a felhasználók előtt, mint a kevésbé reagált posztok (Bucher 2012).

Következésképpen a Facebookon az állampolgárok, illetve az ő reakcióik szolgáltatják azt a csatornát, amelyen keresztül az üzenet eljuttatható a választók szélesebb rétegeihez. Éppen ezért a politikusok azáltal tudnak csatornát szerezni az üzenetükhez, ha azokkal egyúttal reakcióra is bírják a felhasználókat (Bene 2020). Kutatók bizonyították, hogy a politikusok közösségi média kommunikációjának valóban egyik legfontosabb célja a felhasználói reakciók kiváltása (Bene – Somodi 2018, Kelm 2020). Ezért aktívan monitorozzák a felhasználók interakciós igényeit (McGregor 2020), és igyekeznek ezekhez igazítani kommunikációjukat (Barberá et al. 2019, Ennsner-Jedenastik et al. 2021). Ráadásul azt is tudjuk, hogy a felhasználói reakciók kiváltása hatással van a választási teljesítményre (Bene 2020), illetve a professzionális médiában való láthatóságra is (Kruikemeier et al. 2018). Ahogyan azonban a professzionális média reakcióra való hatásgyakorlás célja átalakította a politikai kommunikáció formáját és tartalmát, úgy az is okkal feltételezhető, hogy amikor egy másfajta logikához, az állampolgári kommunikáció logikájához igazodnak a politikai szereplők, akkor az is jelentős változásokat eredményez. E folyamatra, mely során a politikusok a közösségi média virális terjedési logikájához igazodnak, a mediatiszáció analógiájára a politikai kommunikáció 'viralizációjaként' hivatkozhatunk (Bene 2020). Ezért érdemes feltárni, hogy miben is áll ez az átalakulás.

Elméleti szempontból legalább két olyan érv is felhozható, amelyek miatt a politika viralizációja jelentős változásokat eredményezhet a politikai kommunikáció működésében. Először is, a politikai kommunikáció célja ezidáig alapvetően a vá-

³ Az általános kommunikációs válaszként felfogott reakció, és a Facebookon a lájkolást is magába foglaló, de további hangulati válaszelemekkel kiegészített reakciógombok közötti különbségtétel érdekében utóbbiakat nagybetűvel írom a szövegben.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

lasztók kognitív és affektív kapacitásaira való hatásgyakorlás volt, az állampolgárokat elsődlegesen „nézőközönségként” („spectators”) kezelték, akiket meggyőzni, lenyűgözni, szórakoztatni és érzelmileg felfokozott állapotban tartani kell (Coleman – Blumler 2009). Ezzel szemben a közösségi médiában a hatékony kommunikációnak az állampolgárok – a politikai kommunikáció tömegmédiá-központú működése által nagyrészt figyelmen kívül hagyott – kommunikatív és társas természetét is célba kell vennie, és a nézőközönséget résztvevőkké kell változtatnia (Kreiss 2016: 5). Másodszor, okkal feltételezhető, hogy az állampolgárok máshogy reagálnak a politikai tartalmakra, mint a professzionális médiatermékek. A média kommunikációját kereskedelmi érdekek, professzionális normák, vagy éppen politikai szempontok irányítják (lásd: Strömbäck – Esser 2014), ezért e faktorok határozzák meg, hogy az újságírók milyen politikai üzenetekre reagálnak. Ezzel szemben a választók politikai kommunikációját a közösségi oldalakon elsődlegesen expresszív motivációk vezérlik, ami azt jelenti, hogy kommunikációjuk során személyes identitásukat jelenítik meg és formálják erős és gyenge kapcsolataik előtt (Bennett – Segerber 2013, Svensson 2011). Az eltérő motivációk feltehetően különböző kommunikációs mintázatokat eredményeznek, ezért várhatóan különböző politikai tartalmak is illeszkednek ezekhez. Annak megállapításához, hogy ez a gyakorlatban mit is jelent, empirikus vizsgálatokra van szükség.

Felhasználói reaktivitás a politikai szereplők Facebook-oldalain

Csak a legutóbbi években kezdték el részletesebben vizsgálni, hogy az állampolgárok milyen tartalmakkal lépnek interakcióba a politikai szereplők oldalain (lásd: Bene 2020, Heiss et al. 2019, Keller – Kleinein-von Königsłow 2018, Xenos et al. 2017). Bár az eredmények meglehetősen vegyesek, a meglévő kutatások számos tartalmi faktor állampolgári reakciókra gyakorolt hatását vizsgálták. A növekvő számú kutatás ellenére két lényeges hiányosság megmaradt. Az első hiányosság tartalmi jellegű. Miközben a posztok formájának, retorikájának, érzelmi és stilisztikai jegyeinek valamint politikai keretezésének hatásairól egyre többet tudunk, a téma kérdése jóval kisebb figyelmet kapott. Miközben az eredmények megmutatták, hogy a szöveges (Bene 2020), moralizáló (Sahly et al. 2019), negatív (Bene 2020, Heiss et al. 2019), meglepő (Staender et al. 2019) és a páthosz retorikai eszközét (Gerodimos – Justinussen 2015, Heiss et al. 2019) alkalmazó posztok jobban terjednek a Facebookon, jóval korlátozottabb a tudásunk arról, hogy milyen politikai témákkal lépnek leginkább interakcióba a felhasználók.

A téma dimenziójának figyelmen kívül hagyása azért is meglepő, mert a politikai kommunikáció-kutatás a múltban nagy jelentőséget tulajdonított annak, hogy milyen topikok terjednek jól a mediatizált kommunikációs szférában. Számos médiahatás elmélet mellett érvelt ugyanis, hogy a tömegmédiá azáltal tudja alakítani



TEMATIKUS TANULMÁNYOK – Digitális szociológia

a politikai folyamatokat, hogy egyes témákra erőteljesebben fókuszál, míg másokat figyelmen kívül hagy. A napirend-elmélet szerint a tömegmédiában erősen jelenlévő témák lesznek azok, amelyeket a választók fontosnak tartanak (McCombs – Shaw 1972). A priming-elmélet is azt hangsúlyozza, hogy a tömegmédiában a témákon keresztül hat a választóra, azonban ez a megközelítés az állampolgárok politikai értékelési folyamataira fókuszál, és úgy látja, hogy e kiemelt témák ebben válnak meghatározóvá (Iyengar et al. 1982). A megszakított egyensúly elmélet a média közpolitikai dinamikában játszott szerepére koncentrál, és azt állítja, hogy a média egyes témákon való erőteljes fókuszja „közpolitikai ablakokat” (policy window) nyithat, amelyekben keresztül jelentős, a hagyományos közpolitikai dinamikától idegen mértékű közpolitikai változások mehetnek végbe (Baumgarner – Jones 1993, Boda – Sebők 2018). Összefoglalva tehát, ezek az elméletek mind arra mutatnak rá, hogy a média a tematikus agendáján keresztül tudja leginkább formálni a politikai folyamatokat. A közösségi médiában azonban a napirendet főként a felhasználói interaktivitás mintázatai határozzák meg, ezért szükséges azonosítani azokat a témákat, amelyekkel előszeretettel interaktálnak a felhasználók. A tárgyalt elméletek alapján ugyanis e témák lesznek hatással a választók szélesebb tömegeinek politikai viselkedésére és a politikai folyamatokra.

A második hiányosság módszertani természetű. A vonatkozó kutatások manuálisan kódolták adataikat annak érdekében, hogy a politikai szereplők kommunikációjának felhasználói reaktivitására gyakorolt hatásait feltárják. A kézi kódolás magas belső érvényességű adatokat eredményez, ugyanakkor erőforrásigényes volta miatt mintavételi döntéseket is szükségessé tesz, ami korlátozhatja az eredmények általánosíthatóságát (Guo et al. 2016). A legtöbb tanulmány ezért csak kisszámú politikai szereplőre fókuszál, ahol az eredményeket erősen torzíthatják a kiválasztott szereplők vagy követői bázisuk egyéni sajátosságai (lásd: Gerodimos – Justinussen 2015, Sahly et al. 2019, Staender et al. 2019, Štětka et al. 2019). Miközben ezek a tanulmányok alkalmasak arra, hogy azonosítsák egy-egy kulcsszerepet játszó politikai szereplő kommunikációjának lényeges aspektusait és hatásait, a felhasználói reaktivitás általánosabb mintázatairól nem szolgáltatnak tudással. Jóval kevesebb kutatás vizsgálja ezt a kérdést nagyobb mintán, ami lehetővé teszi az eredmények megfigyeléseken túli általánosítását (lásd: Bene 2020, Heiss et al. 2019, Keller – Kleinen-von Königslo 2018, Xenos et al. 2017), de ezekben az esetekben más jellegű mintavételi döntéseket kell meghozni a kutatóknak (például: rövidebb időszak, legsikeresebb posztok), hogy a kézi kódoláson alapuló adatgyűjtés kivitelezhető maradjon. Minden bizonnyal ezek a mintavételi torzítások állnak a gyakran egymásnak is ellentmondó eredmények mögött.

A közösségi oldalak és az egyre fejlettebb és elérhetőbb automatizált szövegfeldolgozási módszerek lehetővé teszik azonban, hogy minden releváns adatot bevonjunk a kutatásba, leküzdve ezáltal a mintavételi döntésekből fakadó korlátokat (Guo et al. 2016, Sebők 2016). Az egész kiterjedt populációt lefedő „big data” teljes



TEMATIKUS TANULMÁNYOK – Digitális szociológia

vizsgálata könnyebben általánosítható eredményeket produkálhat. Természetesen számos olyan szempont is megjelenik az irodalomban, melyeket meglehetősen nehéz szövegbányászati módszerekkel megragadni (például retorika, stílus), így ezek magas érvényességű vizsgálata továbbra is kézi kódolást igényel. A politikai témák szókinccse azonban rendszerint erősebben elválik egymástól, ezért a szótár alapú és felügyelt gépi tanuláson alapuló módszerek alkalmasak lehetnek ezek azonosítására. E kutatás ezért a mintavételi döntésekből fakadó torzítások kiküszöbölésére törekszik azáltal, hogy a kutatási kérdést egy kiterjedt, a politikai aktorok és posztjaik teljes populációját lefedő adatbázis automatizált feldolgozásán keresztül kísérli megválaszolni.

Mindezek alapján e kutatás kérdése a következő: Milyen politikai témák váltanak ki felhasználói reakciókat a politikai szereplők Facebook-oldalain?

Hipotézisek

A kampány során számtalan téma bukkan fel a politikusok kommunikációjában. E kutatás a 2018-as magyarországi választási kampány legfontosabb témáinak a reaktivitásra gyakorolt hatásait vizsgálja. A kampányokban rendszerint nagy hangsúlyt kapnak bizonyos, közérdeklődésre számot tartó ügyek. A kutatás ezek közül a kampány két legkiemeltebb ügyének, a migrációnak és a korrupciónak, illetve a három leginkább megjelenített hagyományos közpolitikai témájának, az egészségügynek, az oktatásnak és a fejlesztéspolitikának a hatásait vizsgálja. Mivel ezek a témák általában erősen kötődnek bizonyos politikai szereplőkhöz, ezért a kutatás azt is teszteli, hogy a témagazda-hipotézis érvényes-e a felhasználói reaktivitás esetében. E specifikusabb ügyeken kívül mivel maga a kampány, annak eseményei és az ahhoz kapcsolódó mobilizációs erőfeszítések minden választás előtt kiemelt témák, ezért ezek felhasználói interakciókra gyakorolt hatásait is vizsgálja a kutatás.

A 2018-as kampányt két téma dominálta elsődlegesen: a migráció (Bíró-Nagy 2018) és a korrupció (Kmetty 2018). A migráció a kormánypártok elsődleges témája volt, míg a korrupció az ellenzéki pártok kommunikációjában foglalt el kitüntetett helyet. A két téma leginkább szembetűnő közös vonása, hogy jelentőségük nem korlátozódik a kampányidőszakra. A migráció már 2015 óta a kormánypártok legfontosabb mobilizáló témája, és számos kampányt folytattak azóta (például nemzeti konzultáció, népszavazás, plakátkampányok) a migráció ügyére fókuszálva. Hasonlóképpen, a kormányzati szereplőkkel szembeni korrupciós vádak már 2010 óta visszatérőek az ellenzéki oldalon, és több mobilizációs kezdeményezés fűződik ehhez a témához is, úgy mint népszavazási aláírásgyűjtések, tüntetések, petíciók, plakátkampányok és hasonlók. E két téma természetesen nagy népszerűségnek örvend más kampánykörnyezetben is (lásd: Dolezal – Zeglovits 2014).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Stratégiai nézőpontból lényeges kérdés, hogy a politikai szereplők által erősen nyomott témák hogyan tudnak terjedni a közösségi oldalakon. A professzionális hírmédia általában jelentős teret enged a politikai elitek kommunikációjában kiemelt témáknak (Bennett 1990). Azt azonban nem tudjuk, hogy a felhasználók is támogatják-e a politikai szereplők stratégiai elképzeléseit azzal, hogy reakcióikon keresztül felerősítik az általuk erőteljesen nyomott témák láthatóságát, vagy épp-hogy akadályozzák ezt azáltal, hogy figyelmen kívül hagyják az ezekre az ügyekre fókuszáló tartalmakat. Két érv szól amellett, hogy a választók reaktívak e két kiemelt ügy kapcsán a Facebookon. Először is, empirikus kutatások igazolták, hogy a migráció és a korrupció ügyének nyilvánosságbeli markáns jelenléte jelentősen befolyásolta az állampolgárok attitűdjeit és percepcióit (Bíró-Nagy 2018, Kmetty 2018), tehát a pártok általi kiemelésük erősen hatott a választók kognitív és affektív kapacitásaira. Ha az emberek azokra a posztokra reagálnak, amelyek kognitív és affektív szempontból hatással vannak rájuk (Eberl et al. 2020), akkor azt várhatjuk, hogy e témák több reakciót fognak kiváltani az oldalon. A második érv e témák sajátosságaihoz kötődik. Napjaink politikájában egyfajta populista fellendülés („populist zeitgeist” – Mudde 2004) figyelhető meg, és az emberek ma különösen fogékonyak a korrupció és migráció témákra, amelyeket még a nem populista szereplők is gyakrabban értelmeznek populista módon (Ernst et al. 2019).

H1. A migrációval és a korrupcióval foglalkozó posztok több felhasználói reakciót váltanak ki, mint más posztok.

Tradicionalis közpolitikai területek is fontos témái a kampányoknak: a politikai szereplők rendszerint közpolitikai ígéreteket tesznek, illetve riválisaik közpolitikai teljesítményét és terveit bírálják. A legfontosabb közpolitikai területeket a pártok által a kampányban tett ígérek alapján azonosítja a kutatás. Dobos és munkatársai (2018) összegyűjtötték a kampány alatt tett összes írott vagy szóban tett választási ígéretet, és azt találták, hogy a legtöbb ígéret az egészségügy (összes ígéret 8,6%-a), az iskolázottság (7,1%) és a fejlesztéspolitika (6,8%) területére irányult. Természetesen a politikai táborok között jelentős eltérések vannak: az ellenzéki pártok az egészségügyre és az oktatásra fókuszáltak, míg a kormánypártok első számú közpolitikai témája a fejlesztéspolitika volt.

Kutatások azt is igazolták, hogy hagyományos közpolitikai témák és érvelések kevésbé vannak jelen a politikai szereplők közösségi média kommunikációjában (Gerodimos – Justinussen 2015, Štětka et al. 2019). Miközben a közpolitikai tartalmak nyilvános kommunikációban való visszaszorulása általánosabban is jellemző napjaink politikai kommunikációjára (Coleman – Blumler 2009), a professzionális média mégis viszonylag jelentős figyelmet szentel a legfontosabb közpolitikai történéseknek. Azt feltételezhetjük, hogy a közösségi média kommunikáció esetében nem csak a közpolitikai tartalmak kínálata csökken, hanem az ilyen jellegű posztokra való igény is. A közösségi médiában zajló aktivitást jelentős részben a felhasználók expresszív motivációi vezérlik (Bennett – Segerberg 2013), hiszen minden nyilvános



TEMATIKUS TANULMÁNYOK – Digitális szociológia

megnyilatkozásukkal saját magukat jelenítik meg teljes kiterjedt társadalmi hálózatuk előtt. Emiatt ezekre az aktivitásokra identitás-performanszként tekinthetünk, amelyeken keresztül az egyének alakítják mások róluk alkotott benyomásait (Goffman 1959, Svensson 2011). A felhasználók ezért olyan közösségi média tartalmakkal lépnek elsősorban kapcsolatba, amelyek segítenek valamit megjeleníteni saját személyes identitásukból (Ben 2020, Bennett – Segerberg 2013). Okkal érvelhetünk amellett, hogy személyes identitást kifejezni minden bizonnyal jóval nehezebb közpolitikai tartalmakon keresztül, hiszen ezek általában túl specifikusak, összetettek és bonyolultak ahhoz, hogy az egyén identitásának legfontosabb aspektusairól, értékeiről és moralitásáról üzeneteket közvetítsenek. Ráadásul a közpolitikai tartalmak jelentése általában zárt, avagy előre világosan meghatározott. Ezzel szemben a közösségi média kommunikációban a nyitottabb, sokféleképpen értelmezhető, különböző identitások kifejezésére is alkalmas tartalmak a népszerűbbek (Bennett – Segerberg 2013). Mindezen okok miatt a kutatás vonatkozó hipotézise:

H2. A három kiemelt közpolitikai témával, azaz az egészségüggyel, oktatással és fejlesztéspolitikával foglalkozó posztok kevesebb felhasználói reakciót váltanak ki a politikusok Facebook-oldalain.

Az egyes ügyek azonban erősen kötődnek bizonyos pártokhoz. A témagazda-elmélet azt képviseli, hogy a pártok birtokolnak egyes témákat, amelyekben kompetensebbnek és hitelesebbnek látszanak a választók szemében (Petrocik 1996). Mivel a pártok profitálhatnak abból, ha olyan témákkal kötődnek össze, amelyek a választók számára fontosak (Ansolabehere – Iyengar 1994), a kampányban jelentős erőfeszítéseket is tesznek azért, hogy bizonyos ügyek hozzájuk kapcsolódjanak, és ezeket a választók számára is fontossá tegyék (Petrocik 1996). Ahogy arról fentebb már volt szó, a vizsgált magyar esetben is voltak témagazdák: a migráció és a fejlesztéspolitika a kormánypártokkal, míg a korrupció, az oktatás és az egészségügy az ellenzéki oldallal kötődött össze. Azt azonban nem tudjuk, hogy a politikusok követői is valóban e témákra reagálnak legszívesebben. E kérdés megválaszolásához az témagazda-elméletet a felhasználói reaktivitás vizsgálatára adoptálva a kutatás azt feltételezi:

H3. Az ügygazdák több interakciót váltanak ki saját témájukkal, mint a témát nem birtokló szereplők. Tehát a kormánypárti jelöltek migrációval és fejlesztéspolitikával, míg az ellenzéki pártok korrupcióval, oktatással és egészségüggyel kapcsolatos posztjai több felhasználói reakciót váltanak ki.

A kampány azonban nagyrészt saját magáról szól. Az üzenetek jelentős része a politikai versengéssel, a különböző kampányeseményekkel, a szavazók mobilizálásával vagy az ellenfelek támadásával foglalkozik. A választási kampányokról egyre inkább egyfajta „lóversenyként” tudósít a professzionális média, és a stratégiai, nem közpolitikai jellegű aspektusok kapnak nagyobb hangsúlyt (Banducci – Hanretty 2014). Ezáltal a választók is egyre inkább a kampányeseményekre figyelnek. Ráadásul, a kampányeseményekre fókuszáló tartalmakat jóval egyszerűbb értelmezni,



TEMATIKUS TANULMÁNYOK – Digitális szociológia

megérteni és véleményt formálni róluk, így könnyen lehet kapcsolódni hozzájuk. Emiatt az a kutatás várapozása, hogy a felhasználói reaktivitás csak még inkább felerősíti a politika „lóverseny”-jellegét, azáltal, hogy a kampány-üzeneteket teszi még szélesebb körben láthatóvá.

H4. A kampánytémákkal foglalkozó posztok több felhasználói reakciót váltanak ki.

A kampánytartalmak egy kiemelt altípusa a mobilizációs üzenetek, amelyek célja, hogy a felhasználók online figyelmét a politikai szereplők offline cselekvéssé (szavazás, offline eseményen részvétel) változtassák. Az ilyen mobilizációs üzenetekkel való interakció, különösen ezek megosztása, egyfajta politikai aktivizmusként is felfogható, hiszen ilyenkor a felhasználó célja minden bizonnyal saját ismerősei aktivitásra sarkallása. A politikai szereplőknek egyre nagyobb szükségük van az ilyen jellegű alacsony erőforrás-igényű digitális aktivizmusra, hiszen a hagyományos aktivistabázisok mérete jelentősen csökkent az elmúlt évtizedekben. Ezért érdemes vizsgálni, hogy a közösségi média az offline mobilizáció alkalmas terepének számít-e. A vonatkozó empirikus vizsgálatok eredményei azonban ellentmondásosak (lásd: Heiss et al. 2019, Štětka et al. 2019). Ugyanakkor, ahogy arról korábban már szó volt, a felhasználói aktivitás elsődlegesen expresszív motivációkkal magyarázható, míg az offline cselekvésre való felszólítást tartalmazó poszttal való interakció inkább instrumentális motivációkkal, az ismerősök politikai befolyásolásának szándékával, írható le. E kutatás azt feltételezi, hogy a felhasználói kommunikációt ezen a platformon nem politikai célok vezérlik, így a követők jelentős része nem szándékozik a kampány aktív résztvevőjévé válni.

H4a. Az offline mobilizációs üzeneteket tartalmazó posztok kevesebb felhasználói interakciót váltanak ki.

Módszertan

ADATOK

Az adatbázis a 2018-as országgyűlési kampányban képviselői helyért induló jelöltek összes Facebook-posztját tartalmazza. Azokat a jelölteket vettem figyelembe, akik vagy legalább a szavazatok 1%-át elérték a 106 egyéni választókeret valamelyikében vagy a 0,5%-nál több szavazatot szerző 8 országos pártlista első 30 helyének egyikén szerepeltek. E jelöltek 83%-a rendelkezett nyilvános Facebook-oldallal (N = 523). A választás másnapján a Facebook API-on⁴ keresztül töltöttem le e jelöltek összes, a kampány hivatalos kezdőnapjától (2018. 02. 17.) a szavazóurnák zárásig

⁴ Az adatokat a Python-ra írt facebook-page-post-scraper csomag segítségével töltöttem le, lásd: <https://github.com/minimaxir/facebook-page-post-scraper/blob/master/README.md>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

(2018. 04. 08., 19.00) közzétett posztját. Ez a metódus összesen 48 883 posztot eredményezett 511 jelölttől⁵ a vizsgált 51 napos időszakból, és mivel módszerünk szövegbányászati eszközökre épül, ezek közül a 38 030 szöveges tartalmat is tartalmazó poszt került a végső adatbázisba. A szöveges adatokat a feldolgozás előtt megtisztítottam, így a számok, a központozás, a ragok, a nagyon általános szavak (például: névelők, kötőszavak) és a pártok nevei (stopwords), linkek és hashtagek a szövegekből eltávolításra kerültek.⁶

VÁLTOZÓK

A kutatás elemzési egysége tehát az egyes Facebook-posztok, a függő változók pedig e posztok által kiváltott Reakciók (0 és 52 070 között, átlag = 174; szórás = 860), kommentek (0 – 16 499, á = 15, sz = 130) és megosztások (0 – 25129, á = 43, sz = 303) száma. A független változók a posztok témái, amelyeket szövegbányászati módszerekkel azonosítottam: szótár-alapú metódust alkalmaztam a migráció, a korrupció és a közpolitikai témák detektálására, míg a kampányhoz kapcsolódó és a mobilizációra fókuszáló tartalmak jelenlétét felügyelt gépi tanulási algoritmus segítségével rögzítettem (a főbb szövegbányászati megközelítésekről lásd: Sebők 2016). Az automatizált tartalomelemzésből származó változók érvényességének ellenőrzéséhez a szövegbányászati eredményeket összevettem egy kézíleg kódolt úgynevezett „aranystandard” adatbázissal. A modellek teljesítményét a pontosság (precision), a felidézés (recall) és az F1 értékek segítségével értékeltem. A pontosság azt mutatja meg, hogy az automatizáltan azonosított elemek hány százaléka helyes, míg a felidézés arra világít rá, hogy az adott kategóriába tartozó elemek mekkora részét tudta azonosítani a szövegbányászati metódus. Az F1-érték e két mutató alapján számolható ki, mely a 0 és 1 közötti tartományba eshet, ahol az 1 az aranystandardhoz való tökéletes igazodást, az emberi kódolással megegyező pontosságot és felidézést jelenti.

1600 véletlenszerűen kiválasztott poszt lett manuálisan is kódolva, ami egyszerűen szolgált a szótár-alapú módszer értékeléséhez szükséges aranystandardként, és a felügyelt gépi tanulás tanító halmazaként (training set). További véletlenszerűen kiválasztott 400 kézzel rögzített poszt a felügyelt gépi tanulási algoritmus teljesítményének értékeléséhez használt aranystandardként szolgált.⁷ A kódoló nem ismer-

⁵ 12 olyan jelölt volt, akiknek a posztjait nem lehetett az API-on keresztül letölteni, ezért a végső adatbázis 511 jelöltet tartalmaz.

⁶ Ehhez az R-re írt 'tm' csomagot alkalmaztam (Feinerer et al. 2008).

⁷ Az adatokat Burai Krisztina kódolta. A kézi kódolás megbízhatóságának ellenőrzésére a tanulmány szerzője is lekódolt 250 posztot a mintából, ez alapján pedig kiszámítható a Krippendorff's alpha értéke, amely minden kategória esetén megfelelő szintet mutatott (migráció = .82, korrupció = .81, egészségügy = .90, oktatás = .79, fejlesztéspolitika = .85, kampány = .74, mobilizáció = .87).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

te a témák azonosítására használt szótárakat, ezért a kézi kódolás független volt a szövegbányászati megközelítéstől.⁸

A témák szótáralapú azonosítására minden témához össze kellett állítani egy témaspecifikus szótárt. A szótár elemeinek kiválasztásánál két fő szempontot kellett érvényesíteni: egyrészt a kiválasztott szavak és szókapcsolatok valamelyikének a témához kapcsolódó posztok nagy többségében meg kell jelennie (magas szintű felidézés), másrészt a kiválasztott elemek csak elvétve bukkannak fel olyan posztokban, amelyek nem érintik az adott témát (magas szintű pontosság). Szerencsére a politikai ügyek általában eléggé specifikus szókinccsel rendelkeznek, ezért lehetséges olyan szavakat és szókapcsolatokat találni, amelyek (1) nagyon jellemzőek egy adott témára; (2) és kizárólag az adott téma kapcsán bukkannak fel (Eberl et al. 2020).

A szótár-összeállítás induktív és deduktív elemeket egyaránt tartalmazó iteratív folyamat eredménye volt (hasonló megközelítésért lásd Muddiman et al. 2019). Első lépésben intuitív alapon meghatároztam egy-egy témaspecifikus kulcsszót, úgy mint az „oktatás” és az „egyetem” az oktatáspolitikára, illetve a „korrupció” a korrupció témánál. Ezt követően az adatbázist leszűkítettem azokra a posztokra, amelyek e kulcsszavakat tartalmazták, majd e témaspecifikus szűkített adatbázis szógyakorisági táblázatát vizsgáltam meg. Ebből feltárultak azok a további szavak, amelyek az adott témákat tartalmazó posztokban különösen gyakran jelentek meg (1. feltétel). Mielőtt azonban új elemet adtam volna a szótárhoz, a kiválasztott szót tartalmazó posztok egy kisebb mintáját (N = 50) is átfutottam a teljes adatbázisból, hogy kiderüljön, hogy a szótárhoz adni kívánt elemet milyen kontextusban használják a Facebookon. Ez azért volt fontos, mert néhány esetben e vizsgálat arra világított rá, hogy bár az adott szó valóban gyakran jelenik meg a kiválasztott témában, de más témák kapcsán is visszatérően használják, így a szótárhoz adásával más témákat is ide rögzítene a módszer (2. feltétel). Egyes esetekben ez az ellenőrzés azt mutatta meg, hogy az adott szó más, nem a témához kapcsolódó szónak a része, ezért a szótárhoz csak úgy adhatjuk hozzá, ha kivételt is rögzítünk (például: az iskolát jelölő „iskol” szó része a „miskolc” szónak is). A kibővített szótár alapján aztán újra létrehoztam a témára fókuszáló szűkebb adatbázist és annak szógyakorisági tábláját, és e folyamatot egészen addig ismételttem, amíg a szótárakat már nem lehetett értelmes módon bővíteni. A szótárak véglegesítése után ezek alkalmassá váltak a Facebook-posztokban megjelenő témák azonosítására. Azokat a posztokat tekinthetjük egy adott témával foglalkozónak, amelyekben a szótár bármelyik eleme megjelenik. Fontos megjegyez-

⁸ A kézi kódoláshoz a kódutasítás a következő volt: bármilyen utalás a poszt szövegében (korrupció) a közpénzek illegális vagy etikailag kifogásolható elköltésére; (migráció) illegális vagy legális migrációra, annak okaira vagy következményeire; (egészségügy) egészségügyi rendszerre, illetve az ahhoz kapcsolódó kihívásokra; (oktatás) az oktatási rendszerre, illetve az ahhoz kapcsolódó kihívásokra; (fejlesztéspolitika) bármilyen fejlesztésre, beleértve a fejlesztési ígéreteket, illetve az átadási/felavatási eseményeket; (kampány) a kampányra, a választásokra vagy a politikai versengésre; (mobilizáció) bármilyen offline cselekvésre való felhívásra.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ni, hogy ez a módszer csak az adott téma jelenlétét vagy hiányát rögzíti, azt már nem mutatja meg, hogy az mennyire játszik fontos szerepet a poszt egészében. Ez azt is jelenti, hogy egy poszt egyenél több témával is foglalkozhat.

1. táblázat. A szövegbányászati módszerek teljesítményének ellenőrzése
(az aranystandarddal való összehasonlítás alapján)
(N = 1600 a szótár-alapú és N = 400 a felügyelt gépi tanulás alapú módszereknél)

	Pontosság	Felidőzés	F1	Előfordulás			Módszer
				teljes	kormány- párti jelöltek	ellenzéki jelöltek	
Korrupció	84%	72%	0.78	4.8%	0.1%	6.7%	Szótár-alapú
Migráció	77%	100%	0.87	5.4%	9.6%	3.7%	Szótár-alapú
Egészségügy	76%	92%	0.83	4.5%	3.1%	5.1%	Szótár-alapú
Oktatás	80%	92%	0.86	4.4%	4.8%	4.3%	Szótár-alapú
Fejlesztéspol.	86%	80%	0.83	5.4%	12.1%	2.7%	Szótár-alapú
Kampány	82%	72%	0.77	39.9%	30.3%	43.8%	Felügyelt gépi tanulás (AOC = 0.86; alpha = 0.1)
Mobilizáció	81%	67%	0.73	10.3%	9.3%	10.6%	Felügyelt gépi tanulás (AOC = 0.81; alpha = 0.6)
Negativitás (kontrollváltozó)	71%	58%	0.64	21%	10.9	25.2%	Felügyelt gépi tanulás (AOC = 0.93; alpha = 0.2)

Forrás: Saját szerkesztés

A kézilleg kódolt aranystandarddal való összevetés azt mutatja, hogy a témaspecifikus szótárak meglehetősen jól funkcionáltak. A pontosság-értékek 76% és 86% között, míg a felidőzés-értékek 72% és 100% között vannak. A legalacsonyabb F1-érték így 0,78 (korrupció), ami még mindig magas szintű érvényességet jelent (1. táblázat). Az eredmények ugyanakkor azt is mutatják, hogy a korrupció témája szinte láthatatlan volt a kormánypárti képviselőjelöltek kommunikációjában, ezért e változó esetében nincs értelme a H3 hipotézis által felvázolt interakciós hatást vizsgálni – a korrupciós posztok ugyanis szinte kizárólag az ellenzékhez tartoznak.

A kampány és a mobilizáció témákat már jóval nehezebb specifikus szavakhoz kapcsolni, ezért ezek jelenlétét felügyelt gépi tanulási módszerekkel azonosítottam. Ennek érdekében az adatbázist a tf-idf súlyozás⁹ segítségével úgynevezett doku-

⁹ A tf-idf egy olyan súlyozási technika melyben magasabb érték tartozik a gyakrabban használt kifejezésekhez, de a növekedés nem egyenesen arányos, mert minél gyakrabban fordul elő az adott kifejezés, annál kisebb az egy egységnyi előforduláshoz tartozó növekedés (Kwarler 2017: 100).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

mentum-kifejezés mátrixá alakítottam, amelyben a sorokban az egyes posztok, az oszlopokban pedig a posztokban megjelenő kifejezések jelennek meg. A felügyelt gépi tanuláshoz a logisztikus ElasticNet regressziós modellt alkalmaztam¹⁰ (Kwartler 2017), és a modelleket ahhoz az alpha paraméterhez optimalizáltam, amelyek a legjobb eredményt mutatták az aranystandardhoz (N = 400) való illeszkedés tekintetében. A modellek teljesítményének értékelésekor az látható, hogy a felügyelt gépi tanulási módszer magas pontosságú eredményeket adott ki (82% a kampány és 81% a mobilizáció esetében), míg a felidézés terén valamivel gyengébben, de még mindig elfogadható szinten teljesített (1. táblázat).

Tehát a szövegbányászati módszerek a szöveges adatokon jól működtek, azonban a módszer legnagyobb korlátja, hogy a posztokat pusztán a szöveges tartalom alapján kategorizálja, miközben a közösségi média-kommunikációban fontos szerepet játszanak a nem szöveges tartalmak is. A vizuális elemek figyelmen kívül hagyásából következő torzítás mértékének becslése érdekében a kézzel kódolt minta egy kisebb része (N = 186) oly módon is rögzítve lett, hogy a kódoló a poszt egészét, így a poszthoz tartozó vizuális elemeket is figyelembe vette. A csak a szövegre fókuszáló és a poszt egészét vizsgáló kódolás összevetéséből az derült ki, hogy a szövegen túli elemek figyelembevétele a témák kategorizálásában nem eredményezett jelentős eltéréseket. A migráció, az egészségügy és az oktatás esetében nem volt olyan poszt, aminek kódolása különbözött volna a két adatbázisban, míg a korrupció és a fejlesztéspolitikánál csak egy poszt kapott eltérő kódot (1%-os torzítás). Erősebb torzítás volt megfigyelhető a mobilizáció (3%) és a kampány-téma (6%) esetében, de ezek a szintek is elfogadhatóak. Következésképpen, a szövegalapú kategorizálás magas érvényességű eredményt kínál, hiszen az egyéb kommunikációs eszközök figyelembevétele nem javította jelentősen az adatok érvényességét.

Az érvényes eredmények eléréséhez számos, a felhasználói reaktivitásra potenciálisan hatást gyakorló egyéb változó kontroll alatt tartására is szükség van. A jelöltek szintjén a nem (1 = nő, összes poszt 15%-a), politikai hovatartozás (1 = kormánypárti jelölt, 29%), helyi politikai pozíció (1 = helyi polgármester vagy képviselőtestületi tag, 36%), inkumbencia (1 = parlamenti képviselő, 36%), és a jelöltség helye (1 = csak pártlista, 9%) lett kontroll alatt tartva. Az oldal szintjén a követők száma (kampány végén, 21 és 608 119 között, átlag = 16 064, szórás = 58 199), és a kampány alatt közzétett posztok száma került be a modellekbe. A posztok szintjén a közzététel időpontja (ahol 1 a kampány első napja, 51 az utolsó nap), és a poszt strukturális elemei, miszerint a poszt tartalmaz-e képet (1=igen, 52%), videót (1=igen, 21%), külső oldalról megosztott linket (1=igen, 22%) vagy esemény-megosztást (1=igen, 2%), valamint a szöveg hossza (karakterben, 1 és 18 291 között, á = 304, sz = 488), lettek figyelembe véve. A posztok tartalmával kapcsolatban azonban a legrobosztu-

¹⁰ Az elemzéshez az R-alapú glmnet (Friedman et al. 2010), a caret (Kuhn 2008) és a pROC (Robin et al. 2011) csomagokat használtam.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

sabb eredmény a szakirodalomban az, hogy a negatív tartalmak több felhasználói reakciót váltanak ki (lásd: Bene 2020, Heiss et al. 2019, Xenos et al. 2017). A fentebb ismertetett felügyelt gépi tanulási módszerrel tettünk kísérletet a negatív tónusú posztok azonosítására. Ebben az esetben a módszer gyengébb érvényességgel működött, mint a témák esetében, különösen a felidézés terén teljesített gyengén (lásd, 1. táblázat): tehát míg a legtöbb negatívként azonosított poszt valóban negatív, a valóban negatív tartalmak egy nem elhanyagolható hányadát nem ismerte fel a modell. Azonban figyelembe véve e faktor elméleti fontosságát, az eredmények robusztuságát növelendő, egy külön modellben az ily módon azonosított negativitás hatását is kontroll alatt tartom. Ráadásul, más kutatásokban ehhez hasonló érvényességi szinten mért érzelmi változókat független változóként is felhasználják (Eberl et al. 2020).

Mivel az adathalmaz többszintű struktúrába rendeződik, és a függő változók túlszórta count-jellegű változók, ezért többszintű negatív binomiális regressziómodellel teszteltem a hipotéziseket, ahol az oldalak szintjén random intercept-et alkalmaztam.¹¹ Négy modell tartozik minden interaktivitás-típushoz: az első az interakciós együtthatók nélkül, csak a független változókat tartalmazza. Az interakciós hatások a második modellben jelennek meg. Mivel a mobilizáció változó fogalmilag a kampány-változó altípusa, ezért ennek hatását külön modellben teszteltem, tehát míg az első két modellben a kampány változó szerepel, és a mobilizáció változó nem, a harmadik modellbe a mobilizáció változó kerül be, és a kampány változó pedig ki. Végezetül a negativitás hatását, a változóval kapcsolatos érvényességi fenntartások miatt, külön modellben kontrollálok, így az csak a negyedik modellben jelenik meg.

Eredmények

A modellek eredményei a 2., 3. és 4. táblázatokban láthatóak. A hipotézisek tekintetében az eredmények meglehetősen vegyesek. A legegységesebb eredmény a korrupció témájához kapcsolódik: a korrupcióval foglalkozó posztok mindegyik dimenzióban több interakciót eredményeznek. A másik fő kampánytéma, a migráció kapcsán azonban már összetettebb a kép. Az interakciós együtthatók nélküli modellekben az látható, hogy a migrációval foglalkozó posztok kevesebb Reakciót és kommentet kapnak, viszont szignifikánsan gyakrabban osztják meg őket. Ebből tehát úgy tűnik, hogy míg a korrupció egy interakció-növelő téma, a migráció inkább viszsztatartó erejű e tekintetben. Az interakciós együtthatók azonban árnyalják a képet. A párthovatartozás és a migráció témája mindegyik interakció-típusnál negatív kapcsolatot mutat, ami éppen ellentétes azzal, amit a témagazda-hatás értelmében várunk. Meglepő módon ugyanis az ellenzéki politikusok oldalain megjelenő migrációs

¹¹ Az elemzéshez az R-alapú glmmADMB (Fournier et al. 2012) csomagot használtam.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

posztok több reakciót, kommentet és megosztást váltottak ki más posztokhoz képest, miközben a kormánypárti politikusok migrációs posztjaira a felhasználók éppen hogy még jóval kevésbé is reagáltak, kommentelték vagy osztották meg őket más tartalmakkal összehasonlítva (1. ábra). Tehát a migráció témája kapcsán egyfajta fordított témagazda-hatás rajzolódik ki: a témagazdák követői kevésbé reagálnak erre a tartalomtípusra, míg a nem témagazdák éppenhogy sikeresebbek riválisuk saját témájában.

A közpolitikai témák kapcsán is vegyes eredményeket látunk. A H2-t az oktatáspolitikai esetében teljes mértékig, míg az egészségügynél csak részben fogadhatjuk el. Az oktatással foglalkozó posztok szignifikánsan kevesebb Reakciót, kommentet és megosztást kaptak, míg az egészségüggyel kapcsolatos tartalmakat kevésbé kommentelik, ugyanakkor a megosztások terén népszerűek, míg a Reakcióknál nincs szignifikáns kapcsolat. Ezzel szemben a fejlesztéspolitika témája más mintázatot mutat. A fejlesztéspolitikával kapcsolatos posztok szignifikánsan több kommentet és megosztást kapnak, a Reakciók esetében pedig nem figyelhető meg együttjárás. Az interakciós együttthatók azonban itt is árnyalják az eredményeket. A témagazda-hatás egyedül az oktatás kapcsán érvényesül, és itt is csak a kommentek és a reakciók esetében. Míg az oktatással kapcsolatos posztok mindkét politikai táborban népszerűtlenekek, a kormánypárti képviselők oldalain még kevesebb figyelmet kapnak, mint az ellenzéki politikusoknál. A megosztások esetében viszont ismét a fordított témagazda-hatás rajzolódik ki: miközben mindkét táborban kevesebb megosztást kapnak az oktatással kapcsolatos posztok, ez a tendencia erősebb az ellenzéki politikusoknál, tehát az ellenzéki politikusok oktatással kapcsolatos bejegyzéseit még kevésbé osztják meg, mint a kormánypárti jelöltek esetében (1. ábra). Az egészségügy és a fejlesztéspolitika témáját szintén jelentős részben a fordított témagazda-hatás jellemzi. A kormánypárti politikusok követői nagyobb arányban reagálnak, kommentelnek és osztanak meg egészségüggyel kapcsolatos tartalmakat, bár a kommentelés esetében a kapcsolat nem szignifikáns. Mindeközben, az ellenzéki követők a fejlesztéspolitikával kapcsolatos tartalmakat reagálják és kommentelik magasabb arányban, mint a kormánypártiak.

A kampány és mobilizáció témákhoz kapcsolódó eredményeket jóval könnyebb értelmezni. A kampánytémák minden dimenzióban több interakciót váltanak ki, míg a mobilizációra törekvő posztokat szignifikánsan kevesebben reagálják, kommentelik és osztják tovább. Ezek az eredmények azt jelzik, hogy az emberek valóban erősen érdeklődnek a kampány iránt, és a kampányperiódusban különösen nyitottak a választásokkal kapcsolatos tartalmakra, de nem kívánnak az választási versengés aktív résztvevőivé válni. Ezért mind a H4, mind a H4a erős megerősítést nyert. A negativitás beléptetése a modellekbe nem változtatta meg a hatások szignifikanciáját, irányát és nagyságát.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

2. táblázat. A jelöltek posztjaira való Reakciók számát magyarázó random-intercept negatív binomiális regressziómodellek együtthatói

	1. modell	2. modell	3. modell	4. modell
migráció	-0.22 (.02) ^{***}	0.02 (.03)	0.02 (.03)	0.03 (.03)
korrupció	0.11 (.02) ^{***}	0.10 (.02) ^{***}	0.11 (.02) ^{***}	0.10 (.02) ^{***}
egészségügy	-0.01 (.02)	-0.09 (.03) ^{**}	-0.09 (.03) ^{***}	-0.09 (.03) ^{**}
oktatás	-0.07 (.02) ^{**}	-0.02 (.03)	-0.03 (.03)	-0.02 (.03)
fejlesztéspolitika	0.00 (.02)	0.07 (.03) [*]	0.03 (.03)	0.07 (.03) [*]
kampány	0.08 (.01) ^{***}	0.09 (.01) ^{***}		0.09 (.01) ^{***}
mobilizáció			-0.11 (.02) ^{***}	
párt*migráció		-0.52 (.04) ^{***}	-0.49 (.04) ^{***}	-0.52 (.04) ^{***}
párt *egészségügy		0.18 (.05) ^{***}	0.18 (.05) ^{***}	0.18 (.05) ^{***}
párt *oktatás		-0.16 (.05) ^{***}	-0.18 (.05) ^{***}	-0.16 (.05) ^{***}
párt*fejlesztéspolitika		-0.09 (.04) [*]	-0.08 (.04) [#]	-0.09 (.04) [*]
szöveg hossz	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}
kép	0.20 (.03) ^{***}	0.20 (.03) ^{***}	0.20 (.03) ^{***}	0.20 (.03) ^{***}
videó	-0.13 (.03) ^{***}	-0.13 (.03) ^{***}	-0.13 (.03) ^{***}	-0.13 (.03) ^{***}
link	-0.20 (.03) ^{***}	-0.20 (.03) ^{***}	-0.21 (.03) ^{***}	-0.20 (.03) ^{***}
esemény	-1.72 (.05) ^{***}	-1.71 (.05) ^{***}	-1.74 (.05) ^{***}	-1.72 (.05) ^{***}
nap	0.01 (.00) ^{***}	0.01 (.00) ^{***}	0.01 (.00) ^{***}	0.01 (.00) ^{***}
posztok száma	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}
követők száma	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}	0.00 (.00) ^{***}
nem	-0.10 (.14)	-0.10 (.14)	-0.10 (.14)	-0.10 (.14)
párt	0.63 (.15) ^{***}	0.67 (.15) ^{***}	0.65 (.15) ^{***}	0.66 (.15) ^{***}
helyi pozíció	0.28 (.11) [*]	0.28 (.11) [*]	0.28 (.11) [*]	0.28 (.11) [*]
inkumbens	0.65 (.14) ^{***}	0.65 (.14) ^{***}	0.65 (.14) ^{***}	0.65 (.14) ^{***}
csak listás jelölt	0.05 (.17)	0.05 (.17)	0.06 (.17)	0.05 (.11)
negativitás				-0.02 (.01) [#]
konstans	2.47 (.10) ^{***}	2.46 (.10) ^{***}	2.49 (.10)	2.46 (.10)
random intercept vari- anciája	1.176 (1.085)	1.173 (1.083)	1.174 (1.084)	1.174 (1.083)
Log-likelihood	-184781	-184683	-184702	-184682
diszperzió-paraméter	1.49 (.01)	1.50 (0.1)	1.50 (.01)	1.50 (.01)
AIC	369606	369418	369456	369418
N (1. szint)	37284	37284	37284	37284
N (2. szint)	507	507	507	507

Megjegyzés: Zárójelben a standard hiba. #p < .10; *p < .05; **p < .01; ***p < .001
Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

3. táblázat. A jelöltek posztjaira való kommentek számát magyarázó random-intercept negatív binomiális regressziómodellek együtthatói

	1. modell	2. modell	3. modell	4. modell
migráció	-0.06 (.03) [#]	0.09 (.05) [#]	0.10 (.05)[*]	0.07 (.05)
korrupció	0.12 (.04)^{**}	0.10 (.04)^{**}	0.11 (.04)^{**}	0.07 (.04)[*]
egészségügy	-0.09 (.04)[*]	-0.16 (.04)^{***}	-0.17 (.04)^{***}	-0.16 (.04)^{***}
oktatás	-0.30 (.04)^{***}	-0.19 (.05)^{***}	-0.19 (.05)^{***}	-0.18 (.05)^{***}
fejlesztéspolitika	0.08 (.04)[*]	0.25 (.06)^{***}	0.18 (.06)^{**}	0.25 (.06)^{***}
kampány	0.19 (.02)^{***}	0.19 (.02)^{***}		0.20 (.02)^{***}
mobilizáció			-0.12 (.03)^{***}	
párt*migráció		-0.34 (.07)^{***}	-0.28 (.07)^{***}	-0.34 (.07)^{***}
párt *egészségügy		0.12 (.09)	0.11 (.09)	0.12 (.09)
párt *oktatás		-0.41 (.08)^{***}	-0.45 (.08)^{***}	-0.41 (.08)^{***}
párt*fejlesztéspolitika		-0.28 (.07)^{***}	-0.26 (.07)^{***}	-0.26 (.07)^{***}
szöveghossz	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}
kép	-0.16 (.05)^{***}	-0.16 (.05)^{***}	-0.16 (.05)^{***}	-0.15 (.05)^{***}
videó	-0.24 (.05)^{***}	-0.25 (.05)^{***}	-0.25 (.05)^{***}	-0.25 (.05)^{***}
link	-0.21 (.05)^{***}	-0.21 (.05)^{***}	-0.23 (.05)^{***}	-0.23 (.05)^{***}
esemény	-2.64 (.10)^{***}	-2.64 (.10)^{***}	-2.66 (.10)^{***}	-2.62 (.10)^{***}
nap	0.01 (.00)^{***}	0.01 (.00)^{***}	0.01 (.00)^{***}	0.01 (.00)^{***}
posztok száma	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}
követők száma	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}	0.00 (.00)^{***}
nem	-0.16 (.14)	-0.15 (.14)	-0.15 (.14)	-0.15 (.14)
párt	-0.19 (.15)	-0.15 (.15)	-0.17 (.15)	-0.13 (.15)
helyi pozíció	0.37 (.11)^{***}	0.38 (.11)^{***}	0.38 (.11)^{***}	0.37 (.11)^{***}
inkumbens	0.61 (.14)^{***}	0.61 (.14)^{***}	0.61 (.14)^{***}	0.61 (.14)^{***}
csak listás jelölt	0.18 (.17)	0.18 (.17)	0.18 (.17)	0.17 (.17)
negativitás				0.15 (.02)^{***}
konstans	0.18 (.11) [#]	0.17 (.11) [#]	0.25 (.11) [*]	0.15 (.11)
random intercept varianciája	1.157 (1.076)	1.155 (1.075)	1.157 (1.076)	1.149 (1.072)
Log-likelihood	-92840.5	-92805.7	-92866.4	-92776.1
diszperzió-paraméter	0.61 (.01)	0.61 (.01)	0.61 (.01)	0.61 (.01)
AIC	185725	185663.4	185784.8	185606.2
N (1. szint)	37554	37554	37554	37554
N (2. szint)	507	507	507	507

Megjegyzés: Zárójelben a standard hiba. [#]p < .10; ^{*}p < .05; ^{**}p < .01; ^{***}p < .001

Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

4. táblázat. A jelöltek posztjaira való megosztások számát magyarázó random-intercept negatív binomiális regressziómodellek együtthatói

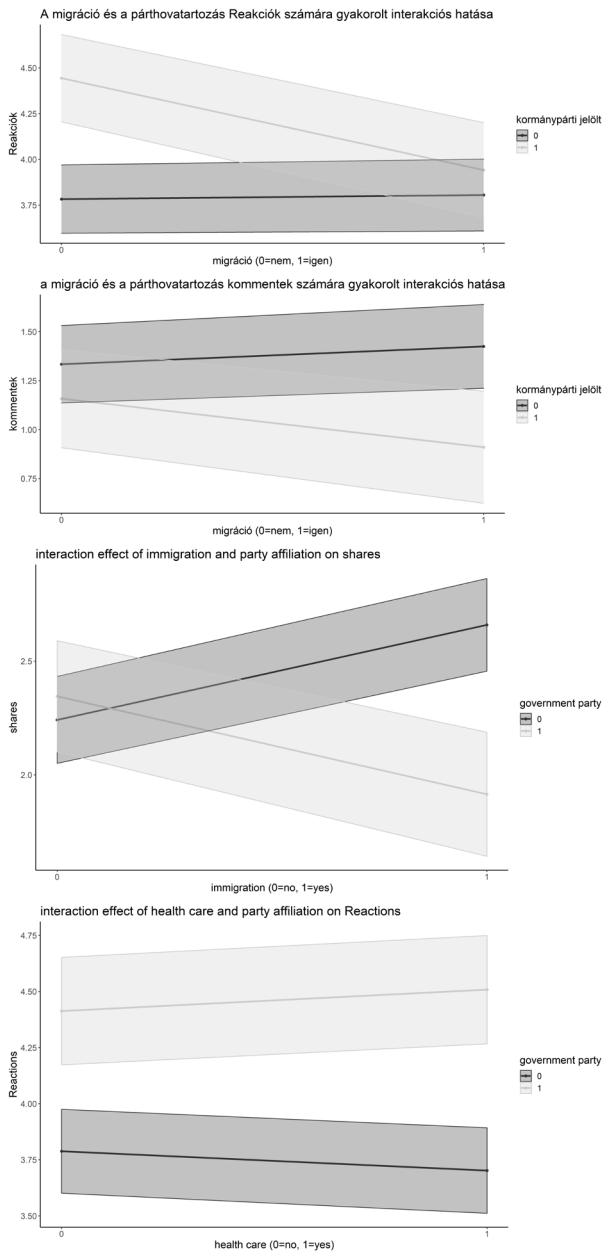
	1. modell	2. modell	3. modell	4. modell
migráció	0.05 (.03)*	0.42 (.04)***	0.42 (.04)***	0.39 (.04)***
korruptió	0.22 (.03)***	0.21 (.03)***	0.22 (.03)***	0.19 (.03)***
egészségügy	0.12 (.03)***	0.06 (.03)#	0.05 (.03)	0.06 (.03)#
oktatás	-0.15 (.03)***	-0.23 (.04)***	-0.23 (.04)***	-0.23 (.04)***
fejlesztéspolitika	0.24 (.03)***	0.17 (.04)***	0.17 (.04)**	0.16 (.04)***
kampány	0.11 (.01)***	0.12 (.01)***		0.13 (.01)***
mobilizáció			-0.11 (.02)***	
párt*migráció		-0.85 (.05)***	-0.80 (.05)***	-0.85 (.05)***
párt *egészségügy		0.21 (.07)**	0.20 (.07)**	0.21 (.07)**
párt *oktatás		0.17 (.06)**	0.15 (.06)*	0.17 (.06)**
párt*fejlesztéspolitika		0.09 (.06)	0.10 (.06)#	0.09 (.06)
szöveghossz	0.00 (.00)***	0.00 (.00)***	0.00 (.00)***	0.00 (.00)***
kép	0.40 (.04)***	0.42 (.04)***	0.42 (.04)***	0.42 (.04)***
videó	0.51 (.04)***	0.53 (.04)***	0.53 (.04)***	0.52 (.04)***
link	0.49 (.04)***	0.50 (.04)***	0.49 (.04)***	0.48 (.04)***
esemény	-7.05 (.39)***	-7.03 (.39)***	-7.03 (.39)***	-7.01 (.39)***
nap	0.00 (.00)***	0.00 (.00)***	0.01 (.00)***	0.00 (.00)***
posztok száma	0.01 (.00)***	0.01 (.00)***	0.01 (.00)***	0.01 (.00)***
követők száma	0.00 (.00)***	0.00 (.00)***	0.00 (.00)***	0.00 (.00)***
nem	-0.31 (.14)*	-0.31 (.14)*	-0.30 (.14)*	-0.30 (.14)*
párt	0.05 (.15)	0.08 (.15)	0.06 (.15)	0.10 (.15)
helyi pozíció	0.47 (.11)***	0.47 (.11)***	0.47 (.11)***	0.47 (.11)***
inkumbens	0.71 (.14)***	0.71 (.14)***	0.70 (.14)***	0.70 (.14)***
csak listás jelölt	0.02 (.17)	0.03 (.17)	0.04 (.17)	0.03 (.17)
negativitás				0.15 (.02)***
konstans	0.38 (.10)***	0.35 (.10)***	0.40 (.10)***	0.33 (.10)**
random intercept varianciája	1.206 (1.098)	1.200 (1.095)	1.197 (1.094)	1.191 (1.092)
Log-likelihood	-129929	-129790	-129820	-129740
diszperzió-paraméter	0.91 (.01)	0.92 (.01)	0.92 (.01)	0.92 (.01)
AIC	259902	259632	259692	259534
N (1. szint)	37486	37486	37486	37486
N (2. szint)	507	507	507	507

Megjegyzés: Zárójelben a standard hiba. #p < .10; *p < .05; **p < .01; ***p < .001

Forrás: Saját szerkesztés

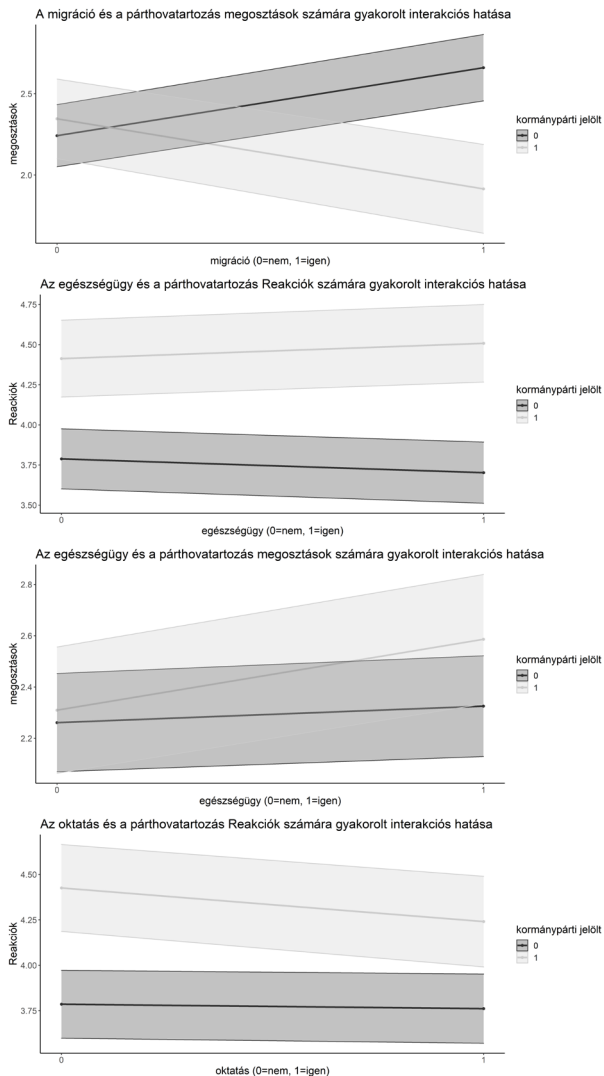


TEMATIKUS TANULMÁNYOK – Digitális szociológia

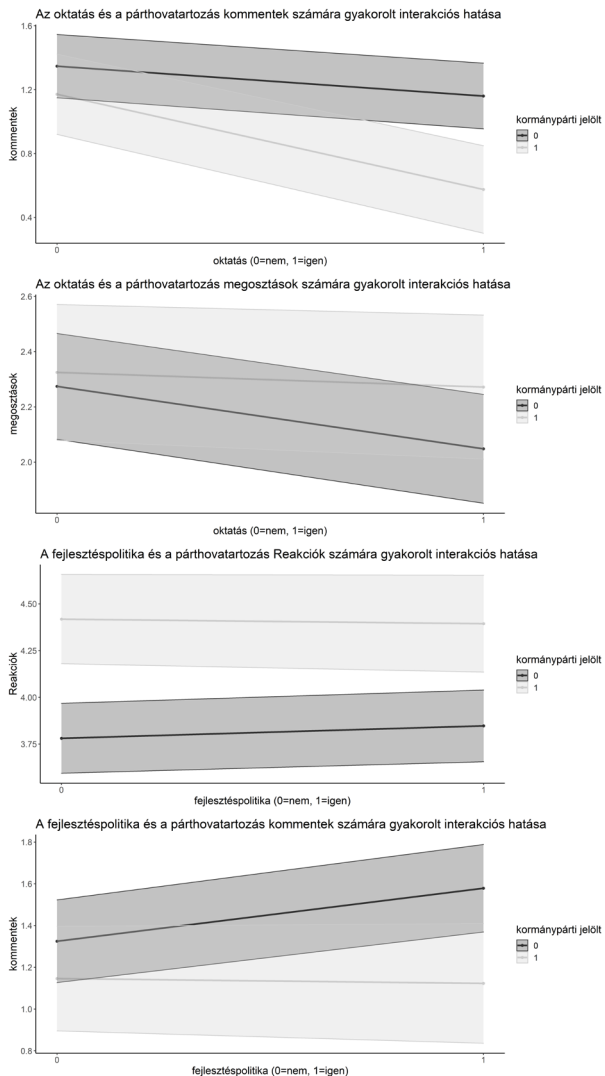




TEMATIKUS TANULMÁNYOK – Digitális szociológia



TEMATIKUS TANULMÁNYOK – Digitális szociológia



1. ábra. Szignifikáns interakciós hatások
Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Összegzés

A politikai kommunikáció viralizációjának tézisére támaszkodva a tanulmány kiindulópontja az volt, hogy a politikai kommunikáció egyre inkább az állampolgári kommunikáció sajátosságaihoz igazodik, mivel az állampolgárok reakciói képesek a politikai szereplők üzeneteit széles választói rétegekhez eljuttatni a közösségi oldalakon. Napjaink politikai kommunikációjának megértéséhez ezért fontos feltárnunk az állampolgári reaktivitás működését. A tanulmány az egyre bővülő szakirodalom két lényeges hiányosságát kísérelte meg pótolni a 2018-as magyarországi országgyűlési választási kampány kontextusában végzett vizsgálattal. Először is, a kutatás elsődleges fókuszja az egyébként eddig kevésbé vizsgált politikai témák kérdése: arra koncentrált, hogy a különböző politikai témák hogyan hatnak a felhasználói reakciókra a politikusok oldalain. Másodszor, automatizált tartalomelemzési eszközök alkalmazásával a hipotéziseket a korábbi kutatásokhoz képest jóval kiterjedtebb és átfogóbb adatbázison tesztelte, elkerülve ezáltal a mintavételi döntésekből fakadó általánosíthatósági korlátokat. A tanulmány bemutatta, hogy a szótár-alapú és a felügyelt gépi tanulásra támaszkodó szövegbányászati módszerek alkalmasak a Facebook posztokban felbukkanó főbb politikai témák magas megbízhatósági szinten való azonosítására.

A hipotézisek felállításához a kampány vezető témáit a kutatás három csoportra bontotta: kiemelt, közpolitikai és kampány témák lettek megkülönböztetve. Bár a hipotézisek ezekre a témacsoportokra vonatkoztak, az eredmények azt mutatják, hogy e témák a felhasználói aktivitásra való hatásuk tekintetében nem feltétlenül mozognak így együtt.

Az első hipotézis a korrupció kapcsán teljes megerősítésre talált, hiszen ez a téma több reakciót váltott ki minden dimenzióban, a migrációnál azonban árnyaltabb a kép. Ekkor meglepő módon csak az ellenzéki politikusok posztjainál igazolódott a feltevés, az ő migrációs tartalmaik valóban több reakciót váltanak ki. A kormánypárti politikusoknál azonban a migrációs tartalmak kevesebb felhasználói interakciót eredményeztek, mint más tartalmak, ami különösen meglepő annak fényében, hogy a kormánypártok 2018-as kampánya nagymértékben erről a témáról szólt. Összességében tehát azt mondhatjuk, hogy általános szabályként nem lehet kijelenteni, hogy a kiemelt kampánytémáknak reaktivitást erősítő hatása van.

Hasonlóan összetett a helyzet a második hipotézis esetében is. Az egészségügy, és még inkább az oktatáspolitikai nem számít túl népszerű Facebook témáknak, azonban azt nem állíthatjuk, hogy a közpolitikai témák általánosságban népszerűtlenek, hiszen a fejlesztéspolitika képében találtunk olyan szakpolitikai ügyet, ami sikeres a felhasználói interakciók kiváltásában. Ismét tehát azt látjuk, hogy míg a hipotézis részben igaznak bizonyul, a közpolitikára vonatkozó általános állításként már nem állja meg a helyét.

Ezek az eredmények azt jelzik, hogy a politikai témák egyéni jellege is számít. Bár e különböző karakterjegyek hatása további vizsgálatokat igényel, néhány előzetes



TEMATIKUS TANULMÁNYOK – Digitális szociológia

feltevést meg lehet fogalmazni. Először is, a témák különbözhetnek abban a tekintetben, hogy milyen érzelmi reakciók kapcsolódnak hozzájuk. A korrupció például egy általánosan helytelenített politikai magatartásforma, amelyhez leggyakrabban a düh érzése kapcsolódik, ami egy olyan érzés, amit viszonylag könnyű másokkal megosztani: az emberek gyakran beszélnek egymással azokról a dolgokról, ami felűhíti őket. Ezzel szemben a migráció érzelmileg már egy jóval diffúzabb ügy, melylyel kapcsolatban az emberek érzelmileg inkább megosztottak. A domináns negatív érzélem azonban itt talán a félelem lehet, mely érzést már jóval nehezebb a kiterjedt ismerősi hálózat előtt felvállalni, hiszen a félelem általában a gyengeség jeleként értelmeződik. Másodsor, a felhasználói aktivitás szintjére hatással lehet az adott téma földrajzi közelsége is. A fejlesztéspolitika sikere talán abban rejlik, hogy az ezzel kapcsolatos posztok gyakran helyi fejlesztésekkel kapcsolatos helyi szinten releváns információk, amelyeket könnyen meg lehet osztani és vitatni a környékbeli ismerősökkel. Ezzel szemben az egészségügy és az oktatáspolitikai esetében az jelenthet nehézséget, hogy a politikai szereplők e témákat sok esetben rendszerszinten tárgyalják. Későbbi tanulmányoknak érdemes tesztelni ezeket a feltevéseket.

Az eredmények csak a kampány- és mobilizáció témával kapcsolatos hipotéziseket támasztották teljes mértékben alá. Kampányidőszakban a követők előszeretettel interaktálnak kampánytartalmakkal még úgy is, hogy ezekből óriási a kínálat. A politikai versengés ezért egy olyan témának tűnik, amihez könnyen kapcsolódnak a választók. Ugyanakkor aktív résztvevővé nem akarnak válni, hiszen a választókat offline aktivitásokra felhívó tartalmakat nem szívesen promotálnak reakcióikkal, kommentjeikkel és megosztásaikkal.

A legváratlanabb és éppen ezért legizgalmasabb eredmény azonban a témagazda elméletéhez kapcsolódik. Az ebből az elméletből levezetett hipotézis csak elvétve teljesül: egyedül az oktatáspolitikai témája az, ahol a témagazda politikai oldal politikusai valamivel több interakciót képesek kiváltani, mint a másik oldal ugyanabban a témában közzétett posztjai. A korrupcióhoz kapcsolódó eredmény is valamelyest ide sorolható még, hiszen az ide tartozó posztokat szinte csak a témagazda ellenzéki politikusok tették közzé, így a szignifikáns hatások azt mutatják, hogy jól teljesítettek saját témájukban – itt azonban a kormányoldal teljesítményével nincs mód az eredmények összevetésére a vonatkozó posztok hiányában. Ezentúl azonban a legtöbb esetben éppen fordított témagazdag hatás azonosítható, azaz a témagazdák saját témájukban még gyengébb interaktivitás-kiváltó potenciállal is rendelkeznek, mint a másik oldal ugyanabban a témában. Ráadásul ez a fordított hatás a politikai spektrum mindkét oldalán megfigyelhető: a kormánypárti jelöltek egészségüggyel kapcsolatos posztjai sikeresebbek, míg az ellenzék a migráció és a fejlesztéspolitika terén ért el jobb eredményeket.

Empirikus bizonyítékok hiányában e megfigyelések okairól csak spekulálni lehet. A magyar kontextusban elég jól dokumentálták, hogy a politikai táborok által birtokolt témák valóban jelentős hatást voltak képesek gyakorolni saját szavazóik



TEMATIKUS TANULMÁNYOK – Digitális szociológia

percepcióira, attitűdjeire és politikai viselkedésére. E témák fontosságának megítélésében és értékelésében jelentős eltérések mutatkoznak a kormánypárti és ellenzéki szavazók között (Bíró-Nagy 2018, Kmetty 2018, Political Capital 2017). Ez alapján azt mondhatjuk, hogy ezek a kérdések valóban hatással vannak az emberek kognitív és affektív kapacitásaira. Azonban, ahogy arról fentebb volt szó, a közösségi média aktivitást más jellegű egyéni kapacitások, nevezetesen az expresszív, társadalmi és kommunikatív szükségletek motiválják.

A pártok által birtokolt témákkal kapcsolatos visszafogott aktivitás egyfajta téltődési hatásnak is köszönhető lehet. A kampányban a pártok olyan erővel nyomhatják témáikat, hogy azok a választók számára már nem nyújtanak újdonságot, érdekességet és izgalmat. A felhasználók, akik aktivitásaikkal a saját személyes politikai identitásukat akarják megjeleníteni ismerőseik előtt, vagy legalább érdekes és izgalmas tartalmakat mutatni nekik, aligha tudják e célokra felhasználni ezeket a tartalmat. A kormánypárti jelölt migrációval kapcsolatos, vagy az ellenzéki politikus egészségügy állapotával foglalkozó posztja talán nem elég érdekes ahhoz, hogy ismerőseink figyelmébe ajánljuk, vagy akár csak egy lájkra méltassuk. Ez adhat magyarázatot arra is, hogy miért azok a témák népszerűek ezzel szemben, ami éppen a másik oldalhoz kötődnek erőteljesebben: a szokatlan, az általános sztereotípiáknak valamelyest ellentmondó tartalmakhoz való kapcsolódáson keresztül a felhasználó egyedibbnek, érdekesebbnek és szórakoztatóbbnak tűnhet ismerősei előtt.

A pártosabb felhasználók esetén a fordított témagazdag-hatást az is magyarázhatja, hogy a felhasználó be akarja mutatni saját politikai táboraának azokkal a témákkal kapcsolatos érveit, amelyek a köztudatban inkább az ellenoldallal kötődnek össze. A Facebook hírvonala meglehetősen heterogén környezet (Barnidge 2017), ahol gyakran ütköznek versengő politikai érvek. Egyirányú kommunikáció esetén megfelelő stratégia lehet a saját témákra fókuszálni, azonban egy erősebb diszkurzivitást mutató nyilvános térben, amelyben számtalan vélemény, kinyilatkoztatás és nyilvánosan látható azonnali reakció zsúfolódik össze, fontosabb szerepe lehet a másik oldal témáira való saját válasznak, amelyek ellenérvként szolgálhatnak a politikai viták közösségi oldalakon végtelenül hömpölygő áradatában. A gyakorlat szintjére lefordítva: ha az ellenzéki szavazó számos helyen azt látja, hogy az általa preferált pártok a szabályozatlan migráció pártján állnak, a kormánypárti szavazó pedig gyakran szembesül azzal az érveléssel, hogy az Orbán-kormány tönkretette az egészségügyet, akkor különösen vonzó lehet minden olyan tartalom, ami ezekkel az állításokkal vitába száll, és az ellenzék valódi migrációs felfogását vagy éppen az egészségügy fejlődését mutatja be.

A kutatás elméleti kiindulópontjához visszacsatolva, ezek az eredmények összességében arra mutatnak rá, hogy különböző egyéni kapacitásokra eltérő politikai tartalmak gyakorolhatnak hatást. Míg a politikai kommunikáció egyes típusai hatékonyak a választók kognícióinak és érzelmeinek alakításában, ugyanezek már hatástalanok lehetnek az expresszivitás és a társas kommunikáció formálásában. Más



TEMATIKUS TANULMÁNYOK – Digitális szociológia

szóval: nem ugyanazokról a témákról beszélgetünk, amelyek gondolkodásunkat és érzelmeinket meghatározzák – legalábbis a közösségi média nyilvános terepein. A kormánypárti szavazók számára kétségkívül fontos téma a migráció, és érzelmileg is felkavarja őket a kérdés, a Facebookon mégis jóval kevésbé reagálnak a migrációval foglalkozó üzenetekre. Ha ez a megállapítás helytálló, és különösen, ha igaz a politikai kommunikáció más elemeire is, akkor ez azt is jelenti, hogy a politikai kommunikáció tartalma és formája jelentősen megváltozik a viralizáció trendjének hatására. Azonban több empirikus eredmény szükséges e következtetés levonásához, különösen olyan vizsgálatokra lenne szükség, amelyek a politikai tartalmakra adott kognitív, affektív és expresszív reakciókat hasonlítják össze.

A kutatás limitációjaként említhető meg, hogy a módszer a felhasználói reakciók mögötti motivációk kiszűrésére nem alkalmas: az interakciók mögött „nem őszinte”, azaz politikai érdekek által vezérelt vagy akár politikailag irányított aktivitások is jelen lehetnek. Ugyanakkor a felhasználói reakciók motivációtól függetlenül növekszik a vonatkozó tartalmak láthatóságát, ezért az eredmények a viralizáció elmélete szempontjából való jelentőségét ez a korlát kevésbé érinti. Az eredmények interpretációjában továbbá akkor jelenthetnek gondot az irányított felhasználói aktivitások, ha azok szisztematikus mintázatot követnek a vizsgált tényezők tekintetében. Feltételezhetjük azonban, hogy a politikai szereplők leginkább a saját témáik és mobilizációs üzeneteik promotálásában érdekeltek, így amennyiben az ilyen jellegű felhasználók érdemben hatással lennének az eredményeinkre, akkor pozitív téma-gazda-hatást, és a mobilizációs tartalmak népszerűségét kellett volna találnunk, eredmények azonban éppen ezzel ellentétesek. Felmerülhet az is, hogy az irányított aktivitások az ellenoldal szereplőinek posztjaira fókuszálnak, és egyfajta „trollhatás” állhat a fordított téma-gazda hatás mögött. Stratégiai szempontból azonban nem ésszerű az ellenoldal láthatóságát éppen a saját témák kapcsán irányított aktivitás segítségével növelni. Ráadásul ez maximum az ellentétes álláspontot közvetlenül a posztnál megjelenített kommentek esetében lehet releváns, viszont a kommentek esetében nem látunk markánsan eltérő mintázatokat a Reakciókhoz és megosztásokhoz képest. Ezen okok miatt okkal feltételezhetjük, hogy a feltárt tendenciák nem irányított aktivitások eredményei.

Bár a tanulmány kutatási dizájnja lehetővé tette a mintavételi döntésekből fakadó korlátok elkerülését, és a hipotéziseket egy kiterjedt és átfogó mintán tesztelte, a kutatás kontextusa még mindig akadályt jelent az eredmények általánosíthatóságával kapcsolatban. Először is, a magyar kontextus meghatározta a vizsgált témákat, hiszen a vizsgált kampány legfontosabb témáira fókuszált a kutatás. Bár ezek nem országspecifikusak, és népszerű kampánytémáknak számítanak szerte a világon, természetesen számos egyéb téma hatását is érdemes vizsgálni. Másodszor, a felhasználói reaktivitás mintázatait az adott kontextus számos egyedi vonása alakíthatja, így a politikai kultúra, a történelem vagy éppen az adott politikai szituáció. Egyetlen országra vonatkozó vizsgáltként ezeket a szempontokat itt nem lehet ki-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

szűrni a kutatásból. Nagy szükség van ezért túllépni az egy országra szűkített vizsgálódásokon, és a hipotéziseket több országot lefedő mintán is tesztelni. Végezetül bár a szótár-alapú megközelítés magas érvényességű eredményeket adott, a felügyelt gépi tanulási metódust még érdemes tovább fejleszteni a jövőben.

Irodalom

- Ansolabehere, S. – Iyengar, S. (1994): Riding the Wave and Claiming Ownership Over Issues: The Joint Effects of Advertising and News Coverage in Campaigns. *The Public Opinion Quarterly*, 58(3), 335–357.
- Anspach, N. M. (2017): The New Personal Influence: How Our Facebook Friends Influence the News We Read. *Political Communication*, 34(4), 590–606.
- Banducci, S. – Hanretty, C. (2014): Comparative determinants of horse-race coverage. *European Political Science Review*, 6(4), 621–640.
- Barnidge, M. (2017): Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication*, 34(2), 302–321.
- Baumgartner, F. – Jones, B. D. (1993): *Agendas and Instability in American Politics* (2nd ed.). University of Chicago Press, Chicago
- Bene M. (2020): *Virális politika. L'Harmattan, Budapest*
- Bene M. – Somodi D. (2018): „Mintha lenne saját médiánk...” A kis pártok és a közösségi média. *Médiakutató*, 19(2), 7–20.
- Bennett, W. L. (1990): Toward a Theory of Press-State. *Journal of Communication*, 103–127.
- Bennett, W. L. – Segerberg, A. (2013): *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press, New York
- Bíró-Nagy A. (2018): Politikai lottóötös: A migráció jelentősége a magyar politikában, 2014–2018. In: Böcskei B. – Szabó A. (szerk.): *Várákosok és valóságok. Parlamenti választás 2018. Napvilág Kiadó, MTA TK PTI, Budapest, 269–291.*
- Blumler, J. G. (1965): Producers' Attitudes towards Television Coverage of an Election Campaign: A Case Study. *The Sociological Review*, 13(1), 85–115.
- Blumler, J. G. (2016): The fourth age of political communication. *Politiques de communication*, (1), 19–30.
- Blumler, J. G. – Kavanagh, D. (1999): The Third Age of Political Communication: Influences and Features. *Political Communication*, 16(3), 209–230.
- Boda Zs. – Sebők M. (szerk.) (2018): *A magyar közpolitikai napirend: elméleti alapok, empirikus eredmények*. MTA TK Politikatudományi Intézet, Budapest



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Bucher, T. (2012): Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media - Society*, 14(7), 1164–1180.
- Coleman, S. – Blumler, J. G. (2009): *The Internet and democratic citizenship: Theory, practice and policy*. Cambridge University Press, Cambridge
- Dobos G. – Gyulai A. – Horváth A. (2018): Felesleges erőfeszítések? Választási programok és ígéreték 2018-ban. In: Böcskei B. – Szabó A. (szerk.): *Várákozások és valóságok. Parlamenti választás 2018*. Napvilág Kiadó, MTA TK PTI, Budapest, 317-339.
- Dolezal, M. – Zeglovits, E. (2014): Almost an Earthquake: The Austrian Parliamentary Election of 2013. *West European Politics*, 37(3), 644–652.
- Eberl, J. M. – Tolochko, P. – Jost, P. – Heidenreich, T. – Boomgaarden, H. G. (2020): What's in a post? How sentiment and issue salience affect users' emotional reactions on Facebook. *Journal of Information Technology & Politics*, 17(1), 48-65.
- Ennsner-Jedenastik, L. – Gahn, C. – Bodlos, A. – Haselmayer, M. (2021): Does social media enhance party responsiveness? How user engagement shapes parties' issue attention on Facebook. *Party Politics*, Online First
- Ernst, N. – Esser, F. – Blassnig, S. – Engesser, S. (2019): Favorable Opportunity Structures for Populist Communication: Comparing Different Types of Politicians and Issues in Social Media, Television and the Press. *The International Journal of Press/Politics*, 24(2), 165–188.
- Feinerer, I. – Hornik, K. – Meyer, D. (2008): Text Mining Infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54.
- Fournier D. A. – Skaug, H. J. – Ancheta, J. – Ianelli, J. – Magnusson, A. – Maunder, M. – Nielsen, A. – Sibert, J. (2012): AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27, 233-249.
- Friedman, J. – Hastie, T. – Tibshirani, R. (2010): Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.
- Gerodimos, R. – Justinussen, J. (2015): Obama's 2012 Facebook Campaign: Political Communication in the Age of the Like Button. *Journal of Information Technology - Politics*, 12(2), 113–132.
- Goffman, E. (1959): *The Presentation of Self in Everyday Life*. Anchor, New York
- Guo, L. – Vargo, C. J. – Pan, Z. – Ding, W. – Ishwar, P. (2016): Big Social Data Analytics in Journalism and Mass Communication: Comparing Dictionary-Based Text Analysis and Unsupervised Topic Modeling. *Journalism & Mass Communication Quarterly*, 93(2), 332–359.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Heiss, R. – Schmuck, D. – Matthes, J. (2019): What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, 22(10), 1497–1513.
- Iyengar, S. – Peters, M. D. – Kinder, D. R. (1982): Experimental Demonstrations of the “Not-So-Minimal” Consequences of Television News Programs. *American Political Science Review*, 76(4), 848–858.
- Keller, T. R. – Kleinen-von Königslöw, K. (2018): Pseudo-discursive, mobilizing, emotional, and entertaining: Identifying four successful communication styles of political actors on social media during the 2015 Swiss national elections. *Journal of Information Technology - Politics*, 15(4), 358–377.
- Kelm, O. (2020): Why do politicians use Facebook and Twitter the way they do? The influence of perceived audience expectations. *SCM Studies in Communication and Media*, 9(1), 8–34.
- Klinger, U. – Russmann, U. (2017): „Beer is more efficient than social media” – Political parties and strategic communication in Austrian and Swiss national elections. *Journal of Information Technology & Politics*, 14(4), 299–313.
- Klinger, U. – Svensson, J. (2015): The emergence of network media logic in political communication: A theoretical approach. *New Media & Society*, 17(8), 1241–1257.
- Kmetty Z. (2018): Korrupció percepciója, pártosság, választási részvétel: Hogyan változott a szavazók véleménye a hazai politikai korrupcióról a 2014-2018-as parlamenti ciklus alatt? In: Böcskei B. – Szabó A. (szerk.): *Várákosások és valóságok. Parlamenti választás 2018.* Napvilág Kiadó, MTA TK PTI, Budapest
- Kreiss, D. (2016): *Prototype Politics: Technology-Intensive Campaigning and the Data of Democracy.* Oxford University Press, New York
- Kruikemeier, S. – Gattermann, K. – Vliegenthart, R. (2018): Understanding the dynamics of politicians' visibility in traditional and social media. *The Information Society*, 34(4), 215–228.
- Kuhn, M. (2008): Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1 - 26.
- Kwartler, T. (2017): *Text Mining in Practice with R.* Wiley, Hoboken
- Lundby, K. (ed.). (2009): *Mediatization: Concept, Changes, Consequences.* Peter Lang Inc., International Academic Publishers, New York
- McCombs, M. E. – Shaw, D. L. (1972): The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, 36(2), 176–187.
- McGregor, S. C. (2020): „Taking the Temperature of the Room”: How Political Campaigns Use Social Media to Understand and Represent Public Opinion. *Public Opinion Quarterly*, 84(S1), 236–256.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Mudde, C. (2004): The Populist Zeitgeist. In: Government and Opposition, 39(4), 541–563.
- Muddiman, A. – McGregor, S. C. – Stroud, N. J. (2019): (Re) claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. Political Communication, 36(2), 214–226.
- Petrocik, J. R. (1996): Issue Ownership in Presidential Elections, with a 1980 Case Study. American Journal of Political Science, 40(3), 825–850.
- Political Capital (2017): Gyógyítható-e az egészségügy? Friedrich Ebert Stiftung, Budapest
- Robin, X. – Turck, N. – Hainard, A. – Tibertim N. – Lisacek, F. – Sanchez, J. – Müller, M. (2011): pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77.
- Sahly, A. – Shao, C. – Kwon, K. H. (2019): Social Media for Political Campaigns: An Examination of Trump's and Clinton's Frame Building and Its Effect on Audience Engagement. Social Media + Society, 5(2).
- Sebők M. (szerk.) (2016): Kvantitatív szövegelemzés és szövegbányászat a politikatudományban. L'Harmattan, Budapest
- Staender, A. – Ernst, N. – Steppat, D. (2019): Was steigert die Facebook-Resonanz? Eine Analyse der Likes, Shares und Comments im Schweizer Wahlkampf 2015. Studies in Communication and Media, 8(2), 236–271.
- Štětka, V. – Surowiec, P. – Mazák, J. (2019): Facebook as an instrument of election campaigning and voters' engagement: Comparing Czechia and Poland. European Journal of Communication, 34(2), 121–141.
- Strömbäck, J. (2008): Four Phases of Mediatization: An Analysis of the Mediatization of Politics. The International Journal of Press/Politics, 13(3), 228–246.
- Strömbäck, J. – Esser, F. (2014): Introduction. Journalism Studies, 15(3), 243–255.
- Svensson, J. (2011): The Expressive Turn of Citizenship in Digital Late Modernity. JeDEM - EJournal of E-Democracy and Open Government, 3(1), 42–56.
- Turcotte, J. – York, C. – Irving, J. – Scholl, R. M. – Pingree, R. J. (2015): News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. Journal of Computer-Mediated Communication, 20(5), 520–535.
- Xenos, M. A. – Macafee, T. – Pole, A. (2017): Understanding variations in user response to social media campaigns: A study of Facebook posts in the 2010 US elections. New Media - Society, 19(6), 826–842.



Merre tovább, netgeneráció?

A magyar fiatalok életstílus-alapú szegmensei

SÜTŐ ANNA¹

ABSZTRAKT

Tanulmányom kísérletet tesz a magyar fiatalság életstílus-alapú szegmenseinek újító módszer-tanú, közösségimédia-adatokon alapuló feltárására, beemelve a digitalizáltság dimenzióját az életstíluscsoportok alkotásába. A szegmensek életstílus-attitűdjeinek vizsgálatát segítik a kapcsolódó miliőelméleti megközelítések, és nemzetközi illetve magyar empirikus miliőkutatások áttekintése.

KULCSSZAVAK: életstílus, miliő, ifjúságkutatás, big data, Facebook

ABSTRACT

Where to go, net generation?

Lifestyle-based segments of the Hungarian youth

My study attempts to explore the lifestyle-based segments of the Hungarian youth through an innovative methodology based on social media data, incorporating the dimension of digitization into the creation of lifestyle groups. The examination of the segments' lifestyle attitudes is assisted by a review of the related theoretical milieu approaches, international and Hungarian empirical milieu researches.

KEYWORDS: lifestyle, milieu, youth research, Facebook

Bevezetés

A társadalom digitalizálódása és az internethasználat széles körű, gyors terjedése nagymértékben befolyásolja a társadalmi változások dinamikáját és sajátosságait is. A digitalizáció által katalizált információs forradalom az új lehetőségeknek, ezzel együtt a társadalmi mobilitásnak, élethelyzeteknek is olyan új útjait nyitja meg, amely alakítja a társadalmi struktúrákat is. Különösen igaz ez a fiatalabb generációk esetében, akik digitális bennszülöttként egyre inkább ebben az új környezetben szocializálódnak, és miközben a digitális világhoz való kapcsolódásukat társadalmi

¹ ELTE Szociológia Doktori Iskola, Interdiszciplináris Társadalomkutatások Doktori Program, e-mail: anna.suto@dataalyze.eu



TEMATIKUS TANULMÁNYOK – Digitális szociológia

helyzetük is befolyásolja, a digitálishoz fűződő viszonyuk alapvetően meghatározhatja életstílusukat is. Kérdés, hogy ezen folyamatok mentén milyen szegmensei rajzolódhatnak ki a mai magyar fiatalságnak.

Elsősorban feltáró jellegű tanulmányom során a magyar fiatalok életstílus-alapú csoportjainak (homogén struktúráinak) feltárására teszek kísérletet a közösségi média adatok egy innovatív, nem beavatkozó gyűjtésén, kategorizációján és elemzésén alapuló módszer segítségével, a Facebook aktivitáson keresztül rekonstruálható érdeklődési körök, attitűdök és fogyasztási szokások vizsgálatával. A digitális térben zajló viselkedéseken alapuló, és a digitális fejlődés által leginkább érintett fiatal generációt feltáró vizsgálat indokoltá teszi az életstílus-jellemzők mellett a digitális attitűdök dimenziójának beemelését, amelyet több, jelen tanulmányban bemutatott életstílus- és miliómodell is alkalmaz, melyek megközelítései segítséget nyújthatnak a feltárt szegmensek megértéséhez. A milióelméleti megközelítés célja az objektív társadalmi tényezők mellett a társadalmi cselekvések szubjektív, életvilág-alapú megismerése, melyekhez szükség van életstílushoz kötődő kvalitatív információkra is. Az életstílushoz köthető preferenciák jól rekonstruálhatóak a közösségi média-adatokból, hiszen ezek az aktivitások az érdeklődés és az életstílus digitális lenyomatai, egy bizonyos témájú Facebook-posztra érkezett reakcióból a poszt témáját, vagy a posztoló felület kommunikációját vizsgálva a Facebook-felhasználó ízlésére, értékeire, fogyasztási szokásaira egyaránt lehet következtetni.

Tanulmányomban a magyar fiatalok közösségimédia-aktivitásai mentén elemzem érdeklődésüket, és ez alapján csoportosítom tagjaikat életstílus-alapú szegmensekbe. Az így keletkező szegmensek vizsgálatának segítéséhez több egybecsengető elméleti és empirikus miliómodellt mutatok be és szintetizálok, továbbá megvizsgálom, hogy az általuk kirajzolt társadalmi mintázatok hogyan harmonizálnak a közösségimédia-adatok elemzésével. Az elméleti kiindulópontnak Hradil miliómodellje szolgál, amely egy széles körben elterjedt, és életstílusfókuszú miatt a közösségimédia-adatokból származó eredmények esetében is releváns, operacionálízható dimenziókat tartalmazó elmélet. Az elemzés segítésére két hasonló szemléletű, életstíluscsoportokkal dolgozó, a digitalizálódás dimenzióját is integráló empirikus modell is vizsgálatra kerül a közelmúltból, ezek a DIVSI Internetmiliók 2016-os társadalmi modellépítés kísérlete, valamint Ságvári Bence netgenerációról szóló 2011-es tanulmánya, amely kifejezetten a magyar fiatalokra fókuszál (DIVSI 2016, Ságvári 2011).

Mivel az életstílus egyik fő komponense az értékészlet, és a generációk kohorszélményei formálják a közösen elfogadott, képviselt értékeket, az életstílus vizsgálható generációs megközelítésben is. Vizsgálatom tárgyául azért választottam a 14-36 év közötti fiatalokat, mert ez a korosztály öleli fel azt a két generációt (Z és Y generációk), amelyek digitálisan bennszülöttek, azaz az infokommunikációs technológiák használata már komoly szerepet játszik szocializációjuk során (Strauss – Howe 1992). Az információs társadalom trendjeit követve ezek a generációk már



TEMATIKUS TANULMÁNYOK – Digitális szociológia

hazánkban is felzárkóztak a nyugathoz, egy olyan globális ifjúsági kultúra naprakész résztvevőiként, ahol eltűnik a generációk közötti, nagyságrendben értelmezhető különbség a világ nyugati felétől (Székely 2013). Ezáltal ezen korosztályok esetében kiváltképp izgalmasnak tűnik a nyugati modellek bemelése és az ezek alapján történő nemzetközi összevetés az életstílus-alapú vizsgálatuk során, különösen a digitális folyamatokat is figyelembe vevő elméletek esetében.

A Facebook-adatokon alapuló vizsgálatot a módszer egyedisége és újszerűsége mellett azok digitális attitűdökhöz köthető és egyben életstílushoz kapcsolódó kvalitatív jellege, és a rendelkezésre álló, más módon nem hozzáférhető adatmennyiség indokolja. A Facebookon kifejtett online aktivitások a magyar fiatalok vizsgálatakor különösen relevánsak lehetnek, mivel a közösségimédia-oldal látogatottsága hazánkban kiemelkedő: a 13 évesnél idősebb felnőtt népesség 71%-a átlagosan napi 20 percet látogatja, ezzel jelenleg a második leglátogatottabb online felület Magyarországon a Google után. A látogatók 43%-át a 13-34 évesek teszik ki, a Facebook-interakciók közül csak a posztkedvelések (like) havi átlaga 15 db felhasználóként (Kemp 2021).

A hazai lakosság körében az életstílust tekintve jelentős változások körvonalazódtak az elmúlt évtizedben. Ennek vizsgálatával leginkább marketingorientált kutatások foglalkoztak, azonban ezen hazai életstílus-kutatások esetében az eredmények és a módszertan megismerése legtöbbször csak erősen korlátozottan lehetséges. A fentebb felsorolt trendek kiemelt aktualitását adhatják a jelen tanulmányban bemutatott vizsgálatnak, amely nemzetközi modellek beemelésével, módszertanilag újító módon, közösségimédia-aktivitások vizsgálatával, a gyorsan változó, fiatal digitalizálódó generációk feltárására tesz kísérletet (Törőcsik et al. 2019).

Az életstílus- és miliőelméletek gyökere

A posztindusztriális társadalmakban az elmúlt évtizedekben tapasztalható radikális társadalmi átalakulások következményépp megfigyelhető az emberi életfeltételek drasztikus mértékű változása, melynek része a nyugati társadalmakban az egyéni életutak diverzifikálódása és individualizálódása. Ennek mentén a modern társadalmak szerkezetét egyre inkább olyan életfogalmi tényezők (élethelyzet, életút és életstílus) is meghatározzák, amelyek a rendi és az osztály szerinti rétegződésen túl befolyásolják az emberek társadalmi helyzetét, magatartását (Beck 1997). Ezzel párhuzamosan a társadalom rétegződésének kutatása is egyre nehezkesebbé válik, ahogy egyre körülményesebb lesz a társadalom differenciálódásának áttekintése, és elhalványulnak a nagycsoportok határai (Angelusz 2000). A nyugati társadalmakban végbemenő változások olyan jelentőssé váltak, hogy a struktúraelemzés diskurzusát korábban meghatározó vertikális rétegződés- és osztályelméletek már nem tudták maradéktalanul megragadni a szerkezeti tényezőket, és számos kritika érte



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ezeket a modelleket (Schulze 2003, Luhmann 2006). A korábbi rétegeképző tényezők leváltására több kísérlet született, ezek közül széles körben elterjedtek az elsősorban a német szociológiában megjelenő életstílus- és miliőelméletek, melyek sokdimenziós rétegződésfelfogásnak tekinthetők. A miliőelméletek kialakulására nagy hatással volt az Ulrich Beck német szociológus individualizációs elméletében megfogalmazott (Beck 1983, 1986) elgondolás, amely szerint a növekvő jólét és mobilitás hatására az emberek egyre inkább képesek kiszakadni a tradicionális környezetükből.

A miliőelméletek képviselői szerint a társadalmi változások elemzésekor a korábban vizsgált objektív életkörülményi tényezők mellett figyelembe kell venni az életmódból adódó szubjektív tényezőket is, mert a nagy társadalmi rétegekre specifikus viselkedésmódok befolyásoló ereje csökken, ahogy a társadalom egyre jobban differenciálódik, és növekszik a mobilitás, ezáltal növekszik a szűkebb csoportok hatása, melyet miliőnek neveznek (Hradil 1995). A miliőfogalom ezért magában foglalja az egyénre ható individuális, belső, illetve környezeti, külső szempontokat.

Ezek a megközelítések a 21. században empirikus kutatásokkal is alátámasztott miliőmodelleké fejlődtek (például a „Sinus-Mileus” modell, Bolte 2012), amelyek az egyének csoportosítását világnézetük, életmódjuk, értékorientációik, szociális helyzetük és életstílusuk alapján kísérik meg, a legtöbb esetben Hradil miliőelméletéből kiindulva. Hradil koncepciója szerint a társadalom szerkezete nem osztályok vagy rétegek segítségével, hanem kevésbé merev, több szempontot is integráló heterogén struktúrákkal, azaz miliőcsoportokkal írható le (Bolte – Hradil 1984, Hradil 1995).

A társadalom újfajta differenciálódása és az individualizálódás ugyanakkor nem feltétlenül jelenti azt, hogy a társadalmi struktúra áttekinthetatlenné aprózódik, egyes vélemények szerint éppen a társadalmi miliők alkotta, különböző társadalmi rétegeket átfogó csoportok adják az új nagycsoportok alapjait (Schulze 2003). E nagycsoportok a világ hasonló interpretációja révén, annak értelmezése során, jellemző szemléletmódok mentén közösségekként termelik újra és tartják fenn magukat. A társadalom tagozódási elveinek változásai mellett egyre hangsúlyosabbá válik a korcsoportok szerinti tagozódás, amely az élménytársadalom kialakulásával van összefüggésben, ti. azzal a jelenséggel, hogy a fő létp probléma egyre inkább az élményszerzéshez kötődik, amelyet a generációs élmények és élethelyzetek egyre inkább befolyásolnak (Schulze 1992).

Életstílusmodellek alkalmazása a magyar fiatalság feltárására

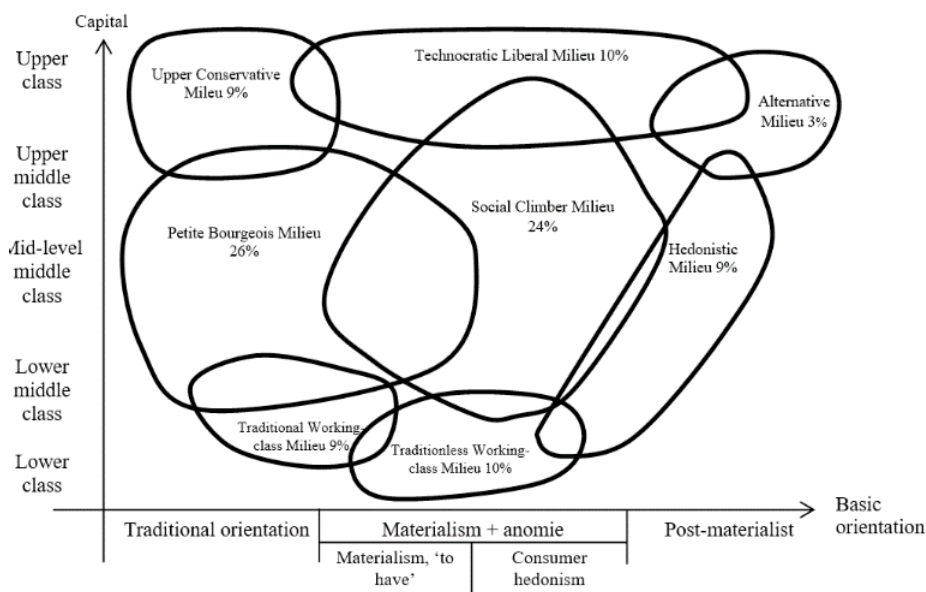
A hradili miliőelmélet gyakran alkalmazott kiindulópont az életstílus által struktúrált (csoporton belül homogén, csoportok között heterogén) kategorizálásnak. A társadalmimiliő-elmélet alapvetése tehát, hogy objektív és szubjektív tényezőket összeegyeztetve életstílus-kategóriákat hoznak létre, majd ezen életstílusok mentén



TEMATIKUS TANULMÁNYOK – Digitális szociológia

elemzik a társadalom szegmenseit. Hradil miliőelmélete is tartalmazza az fent leírt struktúrát, a miliő fogalmát a következőképpen definiálja: „az új társadalomtudományi kutatásban miliőn esetleg heterogén környezeti tényezők (legyenek azok anyagi vagy nem anyagi jellegűek, természetesek vagy társadalom által létrehozottak, gazdaságiak, politikai-adminisztratívak vagy társadalmi-kulturálisak besorolás szerint) összefüggéseit értjük, melyeket meghatározott társadalmi csoportok meghatározott módon ismernek fel és használnak ki, mely meghatározott életmódok kialakulását vonja maga után.” (Hradil 1995: 356)

A hradili szemlélet, amely inkább egy kulturális alapú, horizontálisan differenciálódó rétegződésmodell, a kilencvenes években egyre több kritikát kapott az inkább a hierarchikus társadalmi tagozódásokat, osztály- és rétegelméleteket szem előtt tartó rétegződéskutatók részéről. A modellt mégis sok esetben használják kiindulásként napjainkban is, és jelen tanulmány esetében is hasznos kiindulópontul szolgál. Ez elsősorban annak köszönhető, hogy a fogyasztási és kulturális rétegződési mintázatok vizsgálatával egy érzékenyebb, többszemponútú kutatási stratégiát tesz lehetővé, amelybe jól beemelhetőek a kulturális beágyazódással összefüggő digitális attitűdök is. Aktualitását a komplexitása mellett gyakori hivatkozottsága, empirikus mérhetősége és a társadalomelméletek közötti viszonylagos frissessége adja, amely alkalmassá teszi a fiatalok digitális szegmenseinek megértésére is (Hradil 1995).



1. ábra. A hradili miliők szemléltetése

Forrás: Linke 2017: 39



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Hradil az ábrán látható nyolc csoportot különböztette meg egymástól. A fő dimenziók, melyek a csoport-hovatartozást befolyásolták a társadalmi osztályhelyzet (státusz) és a tradicionális-posztmateriális értékskálán való elhelyezkedés határozta meg. Röviden az értékelköteleződésekről:

- A tradicionális értékek a hagyományos, individualizós feltörekvések előtti, közösségi értékeket képviselik, például a család, nemzet fontossága, vallás, hagyományos társadalmi szerepek.
- A materiális értékrendű emberek igényei anyagi jellegűek, például az anyagi jólét, árstabilitás, biztonság, szépség, hatalom.
- A posztmateriális értékek fókuszában már nem a materiális igények kerülnek, például a szólásszabadság, környezetvédelem, egyéni fejlődés, élménykeresés.
- Az anómiás, értékvesztett vagy értéksemleges állapot az a jelenség, amikor a társadalomban régen jelen lévő értékek már kivesztek, de helyükbe még nem kerültek újak.

A skála bal felén találhatóak a tradicionális értékeket vallók, míg a jobb fele haladva előbb a materiális/értéksemleges, majd a posztmateriális értékekhez kapcsolódók.

- Az alternatív miliő tagjait az egyszerű és hiteles gondolkodás jellemzi. Az eszmei értékek kerülnek előtérbe, mint az önkifejezés fontossága, kreativitás és megfelelő kommunikáció preferálása. Kritikusan állnak a civilizációhoz. Baloldali beállítottságúak és támogatják a béke, szabadság, igazságosság eszméit. Magasan képzettek, vagy még tanulnak és ezért alacsonyabb a jövedelmük, vagy már a munkaerőpiacon vannak, átlag feletti bevétellel.
- A technokrata-liberális miliőbe tartozókat a karrierépítés, teljesítményvágy, siker, önismeret és önfejlesztés jellemzi. Hangsúlyos náluk az objektív szemléletmód és progresszivitás. Eredményorientáltak, a társadalmi presztízs és a jövőbeli családi életben való funkcionális maximalizálása szerint alakítják életüket.
- A felső konzervatív miliő fókuszában a tradíció és presztízs áll. A csoporthoz tartozók számára fontos a magánélet és a család védelme, a társadalmi kötelek, az elköteleződés, társadalmi felelősségvállalás. Előszeretettel kritizálják a társadalomban feltörekvő értékvesztés tendenciáját, a materializmust és a morális romlást.
- A hedonista miliő hajlamos a szórakozás és luxus érdekében való fogyasztásra. Támogatja a függetlenség és a szabadság eszméit. A csoport tagjai szeretik áthágni a szabályokat, egyediségre törekuszenek és a munkát szükséges rosszként definiálják. Alacsony vagy közepes a jövedelmük és iskolai végzettségük.
- A felkapaszkodó miliőt a modern mainstream életstílus jellemzi. Folyamatosan várják a munkájukkal elért eredmények életszínvonalat változtató ha-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tását, melyért tesznek is nap mint nap. Fontos számukra a státusz és annak (vélt vagy valós) szimbólumai és ezek megmutatása mások számára. Iskolai végzettség és jövedelem szempontjából közép és alacsony kategóriában is lehetnek.

- A kispolgári milióhöz tartozókat bezárkózó és tradicionális beállítottság jellemzi. Idegenkednek az újdonságoktól és az ismeretlentől. Mindennapjaikat a kötelességtudat és a konvencionizmus fókuszában élik. Fontos számukra a biztonság. Konzervatív-kispolgári életet élnek, közepes vagy átlag feletti jövedelmi stártuszhoz tartozóként.
- Az új munkásosztály milió tagjai a kétkezi szakmai hivatás jegyében végzik munkájukat, tanulmányaikat. Idejük nagy részét jellemzően ez teszi ki, ezért szabadidejükben passzív kikapcsolódási formák és reális körülmények között tartott hedonizmus jellemzik őket. Átlagos iskolázottság, átlagos jövedelem, átlagos nem szellemi munkakört betöltő emberek alkotják a miliót.
- A tradicionális munkásosztály milió tagjai hátrányos helyzetűek. Az egyszerűség és takarékoság jegyeiben élnek. Fontos számukra a közösséghez tartozás érzése, a szolidaritás. Iskolai végzettségük és jövedelmi szintjük alacsony mértékű. Korlátozott cselekvőképességűek, a társadalmi folyamatok passzív elszennedői.

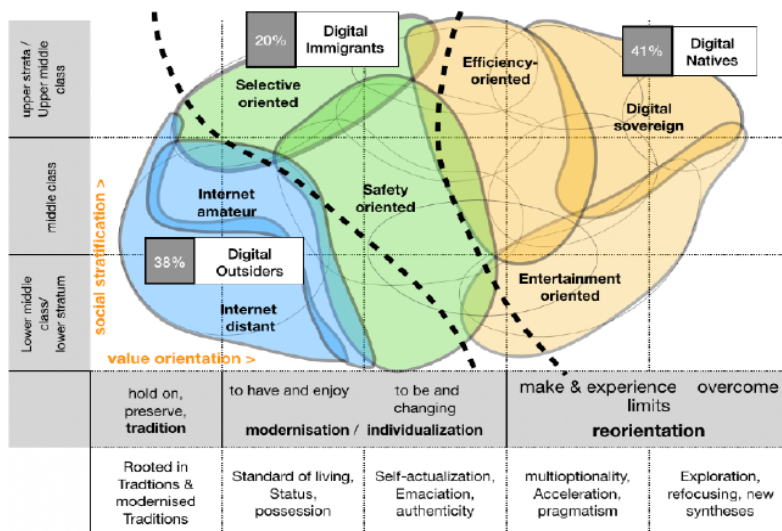
DIVSI Internetmiliók

Hradil fent vázolt modelljéből kiindulva számos további társadalmimodell-építési kísérlet született. A további vizsgálódás céljából egy olyan, hasonló érték- és osztály-hovatartozáson alapuló modell került kiválasztásra, amely a szegmensek megalkotásakor és tulajdonságaik vizsgálatakor a digitalizálódás dimenzióját is integrálja.

A DIVSI (Deutsches Institut für Vertrauen und Sicherheit im Internet) egy 2011-ben alapított (és 2018-ban megszűnt) nonprofit vállalkozás, melynek célja a magabiztos és biztonságos internethasználat megteremtése volt. A DIVSI 2016-ban egy milióalapú megközelítést alkalmazott a német fiatalok digitalizációjának, internethasználatának feltárására. A társadalmi rétegek életstílus alapján történő elkülönülése segítségével szerettek volna választ kapni a társadalom internethez kapcsolódó szokásainak, motivációinak, jártasságának és egyéb más tényezők kérdéseire. Az így kapott csoportokon végeztek további kutatásokat, jellemzőik mélyebb feltárása alapján. A digitalizáció dimenzióját is bevonták, hogy tovább pontosítsák a csoportok tulajdonságait (DIVSI 2016).

TEMATIKUS TANULMÁNYOK – Digitális szociológia

Picture 8: DIVSI Internet Milieus and Digital User Groups (Simus 2017, Borgsted 2016)



2. ábra. A DIVSI Internetmiliók szemléltetése
Forrás: Hellmann 2014: 1

A DIVSI által feltárt hét szegmens jellemzőik leírása alapján megfelelnek a Hradil által feltérképezett társadalmi strukturálódás képének, egy kivétellel, ahol két Hradil-féle miliócsoport egy szegmensevé olvad össze (új munkásosztály és felkapaszkodó miliók):

- A digitális szuverének egy aktív, de kritikus internetező tömeg. Viszonylag jó életszínvonal, magas iskolai végzettség jellemzi, és őszinte érdeklődés az iránt, amit csinál. Tudatos és felelős internethasználók.
- A teljesítményorientáltak átlag feletti iskolai végzettséggel rendelkeznek, töretűek, vállalkozó jelleműek. Magas életszínvonalon élnek. Jellemzően aktív és professzionális internethasználók, életmódjukban és az online világában is tudatosak. Érdeklí őket a tudatos életmód és táplálkozás, az világ megismerése, az utazás. Sok elektronikai eszközt tartanak maguk körül és aktívan használják is őket.
- A szelektív felhasználók körültekintően állnak az internethez, megválogatják, hogy mikor és mire használják, el tudnak szakadni az internettől. Tradicionális, nagycsaládos életet terveznek maguknak. Fontos érdeklődési köreik közé tartozik a sport és a hírolvasás. Nem rajonganak a kultúra iránt. Fontosak számukra a személyes kapcsolatok, az adatvédelem, a maguk és családjuk megvédése az internet veszélyeitől. Nem affinisek az újonnan piacra kerülő termékekre.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- A szórakozásorientáltak feltétel nélkül rajonganak az internet 'fun' része iránt. Keresik az élményeket, a szórakozási lehetőségeket, a különböző eszközöket, kiegészítőket. Gyakran még a szüleikkel élnek (mamahotel). Érdeklik az újdonságok, hajtja őket a kíváncsiság, igazi hedonisták, és fontos szempont számukra a közösségi hálózat kiterjesztése és ápolása.
- A kockázatkerülők csoportjába tartozók átlagos iskolai végzettséggel és jövedelemmel rendelkeznek, legtöbbször alkalmazotti beosztásban dolgoznak. Szabadidejükben sok TV-t néznek, a szépség, lakberendezés, dekoráció és autó témakörök érdeklik őket. Nem értenek az informatikai megoldásokhoz.
- Az internetamatőrök csoport tagjai nagyon visszafogott internethasználók, féltik az adataikat, sokszor nem értik a digitális világot. Átlagos, vagy alatti jövedelemmel és státusszal rendelkeznek. Konzervatív értékeket képviselnek. Kiemelten érdeklődők a közélet iránt. Nem jellemzi őket az elektronikai eszközök halmozása, sem pedig az online vásárlás, vagy bankolás – részben biztonsági megfontolásból.
- Az internettávoliak leginkább elkerülik az internet világát. Idősebb korosztály, akik alacsonyán képzettek és alacsony jövedelemmel rendelkeznek. Nem értik a technika világát. Tradicionális értékeket vallanak, a család, otthonteremtés és munka áll a fókuszukban.

A DIVSI kategorizációja, osztályozási skálái a Hradil által megalkotott, valamint ahhoz hasonló elméleti alapokon nyugszanak, mely kiindulópontként Hradilhoz hasonlóan egy érték- és egy osztály-hovatartozás tengelyen alakította ki a rétegeket, a két elméleti modell összekapcsolása ezért lehetségesnek tűnik.



3. ábra. A hradili milióelmélet és a DIVSI Internetmiliók szerinti szegmensek összekapcsolása jellemzőik alapján

Forrás: Saját szerkesztés



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Mivel a DIVSI Internetmiliók foglalkoznak a digitalizálódás dimenziójával is, a modell alkalmassá válik arra, hogy a digitális attitűdöket is vizsgáló kutatás elméleti támasza lehessen.

A fiatalok életstílus-csoportjai Magyarországon

A fent alkalmazott modellek a német társadalom feltárása alapján keletkeztek, ezért ezek áttekintése után indokoltá válhat egy olyan megközelítés integrálása, amely kifejezetten a magyar fiatalságot tanulmányozza és modellezi hasonló szempontok alapján. Ságvári Bence szociológus 2011-ben az „A net-generáció törésvonalai – Kultúrafogyasztás és életstílus-csoportok a magyar 20-29 év közötti fiatalok körében” című tanulmányában a Magyar Ifjúság 2008² felmérés alapján tanulmányozta a magyar fiatalokat (Ságvári 2011).

Ságvári az eddig tárgyalt szerzőkhöz és tanulmányokhoz hasonlóan homogén struktúrákat keresett a társadalomban, de a fiatalok körére szorítkozva, méghozzá az informatikai eszközök használatának dimenzióján keresztül. Médiafogyasztási és életstílusjegyeik alapján hat csoportot különített el a vizsgált személyek adatbázisán. Szegmentációs szempontjai a következők:

- Társas kapcsolatok (baráti társaság, barátokkal való időtöltés)
- Kulturális fogyasztás (színház, mozi, könyvtár, múzeum)
- Médiafogyasztás (nyomtatott: napilapok, hetilapok)
- Médiafogyasztás (elektronikus: tv, rádió)
- Olvasás (mint a szabadidő eltöltésének formája)
- Anyagi státusz
- Kulturális tőke (iskolai végzettség, nyelvtudás, különórák).

A szegmenseket digitalizáció szerint hierarchikus sorrendben elhelyezte egy felső, egy átmeneti, és egy alsó skálán és felmérte mintabeli nagyságukat. Az általa így módon meghatározott hat szegmens:

- A digitális elit szegmens tagjai stabil anyagi helyzettel rendelkeznek, átlag feletti végzettségűek és jellemző rájuk a családi háttérükből jövő támogatás. Beszélnek idegen nyelveket, ezáltal magas arányukban a nemzetközi médiafogyasztás. Érdeklődnek a közélet, politikai iránt, ezek esetén is nemzetközi viszonylatokban is tájékozódnak. Általában jellemző rájuk, hogy sokat olvasnak, szabadidős tevékenységként, szórakozás, és tájékozódás céljából is.

² A Magyar Ifjúságkutatás egy négyévente elvégzett nagymintás, 18-29 éves korosztályt vizsgáló adatfelvétel, amely társadalomtudományos megközelítés alapján térképezi fel a változásokat és jellegzetességeket. A 2008-as Magyar Ifjúságkutatás elsődleges adatairól részletesen lásd: Bauer Béla – Szabó Andrea (2009) (szerk.): Ifjúság 2008. Gyorsjelentés. Budapest, Szociálpolitikai és Munkaügyi Intézet.

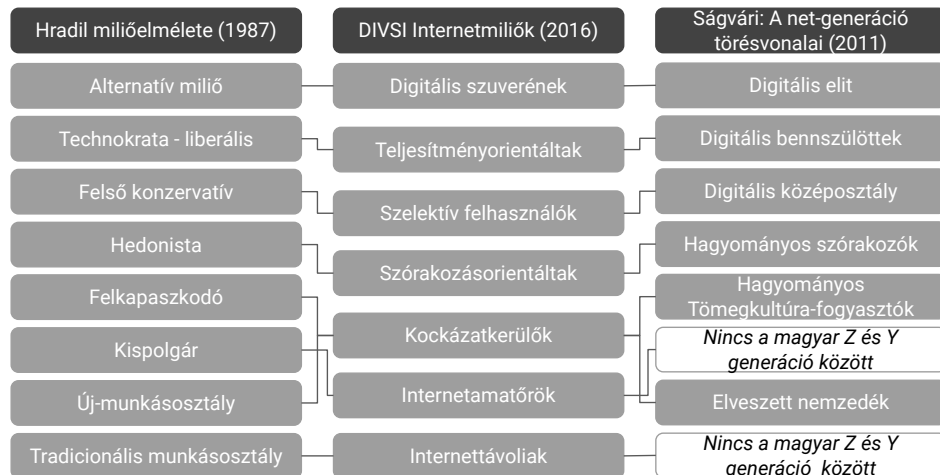


TEMATIKUS TANULMÁNYOK – Digitális szociológia

- A digitális bennszülött kategóriába tartozók átlag feletti iskolai végzettségűek, átlag feletti a státuszúak. Kiemelten aktívak online, jellemzően digitálisan tájékozódnak, informatikai eszközöket halmoznak maguk köré. Egybefolyik számukra a tanulás és a magánélet, vagy a munka és a magánélet. Nem szakadnak el az egyetemi/irodai feladatoktól szabadidejük töltése közben sem, ugyanazokat a (profi) eszközöket használják a nap minden órájában, függetlenül attól, hogy épp dolgoznak, vagy a szabadidejüket töltik. Rendkívül alacsony mértékben néznek televíziót.
- A digitális középosztály tagjaira az átlagos iskolai végzettség és átlagos státusz jellemző. Körülbelül a fele a csoport tagjainak dolgozik, harmada tanul. Közepesen digitalizáltak, szelektíven használják az internetet. Nagyvárosiak, akikre nem jellemző az olvasás, mint időtöltés, átlagos a TV-nézési idejük és kiemelten érdeklődnek a sportok iránt.
- A hagyományos tömegkultúra-fogyasztók csoportja inkább vidéki, tömegkultúra iránt affinis, nem található náluk magaskultúra iránti érdeklődés. Alacsony iskolai végzettségűek, aránylag sok körükben az inaktív, munkanélküli. Magas mértékű náluk a TV-nézés szintje.
- A hagyományos szórakozók körében kevésbé jellemzők a tanulási és szakmai ambíciók, átlag feletti a munkanélküliség aránya. Jellemző rájuk a szabadidős tevékenységek preferálása, sokan helyi kocsmákban töltik leginkább az idejüket, vagy pedig a semmittevést választják. Kevésbé sportolnak. Magas a körükben az okostelefon birtoklásának aránya.
- Az elveszett nemzedék kategóriájába tartozók iskolai végzettsége jellemzően alacsonyabb. Kiemelten jellemző rájuk a TV-nézés, internethasználat tekintetében azonban ebben a csoportban jelenik meg először magasabb arányban az a körülmény, hogy nincs otthon internet-hozzáférése a tagoknak.

Ságvári a Hradil és a DIVSI-féle milióelméletekhez hasonlóan szintén arra a módszerre hagyatkozott, hogy életstílus szerint szegmentálja a résztvevőket, a korábbi modellekben szereplő dimenziók mentén vizsgálódva, ezért lehetségesnek tűnik szegmensrendszerét a nemzetközi példákhoz kapcsolni.

TEMATIKUS TANULMÁNYOK – Digitális szociológia



4. ábra. Hradil miliőelmélete, a DIVSI Internetmiliők és a Ságvári szerinti magyar fiatal szegmensek összekapcsolása jellemzőik alapján

Forrás: Saját szerkesztés

Ságvári életstílus-alapú szegmensei kellő mélységű demográfiai és kvalitatív információt tartalmaznak ahhoz, hogy segítsék a közösségimédia-adatokból kiolvasható attribútumok megértését, és ezáltal egy ilyen típusú adatokon alapuló szegmentációt.

A magyar fiatalok miliőalapú szegmensei

Az elemzéshez felhasznált adatok nyilvános Facebook felületeken történő aktivitások, melyek a Facebook Graph API alkalmazáson keresztül voltak elérhetőek.³ A kutatást megelőzően a Facebook-felületek kvalitatív módon kategorizálásra kerültek a kommunikált témáik alapján, így felület-szegmensek lettek létrehozva belőlük érdeklődési körök, preferenciák, fogyasztási szokások mentén.⁴ Jelen kutatásban ezen felület-szegmensek segítségével kerültek strukturálásra és csoportosításra az aktivitások és a felhasználók, segítségükkel történt a Facebook-aktivitás alapú agglomeratív hierarchikus klaszterezés és az ily módon keletkezett szegmensek feltárá-

³ Az adatgyűjtés során létrehozott adatbázisban az objektumok és a köztük lévő relációk (aktivitások, felhasználók, felületek) anonim formában kerültek tárolásra. A gyűjtött aktivitások a nyilvános Facebook-oldalak által megosztott tartalmakra (poszt) való reakciók (like vagy egyéb a posztra adott reakció).
Időszak: 2017. 01. 01. – 2019. 12. 31., 2.986.683.330 db reakció.

⁴ 67 943 db publikus Facebook felület 426 felületszegmensbe lett rendezve.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

sa. Az elemzésbe bevont, a Z és Y generációknak megfelelő 14-36 éves Facebook-felhasználók leválogatása modellezési eljárással zajlott.⁵

Az adatok elemzése Python nyelvű program segítségével történt, az elemzési minta hierarchikus klaszterezési eljárással került szegmentálásra, a szegmensek feltárása és a klaszterezés eljárása vegyes módszernek mondható abból a szempontból, hogy míg az adatok (reakciók, felhasználók, felületek) gyűjtése, tárolása, az ezek közötti metszetek, átfedések vizsgálata kvantitatív és informatikai eljárásokon alapul, a felület-szegmensek kategorizációja kvalitatív kutatói folyamat. A vizsgálat során így nem elemzési céllal létrehozott adatok és kifejezetten tudományos és kutatói vizsgálati célra készített elemzési keretrendszer egyaránt felhasználásra került (Salganik 2017).

Az elemzési minta leválogatása, klaszterezése és érdeklődési körének vizsgálata a szegmensekbe előzetesen kategorizált nyilvános felületek alapján történt, összesen 67 943 darab felületen, 5 570 766 magyar Facebook felhasználó aktivitása került vizsgálatra a 2017. 01. 01. – 2019. 12. 31. időszakban. Az összesen vizsgált aktivitás 2.986.683.330 adatpont, a Z és Y generációs mintába összesen 81 399 felhasználó került.

A Z és Y generációs minta agglomeratív hierarchikus klaszterezési eljárás segítségével került szegmentálásra,⁶ ezzel a módszerrel lehetséges volt olyan, érdeklő-

⁵ Az életkor meghatározása modellezési lépéseket tett szükségessé, mert a Facebook irányelvei miatt direkt demográfiai adatokhoz nincs hozzáférés. A modellezés két eljárás segítségével történt. Egy korábbi, 2017 áprilisában zajlott kérdőíves adatfelvétel eredményei alapján rendelkezésre állt egy olyan modell, amely egy 30.000 fős adatfelvételi minta eredményeiből kiindulva a demográfiai, pszichológiai és Facebook lábnyom adatokat egymással összekötve képes arra, hogy 71%-os biztonsággal megbecsülje a felhasználók életkorát a Facebook lábnyomukból kiindulva. Az így kapott minta további finomításra került újabb aktivitásalapú modellezéssel, melyhez 1230 Facebook-felület egyedileg leválogatásra és címkézésre került kvalitatív módon aszerint, hogy kifejezetten fiatalabbaknak vagy idősebbeknek szóló tartalmakat közvetít. Ennek során a fiataloknak szóló tartalmakat közvetítő felületeken való aktivitással rendelkezők nagyobb eséllyel kerültek a mintába, a kifejezetten idősebbeknek szóló tartalmakat közvetítő felületeken aktívabbak pedig kizárásra kerültek. A vizsgálati mintába végül tehát olyan Facebook-felhasználók kerültek, akik digitális aktivitásuk alapján a Z és Y generációhoz hasonló érdeklődésűek, és 2017. 01. 01. – 2019. 12. 31. között aktív Facebook felhasználók adják, összesen 189 512 fő, melyből leválogatásra került az elemzéshez kellő mértékű aktivitással rendelkező 81 399 fő.

⁶ Az agglomeratív hierarchikus klaszterezési eljárás egy iteratív eljárás, melynek egy általános lépésében egy kiválasztott módszer által meghatározott páronkénti távolságok közül azt a párt fogja új alkszterként kezelni (az általuk reprezentált felhasználókat uniózni), melynek a módszer alapján meghatározott távolságot minimalizálja (Lior – Maimon 2005).

A klaszterezési eljárásban szükség van egy távolságfüggvényre, ami a felhasználók közötti hasonlóságot méri. Erre a legalkalmasabb a koszinusztávolság (utóbbi nem más, mint a mátrix két sora között, mint vektorok által, bezárt szög koszinuszának 1-ből kivont értéke (Lior – Maimon 2005). A már említett módszerként, ami azt határozza meg, hogy egy adott lépésében az iterációnak mely alkszterekből alkot uniókat, a Ward-féle módszer került felhasználásra (ez az eljárás az alkszterpárokban szereplő userok között fellépő páronkénti távolságok varianciáját minimalizálja (Ward 1963). Az eljárás eredménye egy dendrogram, amelynek segítségével meghatározható, hogy hány klaszter kerüljön további vizsgálatra. A vizsgált klaszterek számának meghatározásában a dendrogram szociológiai értelmezése, azaz az szegmentált adatstruktúrát tekintve interpretálható klaszterek kiválasztása fontos szerepet játszik.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

désüket tekintve homogén szegmensek feltárása, amelyek tulajdonságaik mentén megfeleltethetők a korábban vázolt életstílus-modellek kategóriáinak. Ehhez az adatalapon szétváló szegmensek leíró jellemzésére volt szükség,⁷ amely alapján a vizsgált szegmensek különböznek egymástól az őket alkotó Facebook-felhasználók *további* érdeklődése szerint. A további érdeklődési körök jól mutatják a szegmensek szubkulturális és életstílus karakterét, és segítik azt, hogy a fiatalok mintáján életstílus alapú szubkulturális szegmenseket azonosítsunk be.⁸

A 2016-os, teljes német társadalomra vonatkozó DIVSI-kutatás és a magyar ifjúság szegmenseit vizsgáló 2011-es Ságvári-tanulmányhoz hasonlóan a mintázatok alapján a szegmensek mentén jól karakterizálható csoportok kerültek kialakításra. Majd annak vizsgálata történt, hogy a korábbi kutatások által feltárt szegmensek közül az ezekkel közös tulajdonságaik alapján melyek jelennek meg a jelen tanulmány keretében vizsgált vizsgált Z és Y generációs felhasználók körében.

Az eredmények alapján a Z és Y generációs elemzési minta jól szegmentálható, a kirajzolódó szegmensek érdeklődésük alapján hasonlóak mind a 2016-os német, mind a 2011-es magyar kutatás szegmenseihez, tehát az aktuális magyar közösségi-média-alapú adatok alapján feltárt csoportokmintázatok egybecsengenek az elmúlt évtized német és magyar kutatási eredményeivel.

Az adatok alapján öt szegmens került beazonosításra. A szegmensek ily módon történő feltárása és validálása után annak vizsgálata következett, hogy a korábbi kutatások által feltárt szegmensek és miliók közül közös tulajdonságaik alapján melyek jelennek meg a vizsgált fiatal magyar célcsoport körében.

⁷ A szegmensek leíró jellemzéséhez létrehozott táblák az alábbi információkat tartalmazzák: az elemzésre került felület-szegmenseket és felületeket, a felület-szegmenshez vagy felülethez tartozó összes Facebook-felhasználó számát, továbbá a vizsgált célcsoport metszetét a felületekkel vagy a felület-szegmensekkel (aktivitási arány). Ez utóbbi azt fejezi ki, hogy a célcsoportból hány aktív felhasználó van az adott felületen, vagy felület-szegmensben. Ezen kívül a felület vagy a felület-szegmens méretét (aktív felhasználóinak számát) is figyelembe vevő mutató került felhasználásra annak kifejezésére, hogy a célcsoportba tartozó felhasználók az átlagos magyar Facebook-felhasználónál mely felületen, vagy felület-szegmensben (vagy akár adott Facebook felületen) aktívabbak, és melyek iránt mutatnak kevesebb érdeklődést. Ezt a mutatót a felület, vagy felület-szegmens szegmens és a szegmens tagjainak számát, illetve azok metszetét veszi figyelembe, a megszokott várhatóérték-számítás módszerével. A mutató mértékegység nélküli, standardizált szám. A szegmensek és a felületek ily módon keletkező sorrendje alapján beszélhetünk preferenciákról, érdeklődési körökről, attitűdökről a szegmensben belül.

⁸ A beazonosított szegmensek feltárása érdekében a szegmensekre leginkább jellemző érdeklődési témakörök felület-szegmensek és a körükben leginkább felülreprezentált Facebook felületek, és ezek konkrét témája került vizsgálatra. A vizsgálathoz a feljebb tárgyalt legnagyobb bevonódási arányszámok szerint csökkenő sorrendben kerültek listázásra a szegmensre leginkább jellemző, a szegmens összetételéhez a legtöbb aktív felhasználót adó, a vizsgált időszakban legalább 100 aktív felhasználóval rendelkező Facebook felületek.



TEMATIKUS TANULMÁNYOK – Digitális szociológia



5. ábra. A feltárt szegmensek összekapcsolása Hradil miliőelmélete, a DIVSI internetmiliók és a Ságvári szerinti magyar fiatal szegmensek jellemzői alapján

Forrás: Saját szerkesztés

A szegmensek Facebook lábnyomának szegmens- és felületszintű vizsgálata alapján feltárt öt szegmens és fő jellemzőik az alábbiak:

- **Digitális szuverének:** Jellemzően az alternatív kultúrához kapcsolódó értelmiségi fiatalok, akik viszonylag stabil anyagi jólétben élnek. Kifejezetten progresszív és posztmaterális értékrendet követnek, társadalmilag, szociálisan érzékenyek. Magasan kvalifikáltak, erős a humán érdeklődésük, karrier szempontból ambiciózusabbak, az olvasás fontos számukra. Elsősorban a fővárosban és környékén magas az arányuk. Aktív típusok, kiemelt kultúrafogyasztók, szervesen kapcsolódnak a közéleti kérdésekhez, nemzetközi szinten tájékozódnak, tudatos médiafogyasztók, idegennyelv-tudásuk kiemelkedő. Tudatosan, céljaikhoz igazítva használják a digitális megoldásokat, eszközként tekintenek rá, céltudatosan és otthonosan mozognak az online térben. Sok időt töltenek online, de inkább célzottan használják a netet, sok offline tevékenységük, programjuk van. Nem kiemelten technológia-affinisek, eszközhasználatukat a praktikusság és feladatorientáltság határozza meg. A digitális felületeket elsősorban tájékozódásra, ügyintézésre és kapcsolattartásra használják, szórakozás céljából színvonalas kulturális tartalmakat keresnek.
- **Teljesítményorientált digitális benszülöttek/technokraták:** Vállalkozó szellemű, szakmai világhoz kapcsolódó, önfejlesztő, magas státuszú, jó anyagi helyzetű fiatalok. Inkább liberális értékrend szerint élnek, kiemelten fontos számukra a környezettudatosság. Magasan kvalifikáltak, inkább műszaki irányban, karrier szempontból ambiciózusabbak. Elsősorban a fővárosban és nagyvárosokban magas az arányuk. Életük legtöbb területén tudatosak (étkezés, közlekedés, szakma stb.), törekvők, fontos számukra a közösség elismerése. Gyakran folyik egybe náluk a munka-tanulás és a magánélet. Digi-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tális bennszülöttek, életük legtöbb területe digitalizált, szinte folyamatosan kapcsolódnak valamilyen virtuális platformhoz, eszközhöz, megoldáshoz, digitális profizmus jellemzi őket. Aktív de egyben tudatos nethasználók, fontos számukra a digital well-being, képesek arra, hogy szakmai céljaikhoz igazítsák digitális szokásaikat. Erősen tech-affinisek, innovatív digitális eszközöket használnak, érdeklí őket a technológiai eszközök háttere. A digitális világ legalább annyira valóságuk része, mint az offline világ, a kettő összefonódik számukra.

- *Netfüggő/hedonisták:* Átlagos vagy átlag alatti státuszú, kevésbé önreflexív, jellemzően kevésbé tudatosan élő fiatalok. Értéksemlegesek, kifejezetten élmény- és szórakozásorientáltak, olyan konzumerek, akik könnyen feloldódnak a trendekben. Átlagosan kvalifikáltak, kevésbé jellemzőek rájuk a kiemelkedő szakmai vagy tanulmányi ambíciók. Elsősorban városokban és nagyvárosokban magas az arányuk. Intenzív online tartalomfogyasztók zene, film, gaming témákban. A Z generáció esetén gyakori tevékenység a „lógás” és „semmittevés”, az Y generációnál pedig a közösségi bulizás, alkoholfogyasztás. A digitális világ hétköznapi valóságuk része, a nem digitális világgal összefonódik számukra. Hálófűggők vagy nem tudatos nethasználók, konzumerek, érdeklődnek az új eszközök iránt, de a technikai háttér nem érdekli őket, a trend és az élményfaktor határozza meg irányultságukat. Érdeklődésük központjában a szórakozás áll, így az ehhez kapcsolódó felületeket látogatják leginkább.
- *Kötelességtudó szelektív nethasználók:* Átlagos vagy átlag feletti státuszú fiatalok, akik jellemzően polgári és normakövető családi háttérrel rendelkeznek. Tradicionális-konzervatív értékrend szerint élnek, fontos számukra a nemzet, a család és a hagyományos társadalmi szerepek megőrzése. Átlagosan vagy átlag felett kvalifikáltak. Városokban magas az arányuk. Érdeklődésük központjában a sport és a közélet áll, nem kifejezetten kultúraaffinisek. Fontosak számukra a nem online személyes kapcsolatok. Körültekintően állnak az internethez, szelektív használatként kevésbé vonódnak be az online világba, könnyen el tudnak szakadni a netről, fontos nekik az online biztonság. Limitáltan proaktívak, megszokott, rutinszerű tevékenységeiket végzik online, kevesebb platformot használnak. Nem kifejezetten érdeklik őket az újdonságok, kevésbé technológia-affinisek. A klasszikus digitális termékeket használják elsősorban, mint a TV, konzol, laptop, telefon, a digitális teret leginkább tájékozódásra és kapcsolattartásra használják.
- *Hagyományos tömegkultúra fogyasztók:* Átlagos vagy átlag alatti státuszú, konzumer, mainstream szemléletű, kevésbé tudatos, Y generációs fiatalok. Értékvesztettek, nincs határozott világnézetük, jellemzően a populáris trendekkel sodródnak. Átlagosan vagy alacsonyán kvalifikáltak. Jellemzően vidéki községekben magas az arányuk. Nem szociálisak, intenzív TV-nézők, tömegkultúrát fogyasztanak. Fontos számukra a testedzés, ahogy otthonuk, autójuk megjele-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

nése is. Árérzékenyek, fogékonyak a reklámokra. Nem mozognak otthonosan az online térben, kevésbé aktívak a digitalizált világban. A digitális eszközöket egy-két „klasszikus” funkcióra használják, nem affinisak a technológiai vívmányokra. Inkább a klasszikus digitális eszközöket ismerik és használják, ilyenek a háztartási gépek és a klasszikus tömegkultúra csatornái, mint a TV és a rádió. Az online térben inkább információkat keresnek vagy kapcsolattartásra használják.

Összegzés

A közösségimédia-adatokon végzett, a digitalizáltságot, mint szempontot beemelő vizsgálat eredményei arra utalnak, hogy a magyar Z és Y generáció körében változások figyelhetők meg a digitalizáltság és az életstílus szempontjából, és tulajdonságikat tekintve a nemzetközi modellekéhez hasonló miliőcsoportok (szegmensek) rajzolódhatnak ki. Az eredmények alapján a Z és Y generáció körében a digitalizáltság foka összefügg a társadalmi státusszal és a posztmateriális értékek iránti nagyobb fokú elköteleződéssel is. Körükben teljes mértékben hiányzik a nem digitalizált, tradicionális értékrendű, alacsony státuszú szegmens, és az „internetamatőr”, „kispolgári” szegmens sem jelenik meg markánsan adatlábnymuk alapján. Az értéksemleges, alsó és középosztályhoz tartozó hagyományos tömegkultúra-fogyasztók már csak az Y generációban azonosíthatók be adataison, és már nem jelennek meg a 23 év alattiak körében, akik jellemzően intenzíven használják a digitális eszközöket.

Tanulmányomban azt a kérdést vizsgáltam, hogy a magyar fiatalok milyen életstílus-alapú szegmensekre bonthatóak közösségimédia-aktivitásaik alapján. Hibrid és innovatív módszertani megoldásokat alkalmazva, konkrét életsílus-modellek segítségével, a digitalizáció dimenziójának beemelésével megkíséreltem az adataison kialakított szegmensek feltárását, és a magyar Z és Y generáció körében a nyugati fiatalokéhoz hasonló miliőket és értékváltozásokat kerestem. Bár a vizsgálat során a nyugati modellekéhez hasonló mintázatok rajzolódtak ki, miközben akár relevánsnak is tűnő változások láthatóak a fiatalok körében, mégis érdemes körültekintően kezelni ezeket az összehasonlításokat, és fontos lenne akár más módszertani eszközök segítségével tovább vizsgálni a kutatási témát. Kérdéseket vehet fel az a tény is, hogy miközben vizsgálat a magyar fiatalokra fókuszál, a segítségül használt modellek eltérő körülmények között, a teljes német társadalomra lettek kidolgozva. Bár az interpretáláshoz egy hasonló szemléletű, kifejezetten a fiatalokra fókuszáló magyar kutatás is feldolgozásra került, a részben eltérő rétegződésű nyugati jóléti társadalmat vizsgáló modellek csak korlátozottan hasonlíthatók össze a magyar fiatalok attitűdjein alapuló szegmentálással. Az újszerű, közösségimédia-adatokon alapuló kutatási megközelítés is felvetheti a kérdést, hogy az eltérő módszertan mennyiben mutat más eredményeket. Bár az eltérő módszertanok összehasonlíthatósága korlá-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tozott, a közösségimédia lábnyom alapú eredmények olyan tudományos utat jelölnek ki, amely mentén érdemes tovább vizsgálni a későbbiekben. A digitális aktivitásból rekonstruált miliőszegmensek számos, erre a tematikára fókuszáló további kutatás lehetőségét teremthetik meg, miközben az összefüggések mélyebb feltárása további, más módszertannal és adatforrással dolgozó kutatásokat is indokol.

Irodalom

- Angelusz R. (2000): A láthatóság görbetükrei. Társadalompolitikai tanulmányok. Új Mandátum Könyvkiadó, Budapest
- Beck, U. (1999): Túl renden és osztályon? In: Angelus, R. (szerk.): A társadalmi rétegződés komponensei. Új Mandátum Könyvkiadó, Budapest
- Beck, U. (2003): A kockázat-társadalom. Út egy másik modernitásba. In: Andorka R. (szerk.): Andorka Rudolf Társadalomtudományi Társaság és Századvég Kiadó, Budapest
- Beck, U. (1997): Túl renden és osztályon? Társadalmi egyenlőtlenségek, társadalmi individualizációs folyamatok és az új társadalmi alakulatok, identitások keletkezése. In: Angelusz R. (szerk.): A társadalmi rétegződés komponensei. Új Mandátum Könyvkiadó, Budapest
- Bolte, K. – Hradil, S. (1984): Sozial Ungleichheit in der Bundesrepublik Deutschland. Leske-Budrich, Opladen
- Bolte, K. (2012): Die Sinus Milieus in der VuMA. https://www.manfred-bolte.de/pdf/Sinus_Milieus_in_VuMA_2012.pdf (utolsó elérés: 2021. 11. 20.)
- Deutsches Institut für Vertrauen und Sicherheit im Internet (2016): DIVSI Internet-Milieus 2016 Die digitalisierte Gesellschaft in Bewegung. DIVSI, Hamburg. <https://www.divsi.de/wp-content/uploads/2016/06/DIVSI-Internet-Milieus-2016.pdf> (utolsó elérés: 2021. 11. 20.)
- Hellmann, A. (2014): SINUS-Milieus treffen die netzstrategen. <https://netzstrategen.com/blog/sinus-milieus-treffen-die-netzstrategen> (utolsó elérés: 2021. 11. 20.)
- Hradil, S. (1995): Régi fogalmak és új struktúrák. Miliő-, szubkultúra- és életstílus-kutatás a 80-as években. In: Andorka R. – Hradil S. – Peschar J. (szerk.): Társadalmi rétegződés. Aula Kiadó, Budapest
- Hradil, S. (2010): Társadalmi helyzetek és miliók. Egy fejlett társadalom struktúrájának elemzése. In: Angelusz R. – Gecser O. – Éber M. Á. (szerk.): Társadalmi rétegződés olvasókönyv. Eötvös Loránd Tudományegyetem, Budapest, DOI: 10.20311/stat2018.05.hu0468
- Kemp, S. (2021): DIGITAL 2021: HUNGARY. Kepios Pte. Ltd., Szingapúr. <https://data-reportal.com/reports/digital-2021-hungary>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Linke, K. (2017): Analysis of Socio-Spatial Differences in Germany for the Definition of Online Milieus. *Journal of International Business Research and Marketing*, vol. 2 (No. 6): 37-45. DOI: 10.18775/jibrm.1849-8558.2015.26.3005
- Lior, R. – Maimon, O. (2005): Clustering methods. In: Lior, R. – Maimon, O. (szerk.): *The Data Mining and Knowledge Discovery Handbook*. Springer, New York, DOI:10.1007/0-387-25465-X_15
- Luhmann, N. (2006): *Bevezetés a rendszerelméletbe*. Gondolat Kiadó, Budapest
- Ságvári B. (2011): A net-generáció törésvonalai – kultúrafogyasztás és életstílus-csoportok a magyar fiatalok körében. In: Bauer B. (szerk.): *Arctalan (?) nemzedék: Ifjúság 2000-2010*. Nemzeti Család- és Szociálpolitikai Intézet, Budapest
- Salganik, M. J. (2017): *Bit By Bit: Social Research in the Digital Age*. Princeton University Press, Princeton
- Schulze, G. (2003): A Német Szövetségi Köztársaság kulturális átalakulása. In: Wesely A. (szerk.): *A kultúra szociológiája*. Osiris Könyvkiadó, Budapest
- Schulze, G (1992): *Die Erlebnisgesellschaft*. Kultursoziologie der Gegenwart. Campus, Frankfurt am Main, DOI: 10.25656/01:10445
- Simon, K. (2021): *Digital 2021: Hungary* <https://datareportal.com/reports/digital-2021-hungary> (utolsó letöltés: 2021. 11. 20.)
- Strauss, W. – Howe, N. (1991): *Generations. The History of America's Future, 1584 to 2069*. William Morrow and Company Inc, New York
- Székely L. (2013): *Magyar Ifjúság 2012 Tanulmánykötet*. Kutatópont, Budapest
- Törőcsik M. – Szűcs K. – Nagy Á. – Lázár E. (2019): Életstíluscsoportok Magyarországon a digitalizáció korában. *Replika*, 2019 (111): 63-8. DOI:10.32564/111.5
- Ward, J. H., Jr (1963): Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* Vol. 58 (No. 301): 236–244. DOI: 10.2307/2282967



Közösségi média kommunikáció a digitális egészségügyi térben

A #cysticfibrosis és a #Asthma Big Data összehasonlító elemzése

SIMON SÁRA¹

ABSZTRAKT

A 21. század technológiai környezetében egyre nagyobb jelenléttel képviseltetik magukat a tudatos, információszerző betegek, a korábbi alá-fölérendeltségi viszonyt elkezdte átváltani a kooperáció alapú együttműködés, amelyet számos digitális egészségügyi projekt segít (Koskova 2015). A nagy számban elérhető orvosi információ pedig lehetővé tette, hogy a betegek tájékozódjanak, másodvéleményt kérjenek, orvosaiknak segítsenek és tudatos e-pácienssé váljanak (Meskó et. al. 2017).

Különösen igaz a ritka betegségben érintettekre, ahol az információ megléte még nagyobb jelentőséggel bír, Európában már több mint 30 millió érintett életében. Ezen betegségek esetében a diagnosztikus út sok esetben hosszú, ennek következtében a megszerzett tudásanyag segítséget adhat a megküzdésben, ehhez pedig segítséget adhatnak az online felületek is. A kutatás ezért a közösségi médiában történő orvosi információ szerzését vizsgálta social média listening alkalmazásával. Az #Asthma és a #CysticFibrosis Twitter alapú adatbázisok elemzése hét éves intervallumot foglalt magába, feltáró jellegű vizsgálattal a tweet aktivitás, a szemantikai környezet, a véleményvezérek, valamint az elkötelezettség és a hashtag használat témakörében egyaránt.

KULCSSZAVAK: digitális egészségügy, e-páciensek, Big Data, tisztás fibrózis, asztma

ABSTRACT

Social media communication in the digital medical space

In the environment of 21st century technology, the transformation of information acquisition of health care and patients has had an increasing emphasis. Despite the earlier authoritative doctor-patient relationship, a need for an equal, cooperation-based communication has emerged and there are so many digital healthcare projects to achieve this (Koskova 2015).

Information acquisition on the internet has allowed patients that based on the increasingly available medical information they acquire information about their condition, become part of patient communities, ask for second opinions, and become committed helpers of their doctors in their disease (Meskó et. al 2017).

¹ ELTE Szociológia Doktori Iskola, Interdiszciplináris Társadalomkutatások Doktori Program, simonsara.mail@gmail.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

This can be especially true for patients with rare diseases, where a diagnosis might take even a decade, the patient needs lifelong condition maintenance and treatment, if it is available. While the proportion of patients with rare diseases is low compared to the whole of society, the number of such patients is approximately 30 million in Europe (EURORDIS), which means patients and their relatives need not only a harmonized health care system, but extensive information so that they can live with the rare disease with less difficulty.

The aim of our study was to present the options of information acquisition in the social media, focusing on Twitter, via an interdisciplinary and social approach. In this study therefore we carried out a Big Data based social media analysis based on #Asthma and #CysticFibrosis databases of the Symplur corporation. This study results contain the complete online communication of 7 years (2012-2019) regarding these hashtags. The analysis has few levels including semantic research, stakeholder and hashtag review, engagement, and the whole tweet activity exploration.

KEYWORDS: digital healthcare, e-patients, Big Data, cystic fibrosis, asthma

Bevezetés

A digitális szféra térhódítása az egészségügyben egyre nagyobb mértékben figyelhető meg. Az elérhető technológiai fejlesztések hatással vannak az orvos-beteg viszonyrendszerre és az ellátásra, valamint arra kommunikációs mezőre is, amely a területet övezi. A digitális egészségügy komplexitásában a tudatos, önmagukat menedzselő páciensek egyaránt képviseltetik magukat, ahogy az online tér és a megosztott információ is döntő szereppel bír ebben a folyamatban.

Az orvostudomány mellett a technológia fejlődése is magával hozta az online formában történő információszerzés és az orvosi tudásanyag elérését, ezzel párhuzamosan egyre növekvő számú páciens igényét arra vonatkozóan, hogy az alá-főlérendeltségen alapuló orvos-beteg kapcsolatot felváltsa a partneri viszony, amely magában foglalja a kölcsönös információcserét és a páciensek egészség- és betegségmentésmentjét is.

A tanulmány ezért az egyenrangú, partneri alapon működő kapcsolatot veszi alapul az orvos-beteg reláció megközelítésekor, nem tárgyalja a paternalisztikus modellt, amely tekintélyelvű, az ellenőrzés az orvos kezében van és aszimmetria áll fenn, ez pedig fokozhatja a tévedés lehetőségét orvosi oldalról, valamint a betegek kooperációja is jóval alacsonyabb mértékű lehet, ha nem lehetséges a közös döntéshozatal és bevonódás a saját egészségüket illetően. Nem tárgyalja továbbá a fogyasztói modellt sem, ahol a betegek üzleti megközelítéssel szolgáltatásvásárlási céllal közelítik meg a gyógyítás folyamatát (például az asszisztált reprodukció folyamatában, ahol akár elfogadható attitűd is lehet az orvosi tudásanyag ilyen módon történő felajánlása), saját kezükben tartva a kontrollt, amely szembe mehet a szakmaisággal, továbbá az egészségügyi költségeket indokolatlanul növelhetik (Pilling 2017).

A tanulmány tehát nem az orvos-beteg kapcsolatra, annak történelmi és szociokulturális lépcsőfokaira fókuszál, hanem a tudatos, egyenrangú beteg esetében azok-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ra az információszerzési mintázatokra koncentrálnak, amelyek a közösségi médiában figyelhetők meg, Big Data módszertan alkalmazása mellett. A tanulmánynak nem célja az egészségügyi ellátás finanszírozási formáinak ismertetése sem hazai, sem nemzetközi környezetben, ugyanakkor valóban nem hagyható figyelmen kívül az sem, hogy a finanszírozási formák, esetlegesen a magánellátás adta megnövelt időkeret a betegek ellátására, valamint a kooperatív és innovatív eszközök alkalmazása (betegség-monitorozás, digitális alapú kapcsolattartás, leletértékelés) ezt a partneri kapcsolatot elősegítheti, tehát az, hogy a beteg milyen finanszírozási formában veszi igénybe az adott ellátást, hatással lehet arra, hogy melyik modell tud megvalósulni, illetve a megszerzett információk közlése lehetséges-e, az orvos számára is segítő célú, vagy ellenérzést vált ki.

A szerző szándéka tehát az, hogy bemutassa a digitális egészségügy fogalmi körét és annak jellegzetességeit, valamint azokat betegeket helyezze középpontjába, akik önálló betegségmentesítést folytatnak egészségi problémájuk megoldása érdekében, továbbá részt vesznek a közösségi média kommunikációban betegsükhöz kapcsolódva. Az elméletet Big Data alapú kutatási eredmények egészítik ki, amelyek a közösségi interakciót hivatottak bemutatni egy ritka (cisztás fibrózis) és egy gyakori betegség (asztma) esetében, feltáró jellegű elemzés keretében. A kutatási eredmények a világjárvány előtti időszakra reflektálnak, amely szándékos törekvés, tekintve a pandémia jelenlegi fennállását és az orvosi területre gyakorolt folyamatos változást eredményező hatását, amely egységesen még nem írható le. A bemutatott eredmények ugyanakkor összefoglalóan láttatják a járvány előtti időszakot exploratív jelleggel, ezáltal pedig alapot adhatnak arra, hogy a folytatásban a járvány alatti és azt követő időszakokkal összevethetőek legyenek a későbbiekben.

Ahhoz azonban, hogy a digitális egészségügy és az információgyűjtés online formáiról és tendenciáiról beszélni lehessen, tisztázni kell mindenekeelőtt, hogyan tudjuk ezt az újszerű környezetet meghatározni, továbbá ebben elhelyezni a tudatos pácienseket és a kutatási eredményeket.

Digitális egészségügy – Hogyan írható le a világ egyik legdinamikusabban fejlődő területe?

A digitális egészségügy fogalmának meghatározásakor azonban nemcsak a gazdasági, hanem a kutatási, szabályozási környezetet is érdemes áttekinteni. A digitális egészségügy meghatározásával kapcsolatos összetett aspektusokra Fatehi és munkatársai mutatnak rá, akik 1527 kutatási találat elemzése mellett 95 különböző definíciót azonosítottak (Fatehi et. al. 2020). Ebből az amerikai The US Food and Drug Administration (FDA) a digitális egészségügyet a mobile health, a telehealth, a viselhető eszközök köre, a telemedicina, valamint az egészségügyi információs technológiák, továbbá a személyre szabott medicina összességével jelöli. Az Egészségügyi



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Világszervezet (WHO) meghatározásában pedig a digitális egészségügy hozzájárul a Fenntartható Fejlődési Célok eléréséhez, annak révén, hogy elérhetővé teszi a magas szintű egészségügyi és jóléti szolgáltatásokat szerte a világban.

A fenti nevezetett szerzők friss szakirodalommal kapcsolatos kutatása hangsúlyozza azt is, hogy nincs egységes, elfogadott definíció, melyet az áttekintő elemzésük² is ismertet. A vizsgálat során 855 peer-reviewed és 672 weboldal került azonosításra, és ennek eredményeként látható vált, hogy a cikkek tartalma és a meghatározások szerint az egészségügyi vonatkozás erősebb volt a fogalomhoz kapcsolva, mint a technológia, amely egyben humáncentrikusabb a csak elektronikus és adat alapú megközelítéssel szemben. A cél ugyanis a betegek és az ellátórendszer állapotának és helyzetének javítása, a gyógyítási lehetőségek és eszközök szélesebb körének alkalmazása, és az új digitális lehetőségek révén olyan, akár földrajzi területen történő használata, amely a hagyományos orvos-beteg kapcsolat során nem lenne megvalósítható, gondolva itt a telemedicina eszköztárára is.

Amennyiben tovább bővítjük a meghatározási mezőt, Steinhubl és Topol hozzátesszik, hogy az individualizáció, a személyre szabhatóság, az okostelefonokkal generált adatok világában az egyén egészségügyi szerepe olyan technológiai dimenzióban is értelmezendő, mely az egészségügy demokratizációjához vezethet, fokozva az elkötelezettségét az érintetteknek, miközben a digitális egészségügy eszköztára segíthet abban, hogy a betegek elégedettsége javuljon, a költségek csökkenjenek, valamint az ellátórendszer többi szereplője számára is jobb és hatékonyabb környezetet biztosítson. Fontos azonban figyelembe venni azt is, hogy szektorban ugyanúgy feltűnhet a „hype” jelensége, ezért az ágazat szereplőinek kommunikációja és a szabályozási koncepcióknak éppolyan fontos szerepe van (Steinhubl – Topol 2018).

A digitális egészségügy meghatározása mellett azonban szükséges tárgyalni azt is, hogy miként lehetséges az adaptációja, melynek mozgatórugója egyrészt a piaci környezet és a technológiai ellőállítói oldal, másrészt a felhasználói kör, ugyanakkor a humán komponens és edukációs igény miatt az elterjedési köre változó és egyenlőtlenségeket is mutat.

Jiminez és munkatársainak vizsgálatában az egészségügyi alapellátás állt, azon belül is a digitális egészségügyi kompetenciák³ alkalmazása (Jiminez et al. 2020). A 2020-ban megjelent irodalmi áttekintés jelentősége abban áll, hogy a mindeddig nem elemzett kutatási eredményeik szerint szakadék áll fenn az alapszintű ellátásban használt készségek és az elérhető lehetőségek között, miközben csaknem 17 különböző kompetencia volt azonosítható, amelyek használhatóak lennének az ellátásban. A publikált irodalom 54 százaléka azonban 2005 előtti, ezért hangsúlyozzák, hogy nemcsak a gyakorlati bevezetés, az alapszintű készségek növelése a cél,

² Web of Science, Scopus, Google Scholar, PubMed 2000–2020 között

³ Digital Healthcare Competencies – DHCs



TEMATIKUS TANULMÁNYOK – Digitális szociológia

hanem a tudományos diskurzus is rendkívül fontos lenne, annak érdekében, hogy az új technológiai lehetőségek a fent említett definíció mentén a gyakorlatban is meg tudjanak valósulni.

Mindezek együttese figyelmeztet arra, hogy a technológiai tér dominanciája mellett fontos lenne ezen tudományos ismeretanyag létrehozása és gyakorlása, nem elegendő az alkalmazás és alkalmazhatóság az orvosi praxisban, szükséges ezeknek módszeres elméleti összefoglalása és konzekvenciák levonása annak érdekében, hogy szakszerű és kidolgozott válasz legyen adható a gyógyítás oldaláról is a technológiai kihívásokra.

Orvostudomány egy technológiai és piaci környezetben

A digitális egészségügy vonatkozásában a társadalomtudományi és elméleti kontextuson kívül, érdemes figyelembe venni azt is, hogy rendkívül élénk és nagyforgalmú világpiacról van szó, tehát a végbemenő változásokat nemcsak tudományos, hanem piaci körülmények is befolyásolják.

A globális digitális egészségügy piaci értéke a VynZ Research közlése szerint⁴ ugyanis 2019-ben már elérte a 111,4 millárd dollárt, a VNZ Research elemzése szerint ez az összeg 510,4 millárd dollárra nőhet 2025-ig, miközben a digitális egészségügy piacához szervesen kapcsolódik az mHealth technológia, az Electronic Health Records (EHR), valamint a telemedicina területe is. A végfelhasználók körében kettéosztható a felhasználási terület, kórházi, valamint egyéni felhasználók szerint csoportosítva. A digitális egészségügy legnagyobb piaca Észak-Amerika volt 2019-ben, ugyanakkor egyre feltörekvőbb jelenléttel képviselteti magát a kínai, valamint az indiai piac is, az említett öt éves periódus legnagyobb arányú növekedési várakozása előtt áll, magában foglalva jelentős számú beteget és alacsonyabb orvos-beteg arányt, miközben a technológiai előretörés megkérdőjelezhetetlen.

Amennyiben csoportosítani szeretnénk a digitális egészségügy piacán elhelyezkedő szereplőket, kategorizálni kell alcsoportok szerint is. Az mHealth magában foglalja a csatlakoztatott eszközöket, alkalmazásokat és szolgáltatásokat is. Az EHR egyaránt lehet felhőalapú, miközben a keletkezett adatok elemzése révén mind előrejelző, mind deskriptív Big Data elemzések egyaránt végezhetőek rajtuk, továbbá a telemedicina a távkonzultáció, a monitoring, az oktatás és orvosképzés lehetőségeit egyaránt tartalmazza.⁵

A Statista közlése szerint (1. ábra) csak az mHealth piaca 2016-ban 23 milliárd dollárt ért, várakozások szerint ez közel 190 millárd dollárra fog nőni 2025-re veze-

⁴ www.globenewswire.com

⁵ www.globenewswire.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tő ágazati húzóerőként, miközben a teljes digitális egészségügyi piac 2021-ig, 2020. szeptemberi elemzések szerint el fogja érni 280 millárd USA dollárt.

Kitekintő jelleggel érdemes megemlíteni, hogy a különböző applikációk formájában nemcsak külföldi fejlesztések érhetőek el a felhasználók számára ezen a területen, hanem fontos kiemelni Magyarországon az EESZT rendszerét, amely az egészségügyi adatok és az EHR szektor kiemelt példája hazánkban, de a szintén magyar Medcall⁶ applikáció és egészségügyi szolgáltató cég is lehetővé teszi a digitális egészségügy vívmányainak elérését is, valamint a felírható e-receptek elérhetősége is jelen van már.

Global digital health market size between 2015 and 2025, (in billion U.S. dollars)

	2015	2016	2017	2021	2022	2023	2024	2025
Health IT	-	134	-	280	-	-	-	-
mHealth	-	23	-	-	-	-	-	189
Wearables	0.28	-	-	-	-	-	-	22.3
EHR/EMR	-	23	-	-	-	-	-	38
Personal genomics	-	-	15	-	-	-	-	5
Population Health Management	-	9	-	-	-	-	-	50
Medical Imaging	-	53	-	-	-	74	-	-
Telehealth	-	2.5	-	-	12	-	-	-
Gamification in Healthcare	-	16	-	-	-	-	40	-

Showing entries 1 to 9 (9 entries in total)

1. ábra. Globális digitális egészségügyi piac 2015-2025 között
Forrás: Statista 1

Annak érdekében, hogy a digitális egészségügy bemutatása ne csak az angolszász és az amerikai piacra fókuszáljon, érdemes a kelet-közép-európai régió országait is figyelemmel kíséreni. Ćwiklicki munkatársaival arra kereste a választ, hogy a digitális egészségügy implementációja hogyan történhet meg a leginkább sikeres módon ezekben az országokban, ahol nemcsak piaci hatások, hanem társadalmi és politikai változások is végbementek az elmúlt évtizedekben. Kutatásukban nyolc pillért azonosítottak ebben a folyamatban, beleértve a finanszírozást, az állami szerepvállalást, egészségpolitikát, az adatok feldolgozását és elektronikus rendszerekbe való illesztését, a résztvevő ágazati szereplők oktatását és az erőforrásokat egyaránt. A szerzők kiemelik, hogy a Big Data módszer és az adatok kulcsfontosságúak a digitális egészségügy alkalmazásában. Vizsgálatukban 10 országról tettek megállapításokat:

⁶ Éjjel-nappal elérést biztosít a magyar applikáció több mint 50 szakterület 500 orvosához telefonon és videó konzultáció keretében történő eléréshez, továbbá biztosítási csomagok révén folyamatos rendelkezésre állás és a magánszektorban igénybe vehető ellátás is igényelhető. <https://medcall.cc/>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Bulgária, Csehország, Észtország, Horvátország, Litvánia, Lettország, Lengyelország, Magyarország, Románia, Szlovénia.

Az országok közül kiemelendő néhány, így Csehország esetében elmondható, hogy a jogi környezet megteremtésében komoly előrelépések történtek, lehetséges e-recept kiváltása és elektronikus módon történő felírása, amely a betegek részéről is preferált, ugyanakkor az e-mailen történő egyeztetés az orvosokkal ezügyben leginkább Észtországban fejlett. Bulgáriában az elektronikus egészségügyi rendszer, az e-receptek, valamint az online konzultáció még nem került bevezetésre, az ezzel kapcsolatos nemzeti stratégia azonban elfogadásra került már. Magyarország kapcsán a szerzők a folyamatos fejlődést és a megvalósult EESZT rendszert emelik ki 2017. november 1-től, melyhez a gyógyszertárak is csatlakoztak. Lengyelország esetében pedig lassabb adaptációs folyamatot ismertetnek, melyben az elektronikus kapcsolattartás és az e-receptek teljes körű bevezetés még várat magára. A szerzők megállapításokat tesznek arra vonatkozóan, hogy a területen legfejlettebb ország Észtország, Lettország, Magyarország és Szlovénia. A döntéshozók számára kiemelik az egészségügyi stratégia fontosságát, az e-health lehetőségek támogatását, a jogi és adatvédelmi környezet megteremtését és az edukációt minden résztvevő számára (Ćwiklicki et al. 2020).

Adatok és az egészségügy kapcsolata – A Big Data kora

A digitális egészségügy világában erőteljesen jelen van a tudatos egészségmenedzsmentet folytató betegek igényeinek előretörése, amely erőteljes Big Data szegmens is egyben, a világ adatmennyiségének megközelítőleg 30 százalékát adja a digitális egészségügy piaca, miközben egy beteg átlagosan 80 MB adatot generál az egészségügyi leletei révén, ugyanakkor az adatok sokszínűsége jelenti az egyik legnagyobb kihívást, amely mind az ellátórendszer, mind pedig a mesterséges intelligencia és az algoritmusok fejlesztése (radiológiai elemző szoftverek) során egyaránt felmerülő probléma. Az adatok összessége ugyanis tartalmazza a betegek korábbi eredményeit, a folyamatban lévő ellátási anyagait, a biztosítási információkat. A legnagyobb adatmennyiséget a képek jelentik, valamint a genomikai eredmények közül a komplett genomszekvenált adatsorok. Ezek mellé csatlakoznak azok az alkalmazások, amelyek a mindennapi egészségmérés részei, valamint sportolás során is képesek adatot szolgáltatni. 325 alkalmazás volt még csak elérhető 2017-ben, ez pedig minden évben 25 százalékkal nő, miközben az egészségügy a legalacsonyabban digitalizált piaci terület az elektronikus (például médiapiac, biztosítási piac) alkalmazási mezőben. Az adatok sokszínűsége mellett folyamatos kihívást jelent ennek fejlesztése és a növekvő piaci termékek elérhetővé tétele, a papír alapú betegadatok első szintjétől egészen az adatalapú és AI vezérelt döntéstámogatási rendszerrel rendelkező értékalapú egészségügyig (Gopal et al. 2018).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Az online tér, a közösségi média, az információszerző- és megosztó felhasználók a világháló egészén hatalmas mértékű adatot hoznak létre. Nincs ez másképp az egészségügy vonatkozásában sem, különösen a digitális egészségügy esetében. Ez az adatmennyiség nemcsak a kommunikáció, hanem a teljes szegmensre értendő, ennek léptékeit szükséges láttatni.

A keletkező egészségügyi adatok mennyisége kapcsán Banks rámutat, hogy míg 2013-ban 153 exabyte egészségügyi adat keletkezett világszerte, addig a várakozások szerint 2314 exabyte-ra nő ez az érték 2020-ban, addig az egyik kínai vállalat csak 2019-ben több mint 1 millió képen tanított a mesterséges intelligencia alapú algoritmust diagnosztikus céllal (Banks 2020).

Jelenleg az Egyesült Királyság és az USA jár élen a Big Data adatok felhasználásában az egészségügy területén, ennek példája a Genome Cancer Atlas (TCGA) vagy pedig a UK Biobank, valamint az USA „All of Us” projektje, de érdemes megemlíteni a svédországi rákregisztert, amelyre évente 70 millió dollárt meghaladó költségeket szán a kormányzat kutatási célokból.

Agrawal munkatársaival arra is rámutat egyúttal Topol nyomán, hogy az orvosok nemsokára terabyte méretű adatmennyiségekkel fognak találkozni, amelyet maguk a betegek állítanak elő saját állapotuk kapcsán, és ehhez az EHR rendszernek is alkalmazkodni kell, valamint az adatokat prediktív céllal feldolgozni, miközben a különböző rendszerekből érkező adatokat titkosítani kell, és továbbra is nagy kérdés az adattulajdonlás témája is. A szerzők felhívják a figyelmet, hogy arra a változásra, amely a Big Data alkalmazásával lehetséges, olyan adatbázisok használatával, amelyek a genomikai adatok mellett demográfiai jellemzőkkel is részt vesznek egy daganatterápia felállításában, még várni kell, de a szabályozási környezet kialakításakor figyelembe kell venni a lehetőséget (Agrawal – Prabakaran 2020).

A kutatás során pedig egyetértés mutatkozott Dessewffy Tibor és Láng László megfogalmazásaival a Big Data és a szociológia kérdéseinek megválaszolása kapcsán. Két irányt ismertetnek a szerzők: a gyenge és erős programokat, utóbbi esetében az adatelemzés révén lehetségessé váló komplex gondolkozást, új kérdések feltételét és módszertani önállóságot a hálózati struktúra felderítésére (Dessewffy – Láng 2015).

„A Big Data megteremti a dinamikus, módszertani individualista és hálózati szemléleten alapuló elmélet gyakorlásának tárgyi, empirikus feltételeit. A Big Data erős programja tehát a társadalomtudományok módszertani, elméleti és gyakorlati paradigmaváltását vetíti előre.” (Dessewffy – Láng 2015: 168)

A digitális egészségügy piaci környezetének tárgyalásakor különbséget kell tenni, hogy mit takar a cégek tevékenysége és a pénzügyi törekvései milyen jellegűek és milyen cél szolgálnak. Ezt célszerű megkülönböztetni annak alapján is, hogy a betegségek menedzsmentje vagy a prevenció áll a középpontban.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Az Egyesült Államokban a lakosságnak csaknem felének (43 százalék) van krónikus betegsége, beleértve a gyakori betegségeket (cukorbetegség, stroke, onkológiai betegségek stb.), amelyek az egészségügyi kiadások tekintetében elérhetik a 75 százalékot. A Milken Intézet eredményei (Raghupathi – Raghupati 2018) szerint az egészségügyi prevenció, életmódbéli szokások javítása évente akár 40 millió esettel kevesebbet adna az ellátórendszernek.

Ezek a betegségek sok esetben az élethosszig való fennállásuk miatt szoros terápiás együttműködést, valamint állandó kommunikációt igényelnek (például cukorbetegség), és erre a technológia számtalan formában lehetőséget nyújt ma már, akár okostelefon alkalmazások keretében, ezek tehát a digitális egészségügy piacának fontos szegmensei.

Cohen és munkatársai 1214 céget azonosítottak 2011–2018 közötti vizsgálatukban, olyan 1970 és 2018 között alapított amerikai cégek esetében, amelyek a digitális egészségügyben működnek. Ezek 73,2 százaléka tartotta elsődleges fókuszában a betegségmenedzsmentet és csak 23,8 százaléka a prevenciót, míg a betegségek időben történő detektálást 13 százalék esetében lehetett megfigyelni. Azok a cégek, amelyek csak a betegségmenedzsmentet helyezték előtérbe, a legnagyobb arányban befektetési rátával rendelkeztek, ez 12,8 milliárd dollárt tett ki, összehasonlítva a 4,7 milliárd dollár összegű önálló prevenció fókuszú befektetői költségekkel (Cohen et al. 2018).

Ennek okai többrétűek, de megfigyelhető, hogy a költségvisszatérítések során a preventív jellegű egészségügyi szolgáltatások nem élveznek pénzügyi előnyt, annak ellenére sem, hogy például a diabétesz terápiájánál lényegesen olcsóbb lenne a megelőző, életmódváltás alapú hozzáállás. Miközben az egészségügyi kampányok, köztük a WHO által képviselt álláspont is azt tűzi ki célul, hogy csökkenjenek a költségek és a betegek részvétele a hagyományos ellátás keretein belül, ehhez szükséges lenne a prevenció. A prevenció üzletág a viselhető eszközök és bioszenzorok területén dominálja, ezek közül az FDA két éve engedélyezte az egyik glükóz figyelő eszközt a diabétesz nyomon követéséhez.

Ezen eredmények (a limitált számú cégek ellenére) fontossága abban áll, hogy egyrészt kiténik, hogy az USA mint legnagyobb piac, milyen dominanciával rendelkezik a betegség-menedzsment irányában a prevencióval szemben még mindig, továbbá rámutatnak, hogy a betegek számára előnyös hozzáférhetőség, az adatok rendszerezése és elemzése révén történő ellátás, valamint folyamatos betegségmunka és utánkövetés egyaránt segítheti az ellátás minőségének javulását. A piac bevonása az ellátásba pedig reflektál arra az igényre is, amelyet a betegek támasztanak, mint legnagyobb vásárlóerői a kínált szolgáltatásoknak.

A bemutatott eredményekből látható, hogy a világ egyik legnagyobb egészségügyi piacán még mindig a betegségek és azok kezelése adja a piac pénzügyi erőterét,



TEMATIKUS TANULMÁNYOK – Digitális szociológia

miközben a digitális egészségüggyel foglalkozó cégek működésük révén elvezethetik az ellátórendszert egy jobb működés, illetve a hatékonyság irányába.

E-páciensek – Hogyan jellemezhető a tudatos, informált beteg?

Amint az említésre került, egyre növekvő mértékű a különböző egészségügyi paramétereket mérő eszközök, alkalmazások száma, és ezekkel egyre többen kerülnek kapcsolatba. Ezen felhasználók egy része egészségtudatos és preventív célokkal is rendelkezik, míg mások betegségüket szeretnék hatékonyabban menedzselni, és adatokkal rendelkezni akár a napi vércukor értékükről, akár kardinális állapotukról vagy sporttevékenységükről orvosuk felé.

Mindez elvezet oda, hogy a technológia lehetővé teszi azt, hogy a felhasználók, érintettek, betegek már nemcsak passzív szereplőként, hanem adatszolgáltató, cselekvő aktorként vegyenek részt saját egészségi állapotuk alakításában, és az ezzel összefüggő orvosi tevékenységnek proaktív részesei legyenek.

Ezt a helyzetet részben a piaci szolgáltatások elérhetősége, másrészt az online tér fejlődése és a szabadon elérhető orvosi tartalmak egyaránt eredményezték. Ennek alapján alakulhatott ki az az információkereső és feldolgozó attitűd a betegek részéről, amely az internetes tartalmak bővülése nyomán elvezetett az úgynevezett e-páciensek megjelenéséhez.

A definíció előtt azonban fontos leszögezni, hogy esetükben nem azon betegekről beszélhetünk, akik a különböző internetes keresőszolgáltatások segítségével orvosi tartalmakat keresnek, ezek között nem válogatnak, a forrásokat nem ellenőrzik, és sokszor hitelesítés nélküli, félrevezető egészségügyi hírek és információk között akár veszélyes egészségi állapotot is diagnosztizálnak saját maguknak. A tanulmány-nak nem tárgya az álhírek kérdésköre, sem az, hogy milyen alternatív gyógymódok, lehetőségek állnak a betegek rendelkezésére, illetve ezek körül milyen információs tér helyezkedik el, és milyen veszélyek vannak az egészségügyi hírek és ezek közösségi média világában.

A szerző mindvégig azon betegeket tartja szem előtt, akik ezzel a felhasználói viselkedéssel nem rendelkeznek, hanem állapotukat illetően az orvosi tudást tartják elsődlegesnek, amellet, hogy partneri viszonyt kívánnak elérni. A tanulmány kritikája lehet az, hogy ez a hazai ellátórendszerben, magyarországi viszonyok között mennyire valósulhat meg, az ismertetés azonban globálisan kívánja a jelenséget vizsgálni, krónikus betegséggel élők példáján keresztül.

Az e-páciensek esetében érdemes rámutatni, ahogy Meskó ismerteti, hogy a betegek öndiagnózist felállító, az orvosi tudást felülbíráló csoportja úgynevezett „Googler” és nem e-páciens. Véleménye szerint az orvos feladata az, hogy előbbi utóbbivá alakítsa, tehát olyan hozzáállást alakítson ki a betegek részéről is, melyben a beteg segíteni szeretne orvosának, kooperálni, továbbá betegségmenedzsmentje



TEMATIKUS TANULMÁNYOK – Digitális szociológia

során betegközösségek részeként a betegutak javításában is érdekelt. Ez azonban nem lehetséges anélkül, hogy az orvos maga is ne legyen kompetens a digitális térben, és ne tudja saját javaslatokkal, információszerzési csatornákkal is ellátni a betegeit (Kazai 2012).

Ezek a betegek tehát jelentős mértékben különböznek azoktól, akik egyszerűen információt keresnek az interneten keresztül. A fogalom megalkotója, Dr. Tom Fergusson megfogalmazása szerint maga a kifejezés (e-patient) az angol enable, engaged, empowered, equipped szavak előtagjából képezhető, tehát ezek a betegek elkötelezettek, érdekeltek, egyenlők – derül ki első nagyobb összefoglaló munkájából, ahogy az is, hogy nemcsak érdekeltek egészségüket illetően, hanem másodszakszvéleményt is kérnek állapotukról (Fergusson 2007).

Definíciója szerint: „*E-pácienseknek nevezhetjük azokat: (ide sorolandóak szintén azok is, akik saját panaszaikra keresnek online útmutatást, és azok is, akik barátaik, családtagjuk kapcsán lépnek online térbe e célból) akiknek kettős céljuk van az online egészségügyi tartalmak keresése során: egyrészt jobb és értékesebb információt és el látást szeretnének, másrészt a korábbtól különböző (ez nem jelenti azt, hogy mindig jobb) kapcsolatot a kezelőorvosaikkal.*” (a szerző saját fordítása). (Fergusson 2004: 1148–1149)

Chin már a 2000-es évek elején ismerteti, hogy a hagyományos paternalisztikus modellt felváltja a közös döntéshozatal modellje, melyben az orvosoknak érdemes előnyben részesíteni a partneri viszonyt, és egy olyan betegségmentes tervet létrehozni, amely egyesíti a betegek elképzeléseit, a kezeléssel kapcsolatos elvárásait és megvalósítási lehetőségeit, melyben az orvos már nem lehet a beteg által isteni szereppel felruházott beavatkozó, sem képzett technikai alkalmazott, hanem a beteg számára egyfajta „bölcs barát” pozícióját kell felvenni és ezzel irányítani a beteget a gyógyulás felé (Chin 2002).

Meskó felhívja a figyelmet arra is, hogy a digitális egészségügy által paradigma-váltás történik, amely egyben egyfajta kulturális transzformáció is, amely létrehozza a partneri viszonyt és megteremti a közös döntéshozatali modellt, az ellátás demokratizálódása érdekében. Fontos a kulturális dimenziót is kiemelni, hiszen folyamatosan növekszik az e-páciensek száma, akik másodszakszvéleményt kérnek, tudatosan és elkötelezetten döntenek, adatokkal rendelkeznek, amelyeket megosztanak orvosokkal, miközben az okostelefon-használat egyre növekszik, és a tudásanyag hozzáférhetősége révén az elefántcsonttornyot felváltja a közösségi média (Meskó 2017).

Az e-páciensek tehát a Részvételen Alapuló Orvoslás⁷ szereplői, valamint egy olyan konstruktív, bizalmon és együttműködésen nyugvó egészségügyi rendszer működésében érdekeltek, amelyben saját betegségük megállapítása és a kezelési terv felállítása együttes döntés eredménye, amelyben a saját tájékozottságuk, infor-

⁷ Participatory Medicine – <https://participatorymedicine.org>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

máltságuk segítheti a legmegfelelőbb gyógymód kiválasztását, amelyben az e-páciensek segítik orvosukat.

Dave DeBronkart a világ egyik leghíresebb e-páciense és motivációs előadója rákbetegségének történetét és az e-páciens lét gyakorlatát írta meg *Let Patients Help* című könyvében, melyben a fent nevezett együttműködést és kooperációt szorgalmazza gyakorlati tanácsok keretében a gyógyítás javítása érdekében.⁸

Jelen esetben releváns szerepük abban áll, hogy a kutatás alapjául választott betegségek szereplői élethosszig tartó kezelést igénylő betegek, akik betegségük menedzselésére szorulnak, azt napi szinten gyakorolják, a rendelkezésre álló információ pedig állapotuk fenntartását segíti, támogatja. A betegközösségek, akár online formában, akár közösségi média kommunikáció és egy hashtag alapú közösségszerveződés révén további megtartó erővel rendelkeznek szociológia értelemben is, valamint az érintettek számára fogódzót nyújtanak mentális téren is a betegség elviseléséhez és a teljes értékű élet létrehozásához. A krónikus betegség tehát már eleve ad egy e-pácines attitűdöt, a tanulmány és a szerző (Simon 2021) álláspontja azonban az, hogy a ritka betegségben érintettek modellezhetik ezen viselkedési módot a hosszú diagnosztikus út, az állandó információigény, a speciálisták és terápiás lehetőségek megkeresése okán.⁹ Ezért jelen tanulmányba két olyan, e-páciens tevékenységet folytató csoport online kommunikációja került, amely alkalmas lehet a gyakorlati bemutatásra is.

Az e-páciensekről nemcsak nemzetközi viszonylatban lehet beszélni, hazánkban is jelen vannak. Meskó és Györffy kutatásukban (Meskó – Györffy 2012) rámutattak arra, hogy már 2012-ben az általuk megkérdezett orvosok közel 100 százaléka használta az internetet szakmai vonatkozásban, de a betegek ilyen irányú felhasználása is meghaladta a 80 százalékot. A kutatásban résztvevő 83 orvos mindegyike kiemelte, hogy az internetes tájékozódás, kapcsolattartás nem helyettesíti a személyes orvosi vizitét, ugyanakkor arra is kitértek, hogy a betegek felkészültsége erősítheti az együttműködést, ahogy többen említést tettek arról is, hogy megadják az e-mail elérhetőségeiket, online tartalmakkal segítik pácienseiket hazai környezetben is.

⁸ <https://www.epatientdave.com/let-patients-help/>

⁹ A ritka betegségekre nincs egységes, egyetemes definíció az orvosi szakirodalomban sem, ahogy az egyes országok is eltérően határozzák meg, hogy mely betegségcsoportok tartoznak ide, ugyanakkor léteznek adatbázisok, és ehhez tartozóan nemzetközileg elfogadott BNO kód alapján történő csoportosítás is, amely segíthet a meghatározásban. A definíciós keretek kapcsán elmondható, hogy a ritka betegség prevalenciája Európában 1:2000-hez, valamint 50-75 százaléka gyermekeket érint. Gyakran súlyos, krónikus lefolyású betegségek tartoznak ide, a betegek 30 százaléka meghal 3 éves kor előtt, továbbá a betegségek 80 százaléka genetikai eredetű. Jelenleg 6-8000 között van a ritka betegségek száma, az érintettek pedig Európában elérhetik a 36 milliót, a családtagokkal együtt pedig csaknem 100 millió életet befolyásolnak. (Ritka és Veleszületett Rendellenességekkel élők Országos Szövetsége, www.rirosz.hu)



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Ahogy egyik interjúalanyuk kiemeli:

„Úgy gondolom, hogy a páciensek többsége amúgy is használja az internetet, és ha akarnánk, se tudnánk leszoktatni őket. Ezért úgy kell rájuk hatnunk, hogy az ottani információkat a lehető legjobban tudják felhasználni.” (Meskó – Györffy 2012 : 683)

Ez egyezik az e-páciensek hiteles információra való törekvésével, azzal, hogy tudatos betegként vegyenek részt a folyamatban, és orvosaiknak segítőivé váljanak.

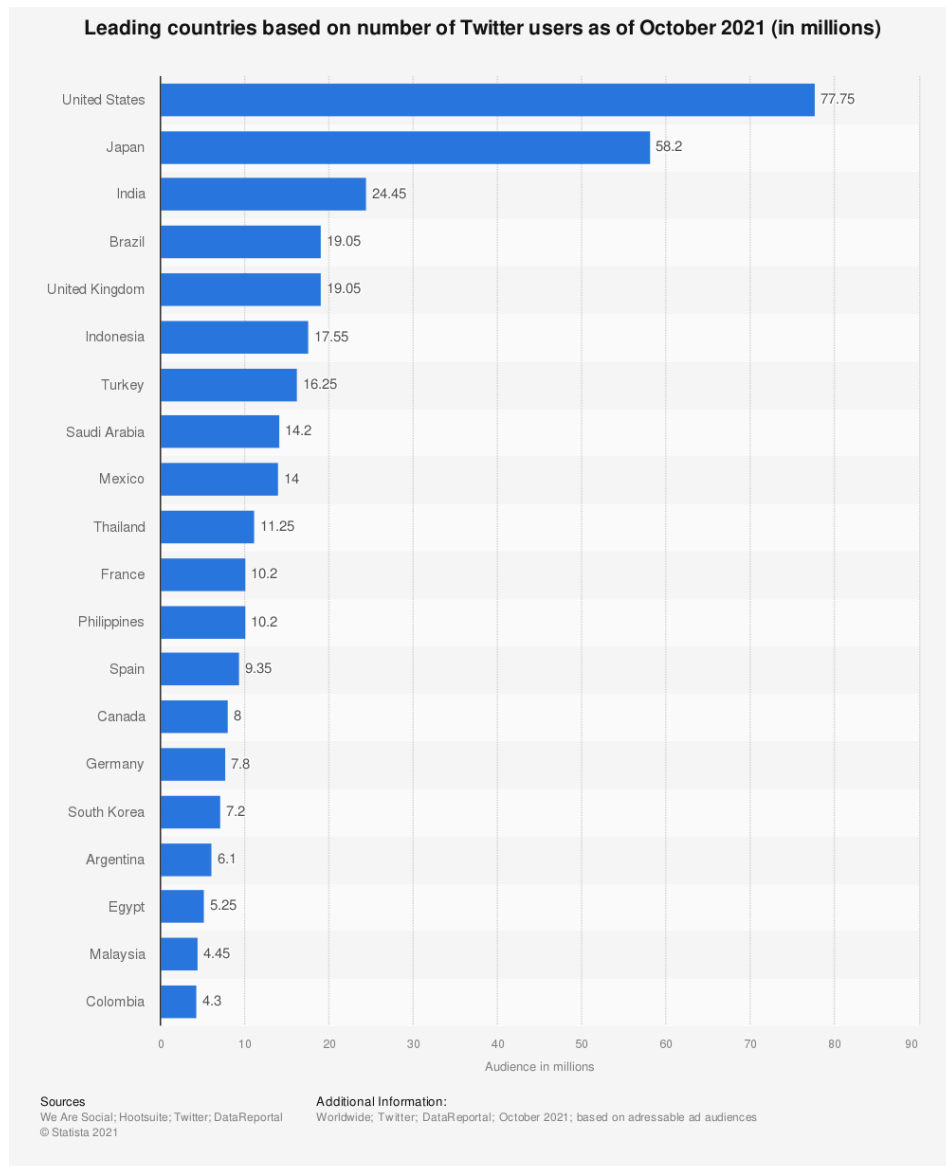
Közösségi média és a gyógyítás – Az információszerzés útjai

Amennyiben e-páciensekről és digitális egészségügyről beszélünk, szükséges az online térben a közösségi média adta lehetőségeket is tárgyalni az orvosi témájú tartalmak esetében.

A közösségi média használatról átfogó jelleggel elmondható, hogy jelenleg a legnagyobb népszerűséggel rendelkező internetes tevékenység, 2020-ban 3,6 milliárd felhasználót is meghaladta ez az érték, várakozások szerint pedig ez 2025-re eléri 4,41 milliárd felhasználót. A Statista közlése szerint átlagosan az internethasználók naponta 144 percet töltenek közösségi oldalon és alkalmazásokban, ez 2015-höz képest már félórával nagyobb időráfordítás. Jelenleg a legnagyobb bázissal rendelkező közösségi média oldal továbbra is a Facebook, 2,7 milliárd aktív, havi felhasználóval (Statista 2).

A Twitter 2006-os megalapítása óta elsősorban az amerikai (77,75 millió felhasználó 2021-ben) és japán piacon (58,2 millió felhasználó szintén 2021-ben) dominál, illetve kiemelt mértékben a közéleti, politikai kommunikációban van jelen (2. ábra), de markánsan domináns közösségi média oldalak között aktív felhasználói száma okán is. Ma már a legtöbb véleményvezér szereplőnek van aktív Twitter fiókja, sok esetben külön szakmai profillal. A felhasználók létszáma folyamatosan növekszik, a 2018-as 335 millióról 2021-ig 421 millióra, és 2020-ban a piaci szereplők több mint fele (53 százalék) a Twittert üzleti célra is használta, ráadásul az internetes cégek közül a kilencedik pozíciót el tudta érni a szolgáltató az Egyesült Államokban. Csak a Facebook, az Instagram és a LinkedIn tudta megelőzni, annak ellenére, hogy a marketing célú használatban a Facebook és a képmegosztó megelőzi, de a növekvő felhasználói létszám mutatja, hogy a felület továbbra is értékes maradt az információ áramlásban, nem beszélve a közéleti relevanciájáról (Statista 4).

TEMATIKUS TANULMÁNYOK – Digitális szociológia



2. ábra. Twitter használó országok nemzetközi megoszlása 2021-ben

Forrás: Statista 3



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Logikus lenne tehát az, hogy felmerüljön, miért nem a Facebook lett a kutatás és az elemzés tárgya, ilyen arányok mellett. Ugyanakkor figyelembe kell venni azt a tényt, hogy a Facebook adatvédelmi rendelkezései az API-n¹⁰ keresztül történő adatgyűjtés esetében is minden esetben az érintett felhasználó engedélyéhez kötik az adatok megszerzését és mentését, védve ezzel az egymással szükségszerűen kapcsolatban álló és saját adatvédelmi beállításokkal rendelkező felhasználók érdekeit.

Ez a Twitteren nem áll fenn, mely egész struktúráját illetően nem a személyes ismeretségre, hanem a követésre épít, akár egymás számára ismeretlen emberek esetében is, továbbá mindenki számára nyilvános.

A Twitter használat esetében már a 2014-es Ebola járvány kapcsán is látható volt, hogy mennyire jelen van az egészségügyben. A Symplur¹¹ közlése szerint a hashtag az akkori időszak egyik legnépszerűbb orvosi címkéje volt, továbbá a felület hatását az is mutatja, hogy lehetőség van arra, hogy konferenciákat élőben közvetítsenek a résztvevők és szakértők által moderált élő chatre, amely a tudományos közösség számára lehetőséget teremt a diszkusszióra, valamint az ágazatban részt vevők rövid üzenetekben nagyon széles közönség számára tudnak információt megosztani, akár még a tudományos publikáció előtt is.

A Twitter ugyanis a követési rendszer alapján azt teszi lehetővé, hogy főleg szakmai használat esetén, nagyon távoli felhasználók is kapcsolatba lépjenek, így az információ szóródása kellően széles lehet, ismeretség nélkül, ami egy speciális téma esetében különösen hasznos lehet, valamint a hashtag önmagában, a kezdetektől fogva lehetőséget ad arra, hogy összegyűjtse témacsoportonként a tartalmakat. Veszélye azonban az orvosi témában pontosan az, hogy a rövid üzenetek hitelessége nem minden esetben biztosított, közvetlen orvosi tanácsra természetesen nem alkalmas, ugyanakkor jelen van az a tény, hogy ellenőrizhető profilú szakértőktől, egészségügyi szervezetektől szabadon kaphatunk információt, közvetlenül, személyes kapcsolat nélkül is (Goff – van den Bergh 2015).

A Twitter és a digitális egészségügy – A mikroblog szerepe

Az egészségügy és az információszerzés kapcsolatában megkerülhetetlen szerepe van a közösségi médiának.

Pershad munkatársaival a közösségi felület szerepét a televízióval veti össze, abban a vonatkozásban is, hogy mindkettő célja az orvosi információ elérhetővé tétele, ugyanakkor utalnak a CNN csatornán korábban futó Dr. Oz showra, amely a szerzők szerint „mágikus tablettaként” tudott hatni a közönségére, bemutatva a tudásanyag mellett például alternatív gyógyászati szereket, jelentős követőbázist elérve az orvos

¹⁰ Application Programming Interface

¹¹ www.symplur.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

műsorvezető karizmájával. Ugyanakkor, amint az Korownyk tanulmánya megmutatja, 400 esetből mindössze 46 százalék volt orvosi tényekkel alátámasztva, tehát jelentős mértékben eltolódott a szórakoztatás és a showműsor elemei felé a tartalom-szolgáltatás. A Twitter egészségügyi ágazat által történő használata viszont nemcsak a direkt kommunikációt teszi lehetővé és elérhetővé az orvosok és a betegek között, hanem lehetőséget ad arra, hogy közösségi és egyben szakmai profilú kapcsolati hálóként is jelen tudjon lenni. Minden érdekcsoportnak, betegeknek, orvosoknak, ágazati szereplőnek egyaránt elérhető, nem limitál szakmai szervezetekbe kizárólagos információelérési lehetőségeket. Fontos azonban kiemelni az álhírek szerepét, amely ellen a szerzők szerint a crowdsourcing erejével próbálnak fellépni, azonban figyelembe kell venni, hogy egy-egy közszereplő megnyilvánulása milyen erővel bírhat. Az a tény azonban, hogy több mint 2000 egészségügyi szolgáltató, legalább 3000 követővel jelen van és naponta tweetelnek, már elegendő alap, hogy a Twitter megkerülhetetlen platformja legyen a digitális egészségügyi térnek (Pershad et al. 2018).

Cisztás fibrózis és asztma – Két betegség, közös metszéspontokkal

A tanulmány ezen pontján, a digitális egészségügy és a közösségi média után érdemes áttekinteni a kutatáshoz választott betegségeket és azok ismérveit, valamint ezeket összevetni az e-páciensek helyzetével. A kutatás két krónikus betegséggel foglalkozó Twitter hashtaget (címkét) elemezve szeretne áttekintést adni a mikroblogon zajló kommunikációról, ugyanis a felhasználói kör ez esetben nemcsak a betegeket jelenti, hanem a betegszervezeteket, az ellátórendszer képviselőit is, ezzel pedig lehetőséget teremt a kooperatív kapcsolatra és a betegedukációra is. A két kiválasztott betegség pedig ez esetben a sajnos világszerte jelenlévő betegség, az asztma és a ritka betegségek közül pedig a szintén tüdő érintettségű cisztás fibrózis.

Az asztma légzőszervi megbetegedés, a tüdőbe vezető főbb légutak, köztük gyakran a hörgők krónikus gyulladása (Zsuga 2019). Az asztma epidemiológiai háttéréről elmondható, hogy világszerte csaknem 300 millió érintettje van, az európai prevalencia 5 és 10 százalék közötti, hazai vonatkozásban 2016-os adatok szerint 300 ezer beteget tartottak nyilván, ugyanakkor ez a szám a fekvőbeteg státuszú és az háziorvosi praxisban előforduló betegeket nem tartalmazza, így ez a szám csak közelítőleg tudja megadni az asztmás betegek számát. Miközben előfordulását tekintve gyakori betegség, mortalitását illetően Európában körülbelül 100 ezer betegből 5 veszíti életet évente (5/100000 asztmás/év körüli) (EMMI irányelv).

Az életmódot befolyásoló tünetegyüttese és kezelési igénye miatt kiemelt figyelmet érdemel mind a betegek, mind pedig az orvosok részéről, ráadásul a kezeletlen allergiában szenvedő betegek akár 20-40 százaléka is érintett lehet a későbbiekben asztmás panaszokban. Az asztmás betegek nagyrésze elsősorban enyhébb tünetek-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ben érintett, amelyet fokozhat fizikai megterhelés is, ugyanakkor a súlyos panaszok fellépése is gyakoribb körökben. Alapos pulmonológiai kivizsgálás előzi meg a diagnózist, a terápia célja pedig a gyulladás visszaszorítása, a tünetek megelőzése, mérséklése (Máriáss 2018).

A WHO közlése szerint 339 milliót is meghaladta a betegséggel élők száma a világon már 2016-ban, a legtöbb esetben halállal idősebb korban végződik, különösen a fejletlen országokban, de általánosan igaz az, sok esetben alul diagnosztizált, és nem jut elegendő terápiás lehetőség a betegeknek (WHO Newsroom 2020).

Szociológiai relevanciáját az a tény adja, hogy egyrészt sok embert érintő, élet-hosszig fennálló betegségről van szó, másrészt nagy számban a munkaképes korosztályt érinti, továbbá a betegség alakulása befolyásolható többirányú megközelítési móddal. Ez azonban nem jelenti azt, hogy a krónikus betegségből eredeztethető állapotromlás, keresőképtelenség ne lenne jelen ebben a betegcsoportban, valamint a kezelés során fellépő költségek egyaránt a betegre és az ellátórendszerre is nagy terhet rónak, különösen igaz ez a sürgősségi ellátásra, a nem megfelelően megválasztott gyógyszerek használatára. A preventív szemléletű ellátás mindkét érdekcsoport számára kedvezőbb eredményeket hozhat, mint az akut ellátás szükségessége (EMMI irányelv).

A cisztás fibrózis az európai populációban leggyakrabban előforduló ritka betegség, mely genetikai eredetű. Elsőként Anderson írta le 1938-ban, amikor a várható élettartam mindössze egy év volt, de a fejlett országokban megfelelő terápia, akár tüdőátültetés révén a 40 éves kort is meghaladhatja napjainkban. Incidenciája meglehetősen eltérő Európában is, 1:1353 Írországban, míg 25 ezer főből legfeljebb egyet érint Finnországban, átlagosan Nyugat- Európában 4500 betegből 1 él cisztás fibrózissal, míg az arány 1:6000-hoz Észak- és Közép-Európában. A világ többi táján a betegek száma kevesebb, de ennek adatgyűjtési és regiszterbeli okai is vannak. A várható élettartamot befolyásolja a tüdő állapotromlása, továbbá demográfiai adatok és infekciók egyaránt szerepet játszanak (Scotet et al. 2020).

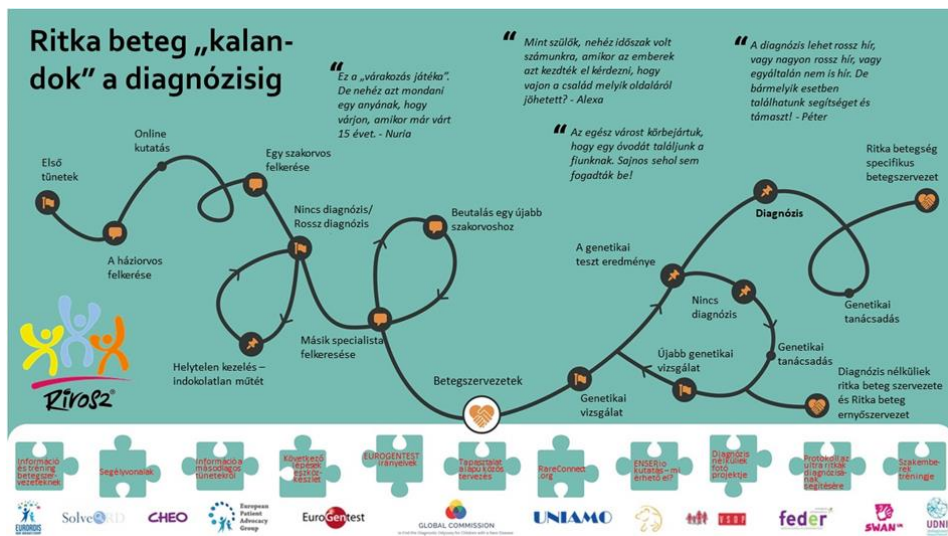
A cisztás fibrózis esetében igaz a ritka betegségben érintettekénél meglévő hosszú diagnosztikus út, ugyanakkor az az asztmával való metszéspontot az is adja, hogy a betegeknek komoly pulmonológiai vizsgálat sorozaton kell átmennie, amíg a panaszok eredete kideríthető, és mindkettő esetében szembe kell nézni azzal, hogy környezeti tényezők ronthatják, és speciális életmód követése segíthet a betegség uralásában, ehhez pedig komoly betegegyesület és információmennyiség szükséges.

A két betegség tehát mutat részterületeket, valamint orvosi szakterületeket, amelyek egyezők (pulmológia, infektológia, belgyógyászat, kiegészítő terápiák, gyógytorna stb.), és akár a kommunikációs mintázat vizsgálat esetében, akár a szemantikai elemzés során egyezések is lehetnek a tartalmakat illetően, például a szavak esetében (hospital, breath, infection stb.).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A ritka betegségek esetén fontos említést tenni arról is, hogy a betegek hosszú betegút és sok esetben kudarcot jelentő, állapot rosszabbodást is eredményező diagnosztikus és kezelési kihívásokkal néznek szembe. Ezzel egyrészt az európai ernyőszervezet, a EURORDIS, másrészt a hazai betegszervezet (RIROSZ) ábrája is foglalkozik. A 3. ábrán látható, mely az európai tapasztalatokat foglalja össze, hogy mind a szülők, mind pedig a betegek számára mennyire megterhelő a bizonytalanság, a sokszor feleslegesen elvégezett beavatkozások megélése, a szociális ellátórendszer hiányossága (speciális óvoda egy-egy kisebb településen), de akár az, hogy másfél évtized is eltelhet addig, amíg a betegségre fény derül.



3. ábra. A ritka betegségben érintettek betegútja és a társzervezetek

Forrás: Ritka és Veszélyes Rendellenességekkel Élők Országos Szövetsége, www.rirosz.hu

A tanulmány és a kutatás éppen ezért arra fókuszál, hogy ebben a környezetben, ahol a bizonytalanság és a szakmai kompetencia hiánya jelen van, milyen jelentőséggel bír az információ, különösen az online tartalmak, melyek bárhol elérhetőek. Az információ és a hiteles orvosi tartalmak lehetnek azok, amelyek a szakadékokat csökkentik, a betegek később e-páciensekké válhatnak, az élethosszig tartó jelleg miatt önmenedzsmentre lesz szükségük, továbbá az információ révén tudnak betegközösségeket találni, segítségre lelteni, ez pedig a ritka betegségek esetén kiemelt fontosságú.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Módszertan – Big Data alapú kutatás

A Big Data alapú, feltáró kutatáshoz social media listening platform¹² került felhasználásra, mely kifejezetten egészségügyi területre fókuszáló, valamint algoritmusai révén többirányú megközelítésre ad módot. A közösségi média figyelése marketing célú felhasználás, márkafelügyelés során széles körben elterjedt, ugyanakkor alkalmas lehet a trendfigyelés és az említések révén kutatási célokra is.

A kutatás mindenkor korlátja természetesen az adatok hozzáférhetősége, annak korlátokkal rendelkező mivolta, valamint az a jelleg, hogy elsősorban deskriptív jellegű közlést tesz lehetővé, oksági válaszokra ezen formában nem alkalmas. Azt azonban meg tudja mutatni, hogy a felhasználói kör milyen mintázatot mutat be kommunikációja révén egy gyakori és ritka betegség esetén, amelynek van tüneti metszéspontja is. A felhasznált adatbázisok mindegyike a szolgáltató cég által hosszasan monitorozott és hozzáférhető volt.

Ahhoz azonban, hogy széles körben bemutatható legyen az említett több részből álló terület, módszertani és kutatási döntéshozatal keretében meg kellett határozni, hogy mely adatbázisok lesznek alkalmasak erre a célra, úgy, hogy a korábban már tárgyalt adathozzáférhetőségi elvet teljesíteni tudják.

A Big Data vizsgálatok előnyei mellett számolni kell azonban fennálló hátrányakkal is, és érdemes említeni azokat a kutatás során felmerülő kérdéseket, amelyek például az adatbázisok kiválasztását kísérik, ez esetben azt a folyamatot mutatják be, hogy mely szempontok szerint lett a két adatbázis kiválasztva.

Az elméleti keret már bemutatta, hogy a digitális egészségügy fogalma alatt többféle megközelítést lehet érteni, valamint több részterületet foglal magába. Ebből következően, ha a digitális egészségügyet szeretnénk a Twitteren vizsgálni, akkor kérdés lehet az is, hogy melyik hashtaget célszerű kiválasztani, úgy, hogy a választott social média listening platform (Symplur)¹³ enged keresőszót vagy konkrét felhasználót is vizsgálni. Változatlan döntési alapvetés volt, hogy ne keresőszó legyen az adatbázisok alapja, ugyanis az komoly kockázatot rejt abban a tekintetben, hogy más, akár oda nem illő téma is az adatbázis része lehet tartalmilag. A hashtagek esetében is ez fennálló probléma a user egyéni hashtag preferenciája okán, azonban

¹² A tanulmány során végig eredeti formában szerepelnek angol kifejezések például a hashtag, a social media listening, a stakeholder szavak, valamint az adatbázisban feltüntetett csoportok nevei. Ezek magyarosítása, tekintve, hogy a nyelvben is átvett a használatuk, nem lenne indokolt, ugyanis félreérthető lenne ezek önkényes elnevezése a szerző által egységes, elfogadott magyar referenciaszavak megléte nélkül.

¹³ A Symplur egy kaliforniai székhelyű közösségi média monitorozással foglalkozó cég, mely egészségügyi témában, a Twitteren gyűjt adatokat, ezeket pedig megvásárolható licenz formájában, kutatási elemzésekhez is közzéteszik. Az oldal létrehozói real-time Big Data analízist végeznek elsősorban piaci szereplők és döntéshozók, akár kormányok számára is. A szolgáltatás kialakítása a Stanford Egyetemmel együttműködésben történt. Lehetőség van kulcsszó, hashtag vagy usernév kiválasztására. Módszertani okokból a hashtag adta a mintavételi egységet, ebben az esetben mind a #Asthma, mind #cysticfibrosis elérhető volt, összehasonlításra módot adott.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

módszertanilag a tweetek és az annak alapján felállított tartalmak esetében jelen van, hogy mindegyik egységesen tartalmazta az adott címkét. Tehát nem kulcsszó volt az adott, kiválasztott fogalom egy másik hashtagnél. Mind az öt esetben a hashtag adta tehát az adatbázist.

A gyakori betegségek mellett, ritka betegségből csaknem 8000 féle ismert (Európai Bizottság), ezért első lépésként a Symplur adatbázisát kellett a hozzáférés után feltárni, hogy mely betegségeket gyűjti hashtagek formájában, ez azonban, mivel keresőszavak révén történik a betegségek listázása, nem segíti a módszertani limitet, hiszen nem jelöli a social média listening relevancia szerint, hogy az adott betegség ritka kategóriába tartozik-e, ezért fordított megközelítés alkalmazása bizonyult célravezetőnek, azaz egy gyakori betegség kiválasztása, amely a populáció nagy százalékát érinti (#Asthma), és ahhoz kellett illeszteni a ritka betegséget és megkeresni, hogy elérhető-e.

Nagy betegpopulációt érintő betegségek között van például az asztma, a cukorbetegség és számtalan onkológia betegség vagy akár a stroke, azonban a fentiekén túl a kutatás során az is vezérelt, hogy olyan adatbázis kerüljön kiválasztásra, amely a nemzetközi irodalomban csak kevésbé elemzett, különösen olyan módon, hogy kontrollcsoportként is szükség volt rá, tehát a kiválasztás célja a ritka betegséghez való társíthatóság is volt. A nagyobb adatbázisok nagyrésze már tárgyalt, ezért a szakirodalmi és publikációs feltárást a Symplur research segítette, ahol a már megjelent Symplur alapú kutatásokat tüntetik fel. Az elemzéshez adatokat a Symplur Signals szolgáltatotta, ezek kiválasztása, összegzése, feldolgozása saját kutatási eredmény a közölt adatok alapján.¹⁴

Az adatbázisok elemzése során megfogalmazott kutatási kérdések arra irányultak, hogy a nagyszámú beteget érintő adatbázis hasonló jellegzetességeket mutat-e a ritka betegséghöz, milyen dinamikát mutatott a kommunikáció a vizsgált hét éves periódusban, melyek a hashtagekhez kapcsolódóan a leggyakrabban előforduló szavak és milyen témákra irányulnak? Vizsgált volt továbbá az is, hogy az influenzaszerek egyéni vagy szervezeti kommunikációt folytatnak-e, és milyen az egyes stakeholder csoportok megoszlása az adatbázisokban a tweetek nyelvi preferenciája mellett, exploratív jelleggel.

Kutatási eredmények – #Asthma és #cysticfibrosis

Az elemzés a Symplur social média listening adatbázisain került elvégzésre 2012. május 11. és 2019. május 11. közötti időtartam vonatkozásában, így lehetőség nyílt hét teljes év adatait összehasonlító jelleggel elemezni.

¹⁴ Az ábrák a Symplur Signals rendszeréből letöltött ábrák, általuk generálva, változtatás nélkül közzéve.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A várakozásoknak megfelelően elmondható, hogy a legtöbb tweetet a #Asthma követői küldték el, és ehhez tartozott a legtöbb felhasználó is (1 094 835 tweet és 384 783 felhasználó). A #cysticfibrosis 660 752 tuellel és 154 782 felhasználóval rendelkezett a vizsgálat időszakban.

A betegségek előfordulási gyakoriságának tükrében ugyanakkor látható, hogy a cisztás fibrózissal élők is jelentős számban képviseltetik magukat a mikroblogon, és a vizsgálati időszakot tekintve nem maradnak el online kommunikációjukban.

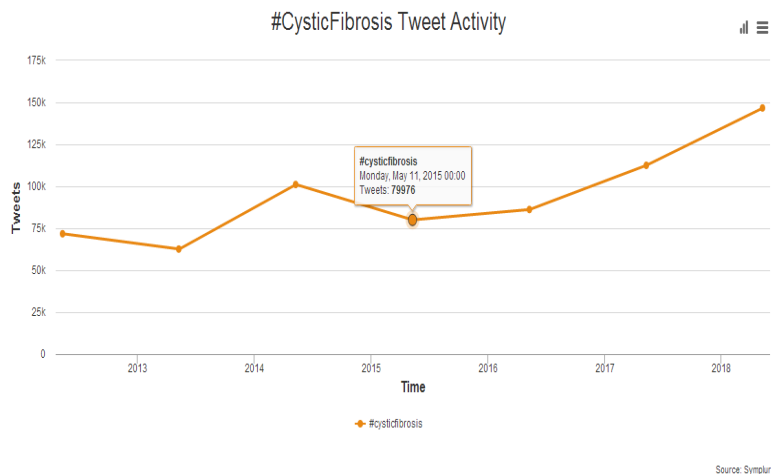
A #**cysticfibrosis** esetében az összes tweet 49 százaléka volt retweet, választ ugyanakkor csak 2,29 százalékban figyelhetünk meg, és a tweetek 50,5 százaléka esetében volt látható, hogy a tartalomban volt beillesztett link is.

A #**Asthma** külön érdekes abban a tekintetben, hogy a retweetek aránya 37 százalék, a linkek vonatkozásában hasonló a megoszlás 70,2 százalék, a válaszok száma pedig 20 365, ami 1,86 százalékot tesz ki.

Az összesített eredmények közül azért erre a három kategóriára fókuszál az elemzés, mert a retweetek aránya az információterjedés kapcsán lehet informatív, azaz, hogy a kommunikáció interaktivitása nem a userek közötti direkt válaszok formájában történik, a linkek pedig a többletinformáció szolgáltatása miatt lehetnek fontosak, azaz, hogy egy adott tweet a felhasználók számára mit közvetít.

A fenti adatok a hét éves periódus összesített adatai, melyben az egyes értékek esetében figyelembe kell venni, hogy egy tweet tartalmazhat egyszerre linket és választ is, illetve retweetelhetők is.

Tweet aktivitás

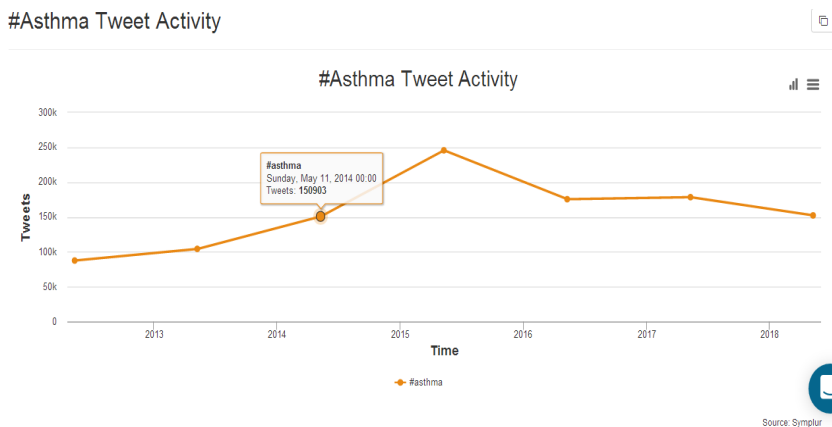


4. ábra. Tweet aktivitás a #cysticfibrosis esetében

Forrás: www.sympLur.com

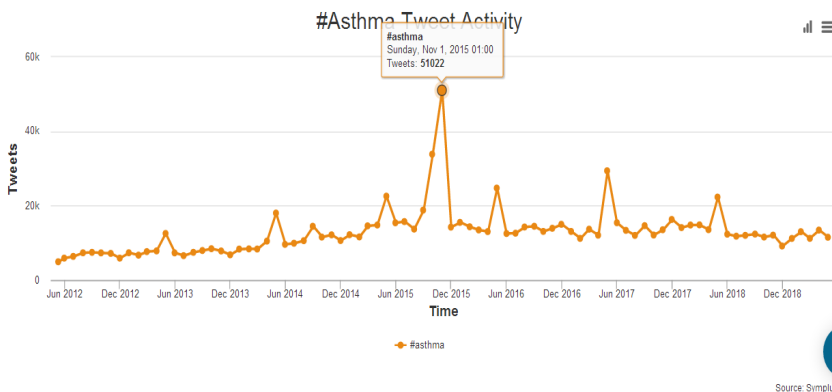
TEMATIKUS TANULMÁNYOK – Digitális szociológia

A cisztás fibrózis évenkénti tweetforgalmában (4. ábra) megfigyelhető, hogy 2012 után (71 727 tweet) visszaesés történik a következő évben (62 607 tweet), azonban a rákövetkező évben a 100 ezret is átlépi ez a szám (101 076), majd ismételen csökkenés mutatkozik 2015-ben (86 175 tweet), ettől kezdődően azonban folyamatos a tweetforgalom intenzitás növekedése, és a hetedik vizsgált évben már eléri a 146 691 tweetet. Tehát ebben az adatbázisban a koronavírust megelőző három évben folyamatosan emelkedett az elküldött tweetek száma.



5. ábra. Tweet aktivitás a #Asthma esetében éves bontásban
Forrás: www.symplur.com

A #Asthma 245 669 tveettel a 4. évben teljesített a legjobban (5. ábra), az addigi felfutó tendenciával, majd ezt követően csökkenni kezdett, és végül 152 294 tveettel zárta az utolsó vizsgálati évet, az azt megelőző 178 477-tel szemben.



6. ábra. Tweet aktivitás a #Asthma esetében havi bontásban
Forrás: www.symplur.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A #Asthma esetében a legnagyobb tweetforgalom (6.ábra) 2015 november 1-jén volt 51 022 tweettel.

Az adatok kapcsán érdemes megjegyezni, hogy a Big Data elemzések közvetlenül is segíthetik a gyógyítást. Angolszász nyelvterületen, méghozzá Ausztráliában, amely ugyan nem jár élen a Twitter-használatban, de az angolnyelvi közösségben igen, és 2016. november 21-én volt a világ legnagyobb epidemiológiai szempontból is katasztrófát eredményező eseménye Melbourne-ben, több mint 2000 akut ellátásra szoruló légzőszervi beteggel a sürgősségi osztályokon. A „Thunderstorm asthma” kialakulása mögött időjárási és környezeti okok egyaránt megtalálhatóak, többek között az eső és a levegőben megtalálható pollenkoncentráció. Thien és munkatársai részletesen elemezték mindazokat a faktorokat, amelyek ezt a fenti, példátlan egészségügyi vészhelyzetet okozták (Thien et al 2018). Silver és szerzőtársai (2018) tárgyalják azt is, hogy a zivatarokat követő koncentráció okozta panaszok és a fennálló asztma elkülönítése diagnosztikus probléma is, és Ausztráliát elsősorban a késő tavaszi, kora nyári időszakban kell e tekintetben kiemelten kezelni. A Big Data adatok pedig segíthetnek a predikcióban a kórházi adatok, a meteorológia előzmények és meglévő eredmények, valamint azok a demográfiai jellemzők is (nem, életkor), melyek jelentősek a betegségben. Ezek együttese adta modelljüket, amellyel előre tudták jelezni a bekövetkező magaskockázatú asztmás periódusokat, és ennek révén mind a lakosság, mind az ellátórendszer felkészítése megtörténhet, reagálásra ad módot. Az adatok mögötti események feltárásához szükséges tehát a betegség epidemiológiai hátterének ismerete is és ennek összekapcsolása az eredményekkel, amelyek segíthetnek magyarázatot adni bizonyos, látható tendenciákra (Silver et al 2018).

Felhasználói viselkedés – További kutatási eredmények

A vizsgált adatbázisok mindegyikére igaz az, hogy nagy számú azon felhasználók köre, akik mindösszesen 1 tweetet küldtek el 7 év alatt, és alacsony azoknak a száma, akik sokat. Ez továbbra is jelenthet passzív követést, de a közösség interakciója alacsony. 1 tweetet a cysticfibrosis felhasználók 75 százaléka küldött el, de ugyanez igaz 72,6 százalékkal a #Asthma esetében is.

A platformnak megfelelően a tweetek több mint $\frac{3}{4}$ -ét angol nyelven írták. A jelenség a digitális egészségügy trendjeivel és nemzetközi jellegével egybevág, a #cysticfibrosis esetében ez némiképp alacsonyabb, de a két adatbázis nem mutat különbséget (74,7 százalékkal).

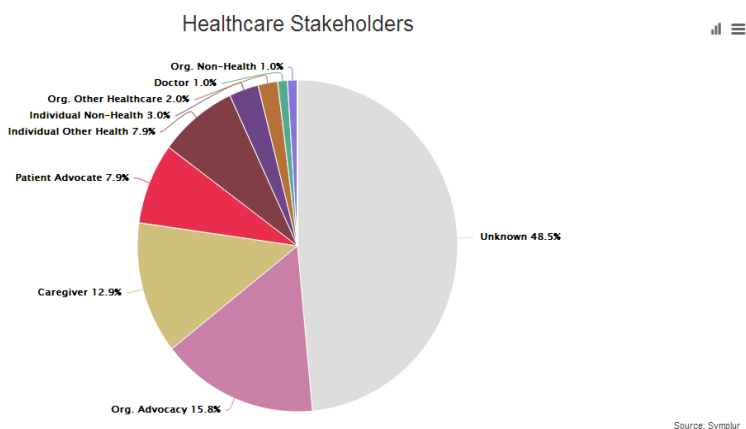


TEMATIKUS TANULMÁNYOK – Digitális szociológia

Influenszerek (véleményvezérek)

Az influencerek esetében módszertani limit volt, hogy az első 100 véleményvezér került a mintába, tekintve, hogy az összes adatbázis esetében csaknem 70 százalék a felhasználóknak 1 tweetet küldött el, tehát sok passzív felhasználó van jelen, ezért a keret ennél nem volt indokolt, hogy nagyobb legyen. A megoszlás tehát a TOP 100 influenszerre vonatkoztatandó.

#CysticFibrosis Influencers by Tweets



7. ábra. Influenszer aktivitás tweetek alapján a #CysticFibrosis esetében

Forrás: www.symplur.com

A cisztás fibrózis esetében a legnagyobb stakeholder csoportot az influenszerek közül, az adatbázisban szereplőknek csaknem a felét, az algoritmus szerint a nem kategorizálható felhasználók adják (Unknown stakeholder kategória a Symplur rangsorolása és kategória képzése szerint, 48,5 százalék (7. ábra). Ez esetben azon általános módszertani felvetések érvényesek itt is, mely szerint anomális viselkedést mutató felhasználókról beszélhetünk, akik között lehetnek robotok, de olyan valós személyek is, akiknek a tevékenysége és a profilja nem elegendő a kategorizáláshoz.

A Symplur külön jelöli a Spam és Unknown, tehát a „nem azonosítható” típusú felhasználói kategóriát, a kettő között átfedés is lehet, ugyanakkor az, hogy egy user unknown kategóriában van, nem jelenti azt, hogy egyéni kategorizáció esetén ne kerülhetne egy másik csoportba. 2020-ban a Symplur ezt már lehetővé tette az előfizetők számára, az algoritmus tanítása érdekében, hogy manuálisan kategóriát lehessen módosítani.

Az 2. legnagyobb csoport az Org. Advocate volt, a 3. a Caregiver 12,9 százalékkal. A betegek 10 százalék alatt voltak (7,9 százalék), az orvosok pedig alig 1 százalé-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ban voltak jelen a TOP 100-as mintában egy ritka betegség esetén, ahol az orvos-beteg találkozás és a hiteles orvosi információ alapvető.

Az eredményt illetően elmondható, hogy a betegedukáció megvalósulása a szervezeti kommunikációs aktorokon keresztül valósulhat meg, azonban az a tény, hogy majdnem 50 százalék ismeretlen felhasználó van a mintában, felveti annak kérdését, hogy a hashtaget használók hiteles információhoz jutása milyen mértékű, tehát tájékozódásra ez a platform kevésbé lehet alkalmas.

Row	Tweets	Thumbnail	Username	Name	Description	Healthcare Stakeholder	Followers
1	15,633		@expatinc	Expat Inc	Making personalized health information available to everyone everywhere for free.	Org. Other Healthcare	1,256
2	8,963		@jodavidson65	JoanneD #nomoreCFangels	hoping/praying for a cure for cystic fibrosis one day soon so our beautiful lass will hv long happy life she deserves. Tweet abt CF. Strictly & general stuff	Individual Non-Health	773
3	7,656		@emmagadie2009	Emma Gadie	My son Oily and my daughter Grace are my whole world. Couldn't imagine my life without them! #OrkambiNow	Unknown	411
4	7,246		@cysticbot	cysticbot	I retweet mentions of Cystic Fibrosis	Unknown	2,410
5	7,191		@strawfie	Strawfie Challenge	A Selfie Campaign in aid of the Cystic Fibrosis Trust (UK)	Org. Advocacy	2,555

8. ábra. Influenszer rangsor tweetek alapján #cysticfibrosis adatbázisnál
Forrás: www.symplur.com

Az első 3 influenszer (8. ábra) között azonban kettő a betegséghez szorosan kötődik, az első helyen (15 663 tweet) egy nem specifikus információközlő cég van, azonban a második helyen, bár nem patient kategóriában, hanem Individual Non-Health stakeholderként, egy érintett, ahogy a harmadik helyen is, méghozzá egy anya, aki a #Orkambi révén is kapcsolódik szorosan a témafolyamhoz. Az Orkambi a Vertex gyógyszergyár által gyártott terápiát adó készítmény a betegek számára, és 2020-tól Magyarországon is elérhető. A gyógyszergyár azonban nincs tweetszám szerint a vezető 3 véleményvezér között, a 4. helyen viszont jelen van egy robot felhasználó, egy cf_bot, amely retweet tevékenységet végez a CF témájában nyíltan a profilja szerint.

Nemzetközi Org. Advocate stakeholderek a retweet szerinti rangsorolás szerint vannak a vezető három pozícióban (9. ábra), egyaránt 25 ezer feletti követőszámmal @CF_Foundation (28 675 követő), @cftrust (25 207 követő). Előbbi globálisan vezető helyet foglal el a terápia iránti kutatás területén, utóbbi pedig az Egyesült Királyságban található, jótékonyági alapú szervezet a cisztás fibrózissal élőkért.

A retweet rangsor és a hiteles információközlés kapcsolatában megfigyelhető, hogy a megosztásokat legnagyobb arányban ellenőrzött, valós szervezetek végzik (9. ábra), ez igaz az első két pozícióra is, a harmadik esetében pedig a Symplur algorit-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

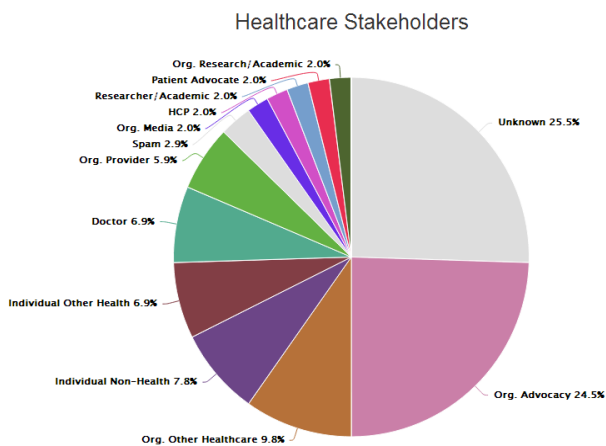
mus tanítása révén Patient kategóriába kerülhetne a felhasználó, ez esetben javítva az arányukat és megmutatva a jelentőségüket a tartalommegosztás terén.

Row	Retweets	Thumbnail	Username	Name	Description	Healthcare Stakeholder	Followers
1	24,917		@CF_Foundation	Cystic Fibrosis Foundation	The Cystic Fibrosis Foundation is the world's leader in the search for a cure for cystic fibrosis.	Org. Advocacy	28,675
2	14,301		@cftrust	Cystic Fibrosis Trust	The Cystic Fibrosis Trust is the only UK-wide charity dedicated to fighting for a life unlimited for everyone affected by cystic fibrosis	Org. Advocacy	25,207
3	6,954		@emmagadie2009	Emma Gadie	My son Oily and my daughter Grace are my whole world. Couldn't imagine my life without them! #OrkambiNow	Unknown	411
4	6,481		@myCFHero	my CF Hero	Mother to a 7 year old girl with the life limiting condition #CysticFibrosis. Fighting for access to medicine #Orkambi Don't make her suffer. #DoTheRightThing	Unknown	636

9. ábra. Influenszer rangsor retweetek szerint a #CysticFibrosis adatbázisnál
Forrás: www.symplur.com

#Asthma

#Asthma Influencers by Tweets



Source: Symplur

10. ábra. Influenszer megoszlás tweetek alapján rangsorolva
Forrás: www.symplur.com



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A legnagyobb adatbázisnál az $\frac{1}{4}$ részt ismeretlen felhasználók adják, a TOP 100 esetén ez 25 user, a tanácsadó szervezetek közelítik meg őket (24,5 százalék), betegek viszont mindössze 2 százalékban vannak jelen, ahogy az orvosok sem érik el a 10 százalékot sem (6,9 százalék) A Researcher/Academic stakeholder csoport 2 százalékot tesz ki (10. ábra).

Összehasonlítva, a #Asthma esetében az ismeretlen felhasználók már nem a felét teszik a mintának, a Patient Advocate kategória aránya viszont itt még alacsonyabb, mindössze 2 százalék a 7. ábrával történő összevetés során, és fontos azt is látni, hogy itt már jelen van a kutatási szféra is, széles spektrumú összetételt mutat a véleményvezérek köre.

A #Asthma esetében elmondható, hogy az első három helyen ismeretlen felhasználók állnak tweetszám alapján. Ez meglepő és továbbra is hitelesítési problémákat felvető eredmény, tekintve, hogy ezen cég, ezen algoritmus alapján az tűnt ki, hogy a legtöbb tweetet azok bocsátják ki az online térbe, akik nem kategorizálhatóak, és a mennyiségi jellegükből adódóan nagyon sok információt osztanak meg, ezekkel pedig a felhasználók találkozhatnak.

Row	Tweets	Thumbnail	Username	Name	Description	Healthcare Stakeholder
1	37,806		@vapeherb	Herb Vape		Unknown
2	18,778		@miraclecuresyou	Miracle Cures	Miracle cures do exist & don't involve Zionist-infected drugs. A living Christ is channeling miracles. If you deserve one then you'll receive. Follow Erol.	Spam
3	16,058		@NoMoreAsthma	No More Asthma	Natural Approach To Curing Asthma	Unknown
4	15,213		@myid_research	MyID Research	I help lead a team of highly talented developers and analysts who have built a very powerful healthcare app which tracks all aspects of your medical history.	Researcher/Academic
5	15,032		@dprescriptio	Free RX	Nurses, doctors, social workers, clinics, hospitals, learn how you can help your patients save on their prescription medication.	Org. Other Healthcare

11. ábra. Influenszer rangsor tweetszám szerint #Asthma esetében
Forrás: www.symplur.com

Azonban ebben az esetben is érdemes a retweet számot tekinteni, méghozzá azért, mert ez egy interaktív kapcsolatot is adhat és orientálja a felhasználókat, itt viszont már az Asthma UK van a vezető helyen, betegszervezetként, moderált tartalommal a felhasználói fióknál, tehát egy olyan csatorna dominál, amely ellenőrzött és a betegek számára is megbízható információszerezés céljára.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A SymplurRank eredménye (12. ábra) a digitális egészségügy és az e-páciensek szempontjából is jelentős, az AsthmaUK az algoritmus szerint is a legnagyobb hatású a beszélgetésfolyamban, azonban a második és harmadik helyre egyaránt orvosokat sorolt be, jelen volt ebben Dr. Dave Stukus gyermekorvos, asztmaspecialista, aki egyben közösségi média szakértő is. Őt követte Ivancevich doktor, a Salvador Egyetem docense, allergológus, aki immunológiai témában is rendszeresen részt vesz a beszélgetésekben. Ez a tájékoztató beteg esetében megnyugtató eredmény, ugyanis hiába a nagy tweet szám, a valós befolyási erőt a betegszervezet mellett két orvos biztosította az első három helyen, ebből következően a szakmaiság nagyobb mértékű lehet és biztosabb a betegedukáció szempontjából. Mindez azon kutatási eredményt is jelenti, hogy a nagy tweetszám nem egyezik a befolyásolási erővel, továbbá a TOP 100 influenzერი eredménnyel és az alacsony orvosi részvétellel ellentétben, a konverzáció jelentős pillérét adták az orvosok, ez pedig a Twitter professzionális használatával egybevág orvosi részről, másrészt a fals tájékoztatót is csökkentheti.

Row	SymplurRank	Thumbnail	Username	Name	Description	Healthcare Stakeholder
1	87.7		@asthmauk	Asthma UK	Every 10 seconds someone has a potentially life-threatening #asthma attack. Our aim is to stop asthma attacks and cure asthma. Monitored Mon-Fri 9-5.	Org. Advocacy
2	84.2		@AllergyKidsDoc	Dr. Dave Stukus	Pediatric allergy/asthma specialist @nationwidekids - @aaaai_org Social Media Medical Editor -Dispelling myths one tweet at a time, often zigging when others zag	Doctor
3	83.5		@Aller_MD	Juan C Ivancevich MD	Allergist - Twittering on #allergy, #asthma & #immunology. Associate Professor of Immunology, del Salvador University, WAO website Managing Editor	Doctor
4	83.5		@atscommunity	American Thoracic	We Help the World Breathe	Org. Advocacy
5	80.7		@AAAAI_org	AAAAI	The American Academy of Allergy, Asthma & Immunology consists of medical professionals devoted to optimal patient care in allergy, asthma and immunology.	Org. Advocacy
6	79.2		@AAFANational	AAFA National	Asthma and Allergy Foundation of America. Community, Advocacy, Research, Education and Support. #asthmacapitals #allergycapitals	Org. Advocacy
7	78.9		@AsthmaCanada	Asthma ca	Asthma Canada's mission is to help Canadians with asthma lead healthy lives through education, advocacy and research.	Org. Advocacy
8	78.9		@accpchest	CHEST	American College of Chest Physicians® (CHEST®), www.chestnet.org	Org. Advocacy

12. ábra. SymplurRank – #Asthma influenzerek esetében

Forrás: www.symplur.com

A Symplur Rank¹⁵ rangsorolása tehát, amely a Google algoritmusához hasonló mérési módszerekkel értékeli, a valódi relevanciát emeli ki, és ebben az esetben a felhasználók közötti kommunikáció és a fiókok, valamint az asztma témájához való

¹⁵ <https://help.symplur.com/en/articles/1891111-simplurrank>



TEMATIKUS TANULMÁNYOK – Digitális szociológia

kapcsolódás okán mégiscsak azt állította, hogy az anomális viselkedést mutató felhasználók bár jelentős mennyiségű tweetet adnak közre, ugyanakkor nem tudtak arányosan akkora hatást kifejteni és agenda-settinget megvalósítani, mint a valódi és hiteles tartalmat közlő betegszervezet és az orvosok, amely ebben az esetben a Twitter szakmai és nemcsak közösségi jellegét ábrázolja.

Ez természetesen nem jelenti azt, hogy más algoritmus ne juthatna más eredményre, ez kizárólag a Symplur adatbázisára, a Symplur algoritmusára jellemző eredmény, azonban ezeket az összehasonlító elemzéseket azért célszerű elvégezni, mert rámutat, hogy a nagy tweetszám önmagában nem elég, ha nincs egy megfelelő kapcsolati hálóba ágyazva, megosztva és csak részlegesen kapcsolódik a main-stream tartalmakhoz.

Szemantikai elemzés – Miről beszélnek a felhasználók

A két elemzett adatbázis során a 500 vezető szóból az alábbiakban a TOP 100 található meg a Symplur adatai alapján, ezek kerültek elemzése tweetszám szerint rangsorolva a kutatás során (1. táblázat). Érdekes eredmény, hogy két adatbázis esetben a TOP 100 szó nem tartalmazza a kórház (hospital) szót. A #cysticfibrosis az egyetlen, ahol nem a hashtag neve van az első helyen, hanem a help szó, amely a ritka betegségek és az online betegközösségek szerepét is mutatja, a segítségkérés és az információkeresés fokozott igényével.

A cisztás fibrózis esetében érdekes, hogy gyógyszernevek közül egyedül az Orkambi jutott 100. helyen a listába, amely a betegek számára Magyarországon is elérhető, fontos gyógyszerkészítmény. A support szó a #Asthma esetén csak a 85., míg a másik adatbázisban a 10. helyen van. A nagyobb adatbázis esetén a research szó a 21. helyen, a ritka betegségben a 43. pozícióban található. Az Asthma esetében a dohányzásra érdemes koncentrálni, ez a másik esetben meg sem jelenik.

A szavak esetében további, részletesebb elemzés is indokolt lehet, kvalitatív esz-közökkel is. Ebben az esetben a Symplur révén nem volt nyerhető olyan elemezhető korpusz, amelyen NLP elemzés (természetes nyelvfeldolgozás módszertana) véghez vihető lett volna, a jövőben érdemes erre is fókuszálni, azonban ez esetben minőségi kritérium, hogy olyan tweetek álljanak rendelkezésre, amelyek több szöveget, kevesebb linket, hangulatjelet (emojit) tartalmaznak, tehát a nyelvi elemzéshez ne töredett tartalom álljon rendelkezésre. Ez elsősorban stakeholder bontás alapján lenne véghez vihető az eddig feltártak alapján, ugyanis láthatóvá vált az elemzés során, hogy a szervezeti kommunikáció (például a betegszervezeteké) nagyobb mértékben ad közre hosszabb, szöveges tartalmat, mint az egyéni felhasználók.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

1. táblázat. A TOP 100 szó a #Asthma és a #CysticFibrosis esetében

	#Asthma	#Cysticfibrosis
	asthma	help
	smoke	cystic
	benefit	fibrosis
	symptoms	people
	help	life
	children	awareness
	air	raise
	people	today
	study	thank
	risk	support
	know	need
	health	day
	treatment	patients
	today	make
	kids	thanks
	patients	great
	use	symptoms
	severe	pls
	thanks	know
	learn	cure
	research	ago
	disease	alerts
	attacks	fight
	tips	share
	inhaler	living
	attack	lost
	control	sign
	care	lung
	day	time
	allergy	petition
	read	little
	triggers	drug
	need	access
	post	fighting
	chronic	join
	life	community



TEMATIKUS TANULMÁNYOK – Digitális szociológia

	check	live
	breathe	story
	make	year
	child	children
	childhood	old
	free	memory
	breathing	research
	common	amazing
	latest	disease
	time	read
	respiratory	week
	school	money
	great	good
	quality	love
	lung	important
	trigger	learn
	better	check
	allergies	years
	good	donate
	home	raising
	reduce	event
	pollution	lives
	cause	team
	management	care
	treat	family
	follow	month
	year	video
	million	hospital
	awareness	come
	lungs	lungs
	cold	news
	high	young
	work	walk
	news	best
	linked	transplant
	manage	age
	improve	add
	living	happy



TEMATIKUS TANULMÁNYOK – Digitális szociológia

	adults	works
	prevent	retweet
	suffer	gene
	sufferers	watch
	important	friend
	love	item
	support	work
	plan	tomorrows
	way	treatment
	talk	cause
	allergic	away
	join	tomorrow
	natural	doing
	conditions	item
	world	sister
	using	hope
	exposure	available
	inhalers	saving
	info	charity
	associated	health
	medication	going
	best	post
	action	patient
	problems	night
	alerts	needs
	clinical	orkambi

Forrás: Saját szerkesztés Symplur adatbázis alapján

Hashtagek

A hashtagek vizsgálata azért fontos, mert ha ezeket címkeként tekintjük egy online mezőben, akkor a tartalom ennek révén kapcsolódik egymáshoz és kap gyűjteményes jelleget, ebből következően pedig érdekes lehet, hogy a felhasználók mely kommunikációs mezőbe kívánják elhelyezni a tartalmat és ehhez mit használnak fel.

A hashtagek esetében az adatbázisok mérete és különbözősége miatt egységes limit akkor lehetséges, ha a kisebbhez is igazodik a tweetszám, az viszont a #Asthma esetében jóval nagyobb elemszámot és TOP rangsorolást eredményezett volna. Ezért a vezető 10 hashtag került elemzésre és a TOP 10-en kívüli értékeket kitekintő jelleggel érdemes említeni, amennyiben releváns.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

#cysticfibrosis

Row ▲	Rank ↕	Hashtag	Tweets ↕
1	1	#cysticfibrosis	643,175
2	2	#OrkambiNow	55,658
3	3	#cf	41,421
4	4	#Orkambi	29,158
5	5	#CureCf	18,002
6	6	#JODIEDIXON	14,585
7	7	#cfawareness	13,155
8	8	#cfaware	10,979
9	9	#GreatStrides	9,113
10	10	#cysticbot	7,861
11	11	#65roses	6,935
12	12	#PrecisionMedicine	6,894

13. ábra. Hashtagek a #CysticFibrosis adatbázisban

Forrás: www.symplur.com

A #CysticFibrosis esetében kétszer is szerepel az Orkambi nevű gyógyszerre utaló hashtag (13. ábra), amely oki terápia, de megjelenik a személyre szabott orvoslás is, amely kapcsolódik a genomikához, és 18 002 tweet esetében volt megtalálható a CF gyógyítása is. Összességében, összevonás után 84 816 tweetben volt jelen az Orkambi, tehát az adatbázis neve után ez a második leggyakoribb, kommunikációs térbe terelő címke. Érdekes, hogy megjelenik a #cysticbot, amiről már az álhírek esetében volt szó, ez esetben felveti a kérdést, hogy a hashtag általi jelölés mit közvetít a felhasználó számára. Erről többletinformációt kvalitatív módszerrel lehetne feltárni a felhasználók megkérdezése révén, az azonban kitűnik, hogy a betegek bár a szavak szintjén ez nem jelent meg markánsan (1. táblázat), itt azonban a terápiás lehetőséget és a gyógyszerkészítményt jelölő hashtaget preferálják, tehát fontosabbnak tartják a csatornát, mint azt, hogy a szövegben ezt megjelölik a szavak szintjén.

#Asthma

A hastagek esetében a nagyobb adatbázis esetében vezet az adatbázis alapját jelentő hashtag (1 020 456 tweet), de megjelenik a 4. helyen #COPD, a 7. a #WorldAsthma-Day, valamint a #respiratory és az #allergy is a TOP 10-ben szerepel. A betegek csoportjában a hastageket bontás után vizsgálva 8. helyen a #depression áll (1448 tweet), ami ugyan kevés, de felhívja a figyelmet a betegség mentális oldalára is, amely a ritka betegségek esetén is igaz.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Földrajzi elhelyezkedés

A geolokáció meghatározása függ attól ilyen jellegű elemzések esetében, hogy a felhasználó a készülékén adott-e hozzáférést a tartózkodási helyéhez. A kutatás ezen esetben feltáró jelleggel tudta megmondani, hogy az érintettek ezt nem tették, tehát nem állapítható meg, hogy mely kontinensről és honnan vesznek részt az interakcióban, ezért az általános Twitter adatokat lehet maximum közölni, ezen két adatbázis esetében mind a két adatbázist tekintve olyan alacsony a felhasználói metaadat, hogy ez a jelenlegi módszerrel nem vizsgálható, ezért a fentiekben nem közölt.

Összegzés

A tanulmányban a digitális egészségügy elméleti megközelítéseinek és piaci viszonyain túl, az e-páciensek kerültek bemutatásra két krónikus betegség Big Data alapú kutatása révén. A digitális egészségügy összetett fogalom, magában foglalja a technológiai és humán aspektusokat, több mint 1700 definíciója közül a tanulmány azt a megközelítést választotta, mely reflektál a felhasználói körre, illetve annak szükségleteire. A digitális egészségügy ugyanakkor a világ egyik vezető piaci környezete, elsősorban az USA dominanciájával. A megjelenő tendenciák legkorábban tehát elsősorban ott figyelhetőek meg, ezt követően nemzetközi viszonylatban is elérhetővé válnak, célszerű ezért figyelemmel kísérni a trendeket az Egyesült Államokban. A fogalmi keret mellett, a tanulmány középpontjában tehát az e-páciensek állnak, szerepük pedig modellezhető a ritka betegségekben érintettek révén, akik tudatos, elkötelezett betegségmentesmentet alkalmaznak széleskörű információszerezés mellett, ennek a folyamatnak pedig megkerülhetetlen színtere lett a közösségi média. Módszertani okokból kifolyólag nem a legnagyobb közösségi média felületre, hanem a Twitterre esett a választás, amely követési struktúrája miatt lehetővé tett összehasonlítható Big Data elemzést social média listening platform bevonása mellett. A kutatás egy ritka és egy gyakori betegséget hasonlított össze, az elemzés mindkét esetben tüdő érintettségű betegséget fedett le. A kutatás megmutatta, hogy a #CysticFibrosis esetében a felhasználói aktivitás emelkedő tendenciát mutatott be a vizsgált 7 év alatt, és ez a trend nem fordult meg, szemben a gyakori betegséget és nagy felhasználói táborát egyesítő #Asthma esetével. A felhasználói elkötelezettség esetében láthatóvá vált, hogy 70 százalék fölött van mindkét esetben a megosztása azoknak, akik mindössze egyetlen tweetet küldtek el a vizsgálati időszak alatt, amely érdekes jelenség, tekintve, hogy social média felületről van szó, ahol a tartalom közzétevése lehetséges. A nyelvhasználatot illetően látható vált, hogy az angol nyelv dominanciája mindkét esetben jelen van a tweeteket tekintve a felület használati szokásainak megfelelően. A véleményvezérek tekintve, a cisztás fibrózis esetében tweet alapú kategorizálás mutatta meg, hogy nagy számban vannak jelen nem azonosítható fel-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

használók, akiknek szerepe lehet a tartalom hitelességében, ugyanakkor más módszertani kategorizálás, a social média listening saját algoritmus besorolása rámutatott a #Asthma esetében, hogy milyen jelentős szerepe van a betegszervezeteknek és a betegeknek az online interakcióban, ennek révén pedig a valóban hiteles információt közlő felhasználók jelenléte is kiemelt. A szavak elemzése során látható vált, hogy a kórház, a gyógyszernevek, a kezelés kifejezések nem dominálnak jelentős mértékben, ugyanakkor az asztma esetében a dohányzás, mint jelentős kockázati tényező, igen. A hashtagek ezzel szemben már reflektálnak a kezelésre, a tisztaság fibrozis második leggyakrabban használt címkéje az oki terápiát nyújtó gyógyszerkészítmény neve. Az elemzés tehát feltáró jelleggel képet adott a mikroblogon folyó kommunikációs mintázatokról, valamint két egészségügyi témájú adatbázist mutatott be. A kutatás további iránya elsősorban kvalitatív jellegű lehet a felhasználókkal történő direkt kommunikáció révén, továbbá a koronavírus járvány alatti kommunikáció összehasonlítása megegyező adatbázisok használatával, annak érdekében, hogy ismereteket nyerjünk a közösségi média használati szokásokról a járvány előtt és alatt, orvosi és társadalomtudományi megközelítésben egyaránt.

Irodalom

- Az Emberi Erőforrások Minisztériuma szakmai irányelve az asztma diagnosztikájának, kezelésének és orvosi gondozásának alapelveiről felnőttkorban. 2018. EüK 13. szám EMMI szakmai irányelvhatályos 2018. 09. 12.
https://www.hbcs.hu/uploads/jogszabaly/2761/fajlok/EMMI_szakmai_iranyelve_asztma.pdf (utolsó letöltés: 2021. 11. 22.)
- Banks, M. A (2020): Sizing up Big Data. *Nature Medicine* 26 (5-6)
<https://www.nature.com/articles/s41591-019-0703-0> (utolsó letöltés: 2021. 11. 22.) <https://doi.org/10.1038/s41591-019-0703-0>
- Chin, J. J. (2002): Doctor-patient Relationship: from Medical Paternalism to Enhanced Autonomy. *Singapore Med Journal* Vol 43(3): 152–155., <https://pubmed.ncbi.nlm.nih.gov/12005343/> (utolsó letöltés: 2021. 11. 22.)
- Cohen, A. B. – Dorsey E. R. – Mathews, S. C. – Bates, D. W. – Safavi, K. (2020): A digital health industry cohort across the health continuum. *Digital Medicine* <https://www.nature.com/articles/s41746-020-0276-9> (utolsó letöltés: 2021. 11. 22.)
- Ćwiklicki, M. – Schiavone, F. – Klich, J. *et al.* (2020): Antecedents of use of e-health services in Central Eastern Europe: a qualitative comparative analysis. *BMC Health Serv Res* **20**, 171. <https://doi.org/10.1186/s12913-020-5034-9>
- DeBronkart, D. (2013): “Let Patients Help”: a patient engagement handbook
<https://www.epatientdave.com/let-patients-help/> (utolsó letöltés: 2021. 11. 22.)



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Dessewffy T. – Láng L (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika* 92–93 (2015/3–4. szám): 157–170.
http://www.replika.hu/system/files/archivum/92-93_11_dessewffy_lang.pdf
(utolsó letöltés: 2021. 11. 22.)
- EMMI irányelv. https://www.hbcs.hu/uploads/jogszabaly/2761/fajlok/EMMI_szakmai_iranyelve_aszmta.pdf (utolsó letöltés: 2021. 11. 23.)
- Fatehi, F. – Samadbeik, M. – Kazemi, A. (2020): What is Digital Health? Review of Definitions. *Studies in Health Technology and Informatics*. <https://pubmed.ncbi.nlm.nih.gov/33227742/> (utolsó letöltés: 2021. 11. 22.)
PMID: 33227742 DOI: 10.3233/SHTI200696
- Fergusson, T. (2007).: White Paper. E-patients. How they can help us heal health care. https://participatorymedicine.org/e-Patient_White_Paper_with_Afterword.pdf
(utolsó letöltés: 2021. 11. 22.)
- Fergusson, T. (2004): The first generation of e-patients. These new medical colleagues could provide sustainable healthcare solutions. *BMJ* 2004 May 15; 328(7449): 1148–1149. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC411079/> (utolsó letöltés: 2021. 11. 22.)
- Global Digital Health Market was Valued at USD 111.4 billion in 2019 and is Expected to Reach USD 510.4 billion by 2025, Observing a CAGR of 29.0% during 2020–2025: VynZ Research (2020.05.23.)
<https://www.globenewswire.com/news-release/2020/05/23/2037920/0/en/Global-Digital-Health-Market-was-Valued-at-USD-111-4-billion-in-2019-and-is-Expected-to-Rich-USD-510-4-billion-by-2025-Observing-a-CAGR-of-29-0-during-2020-2025-VynZ-Research.html> (utolsó letöltés: 2020. 12. 01.)
- Goff, D. A. – van den Bergh, D. (2015): Twitter: A tool to improve healthcare professionals’ awareness of antimicrobial resistance and antimicrobial stewardship. *South African Medical Journal. SAMJ, S. Afr. med. j. vol.105 n.5 Pretoria May. 2015*
http://www.scielo.org.za/scielo.php?pid=S0256-95742015000500036&script=sci_arttext&tlng=en (utolsó letöltés: 2021. 11. 22.)
- Gopal, G. – Crazzolara, C. S. – Toldo, L. – Eberhardt, W. (2018): Digital transformation in healthcare – architectures of present and future information technologies. *Clinical Chemistry and Laboratory Medicine (CCLM) Volume 57: Issue3* <https://www.degruyter.com/view/journals/cclm/57/3/article-p328.xml> (utolsó letöltés: 2021. 11. 22.)
- Jimenez, G. – Spinazzea, P. – Matcharc, D. – Huat, G. K. C. – van der Kleij, R. M. J. J. – Chavannes, N. H. – Car, J. (2020): Digital health competencies for primary healthcare professionals: A scoping review. *International Journal of Medical Informatics*,



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Volume 143, november 2020, 104260 <https://www.sciencedirect.com/science/article/abs/pii/S1386505620311473> (utolsó letöltés: 2021. 11. 22.)
- Kazai A. (2012): Google páciensből e-páciens. Medical Online (2012) http://medicalonline.hu/informatika/cikk/google_paciensbol_e_paciens (utolsó letöltés: 2021. 11. 22.)
- Kostkova, P. (2015): Grand challenges in digital health. Front Public Health. 2015; 3:134. <https://doi.org/10.3389/fpubh.2015.00134>
- Máriáss M. (2018): Asztma tünetei és kezelése. https://www.hazipatika.com/betegsegek_a_z/asztma/41 (utolsó letöltés: 2021. 11. 22.)
- Meskó B. – Drobni Zs. – Béneyi É. – Gergely B. – Gyórfly Zs. (2017): Digital health is a cultural transformation of traditional healthcare. Mealth 2017 Sep 14 ;3:38. doi: 10.21037/mhealth.2017.08.07. eCollection 2017. <https://bit.ly/2si6ulh> (utolsó letöltés: 2021. 11. 22.)
- Meskó B. – Gyórfly Zs. (2012): E-doktorok és e-páciensek Magyarországon. Kvalitív vizsgálat a magyarországi orvosok gyógyítással kapcsolatos internethasználatáról és attitűdjeiről. LAM 2012;22(12):677–683.
- Pilling J. (2017): Orvos-beteg kommunikáció Pilling Jánostól. <https://egeszsegugymarketing.hu/2017/03/13/orvos-beteg-kommunikacio-pilling-janostol/> (utolsó letöltés: 2021. 11. 22.)
- Raghupathi, W. – Raghupati, V. (2018): An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health. International Research and Public Health 15(3),431 <https://www.mdpi.com/1660-4601/15/3/431/htm> (utolsó letöltés: 2021. 11. 22.)
- Scotet, V. – L’Hostis, C. – Férec, C. (2020): The Changing Epidemiology of Cystic Fibrosis. Incidence, Survival and Impact of the CFTR Gene Discovery. Genes (Basel) v.11(6); 2020 Jun PMC7348877 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7348877/> (utolsó letöltés: 2021. 11. 22.)
- Silver, J. D. – Sutherland, M. F. – Johnston, F. H. et al. (2018): Seasonal asthma in Melbourne, Australia, and some observations on the occurrence of thunderstorm asthma and its predictability. *PLoS One*. 2018;13(4) doi:10.1371/journal.pone.0194929 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5896915/> (utolsó letöltés: 2021. 11. 22.)
- Simon, S. (2021): #RareDisease and Information Acquisition Process in the Word of Digital Healthcare: A Big Data-based Analysis on Twitter in the Context of Rare Diseases and E-patients. *Neue Grenzen – New Frontiers*. Budapest, Gondolat Kiadó



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Statista 1: <https://www.statista.com/statistics/387867/value-of-worldwide-digital-health-market-forecast-by-segment/> (utolsó letöltés: 2020. 01. 03.)
- Statista 2: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> (utolsó letöltés: 2021. 11. 22.)
- Statista 3: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> (utolsó letöltés: 2021. 11. 22.)
- Statista 4: <https://www.statista.com/topics/2297/twitter-marketing/#dossier-Keyfigures> (utolsó letöltés: 2021. 11. 22.)
- Steinhubl, S. R. – Topol, E. J. (2018): Digital medicine, on its way to being just plain medicine. *npj Digital Med* 1, 20175 <https://www.nature.com/articles/s41746-017-0005-1> (utolsó letöltés: 2021. 11. 23.)
- Symplur 1: <https://www.symplur.com/blog/ebola-social-media-first-look-facts/> (utolsó letöltés: 2021. 11. 23.)
- Symplur 2: <https://www.symplur.com/healthcare-hashtags/> (utolsó letöltés: 2021. 11. 23.)
- Thien, F. – Beggs, P. J. – Csutoros, D. – Darvall, J. – Hew, M. – Davies, J. M. et.al. (2018): The Melbourne epidemic thunderstorm asthma event 2016: an investigation of environmental triggers, effect on health services, and patient risk factors. *Lancet Planetary Health* Volume 2, ISSUE 6, e255-e263, June 01, 2018 [https://www.thelancet.com/journals/lanplh/article/PIIS2542-5196\(18\)30120-7/fulltext](https://www.thelancet.com/journals/lanplh/article/PIIS2542-5196(18)30120-7/fulltext) (utolsó letöltés: 2021. 11. 22.)
- WHO Newsroom (2020): Asthma <https://www.who.int/news-room/fact-sheets/detail/asthma> (utolsó letöltés: 2021. 11. 23.)
- Zsuga J. (é. n.): Az asztma és tünetei. <https://www.webbeteg.hu/cikkek/asztma/286/az-asztma-es-tunetei> (utolsó letöltés: 2021. 11. 23.)



A depresszió keretezési módjainak vizsgálata online fórumokon természetesnyelv-feldolgozással

MÁTÉ FANNI¹

ABSZTRAKT

A depresszió jelenségének vizsgálata nem újkeletű a szociológiában, de egyre szélesebb társadalmi problémává válása folytán ma is aktuális. A depresszióhoz kapcsolódó biomedikális és pszichológiai aspektus mellett egyre figyelemre méltóbb a szociológiai nézőpont beemelése a depresszió lehetséges okairól zajló párbeszédbe. A depresszióról szóló diskurzus kutatásában az online elérhető szövegek számos új lehetőséget nyújtanak, hiszen a fórumok anonimitása, egyszerű elérése népszerűvé teszi az online segítségkérést az érintettek számára. Kutatásomban természetesnyelv-feldolgozást, azon belül logisztikus regressziót használtam, hogy feltérképezem a depresszió értelmezési kereteinek laikus diskurzusokban való megjelenési mintáztatának gépi tanuló algoritmussal való feltérképezhetőségét. A gépi tanulás által ugyanis lehetővé vált, hogy 67857 hozzászólást elemezzek, amelynek emberi erőforrással való feldolgozása nehezen valósulhatott volna meg. Az elemzés során angol nyelvű, depresszió témájú online fórumok bejegyzéseinek kategorizálást végeztem el a tudományos diskurzus-kategóriák mentén. Első eredményeim azt mutatják, hogy a logisztikus regresszió mérsékelten jobb eredményt nyújtott, mint a véletlenszerű besorolás, egyben az emberi besoroláshoz hasonló eredményt ért el. Annak ellenére, hogy a kutatás angol nyelvű fórumokat elemzett, eredményeim hasznosak lehetnek bárki számára, aki absztrakt szociológiai fogalmak felhasználók által írt online szövegekben való megjelenésével foglalkozik.

KULCSZAVAK: depresszió, természetesnyelv-feldolgozás, mentális egészség, online fórumok

ABSTRACT

Classification of depression-related online forums using Natural Language Processing

The study of the phenomenon of depression is not new in sociology, but since the depression is becoming a wider social problem, it is still a relevant issue today. In addition to the biomedical and psychological aspects of depression, the sociological perspective is becoming more noteworthy in the discourse about the causes of depression. In the research of the discourse on depression, the online texts offer many new possibilities, as the forum's anonymity and accessibility make the online seeking for help popular. In this research, natural language

¹ Eötvös Loránd Tudományegyetem, Társadalomtudományi Kar, email: mate.fanni@tatk.elte.hu



TEMATIKUS TANULMÁNYOK – Digitális szociológia

processing (logistic regression) was applied to find patterns in the definition of depression in lay discourses. These methods make it possible to analyze a large amount of text - which would have been difficult to process with human resources. During the analysis, 67 857 posts of English-speaking online forums were categorized along the categories of the scientific discourse about depression. This study presents the first results, which shows logistic regression classifier performs like the annotators. . Although the research has analyzed English-speaking forums, my findings may be useful to anyone observing abstract sociological concepts in online texts written by users.

KEYWORDS: *depression, Natural Language Processing mental health, online forums*

Bevezetés

A depresszió online szövegekre alapuló kutatása új távlatokat nyit meg a szociológiában. Egyfelől amiatt, mert a platformok és azokon keresztül a hasonló tapasztalatokkal rendelkező személyek könnyen, anonim módon elérhetőek, így népszerűek a depresszióval, szorongással élők egyre növekvő csoportjában. Másfelől az oldalakon keletkező szövegek nem kutatási célból álltak elő, amely kétségtelen előnye, ahogy Dessewffy és Láng is kiemeli (2015), hogy a „hagyományos” kutatások kérdezési szituációjában olyan véleményeket ismerünk meg, amelyeket alakított a kérdezettek elvárásoknak való megfelelésének igénye (lásd: megfigyelői paradoxon (Labov 1972: 61)), sőt a vélemények kérdés nélkül talán meg sem születtek volna. Ezzel szemben az interneten elérhető viselkedések ténylegesek, nem torzítja őket a kutatói igényeknek való megfelelés. A szerzők ugyan elsősorban a strukturált, számszerűsíthető digitális lábnyom kapcsán említik ezt az előnyt, de a megnyilatkozásokra szintén igaz, hogy a véleményeket, hozzászólásokat nem a kutatási szituáció hozta létre, azt külső tényezőként legfeljebb az adott tér normái, szabályai alakítják, de azok is organikus részét képezik a diskurzus terének.

Jelen tanulmányban a depresszió értelmezési módjait, annak okairól és lehetséges kezelési módjairól folyó globális, laikus diskurzust vizsgálom két aspektusból. A kutatás elsődleges célja, hogy hozzájáruljon a természetesnyelv-feldolgozás társadalomtudományi alkalmazhatóságának felméréséhez laikus diskurzusok elemzése nyomán. A vizsgálat abból a feltételezésből indul ki, hogy a szakértői diskurzusban megtalálható keretelési módok a laikus diskurzusokban is megjelennek. A kutatás implicit módon ezt a feltételezést is vizsgálja, amennyiben az osztályozás sikeressége azon is múlik, párhuzamba állítható-e a laikus és a szakértői diskurzus. Annak megértése, hogy az érintettek miként vélekednek a depresszióról, betegségként, lelki, avagy társadalmi problémaként tekintenek inkább rá, a hatékony segítségnyújtás formáit is meghatározhatja. Az online fórumokon zajló párbeszéd tartalmának ismerete amiatt is értékes a szakemberek számára, mert az online fórumok könnyen elérhető információforrásként szolgálnak az érintetteknek.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Tanulmányomban először bemutatom a depresszió fogalmát és különböző (biomedikális, pszichológiai és szociológiai) keretelési módjait, majd depresszióval kapcsolatos szövegeket vizsgáló kutatásokat ismertetek. Kitérek hagyományos kvalitatív tartalomelemzést, és újabb, gépi tanulást alkalmazó kutatásokra is. Az elméleti bevezető után a használt adatbázist, annak előfeldolgozását és az alkalmazott algoritmusokat mutatom be. A tanulmány utolsó részében az eredményeimről és azok lehetséges magyarázatáról, értelmezéséről írok – visszacsatolva ezeket fenti-ekben leírt kutatási kérdésekhez.

A depresszióról

Az Amerikai Pszichiátriai Társaság (APA) 2013-ban kiadott szakkönyve² szerint akkor állítható fel a depresszió diagnózisa, ha az alábbi³ kilenc jellemző közül az első vagy második és további legalább négy jellemző legalább két héten keresztül megfigyelhető:

- levertség
- érdeklődés, öröme való képesség elvesztése
- étvágyban vagy testsúlyban bekövetkező, nagymértékű növekedés vagy csökkenés
- megnövekedett vagy csökkent alvásigény, álmatlanság
- pszichomotoros agitáció (nyugtalanosság) vagy gátoltság (meglassultság)
- fáradtság, energiahány
- csökkent önértékelés, értéktelenség érzése, inadekvát büntudat érzése
- gondolkodásra való képtelenség, döntésképtelenség, koncentrációhiány
- halálra vonatkozó gondolatok gyakorisága, öngyilkossági gondolatok.

Az Egészségügyi Világszervezet (WHO) adatai alapján 2017-ben világszerte több mint 300 millió ember szenvedett depressziótól, s világszinten 2015-ben az összes haláleset 1,5 százaléka öngyilkosság volt, amely a depresszió egyik súlyos következménye lehet (World Health Organization 2017). Az előbbi adat természetesen csak iránymutatás, hiszen az egyes régiók, országok között széles skálán mozog a depresszió diagnosztizálásnak gyakorlata. Az adatokat továbbá azért is óvatosan kell kezelnünk, mert a depresszióhoz és a mentális betegségekhez erős társadalmi stigma tartozik (Rogers – Pilgrim 2005), amely befolyásolhatja az orvoshoz, pszichológushoz fordulás valószínűségét, így a diagnózis felállítását is. Ezen tényezők azonban csak tovább növelnék a megfigyelt előfordulások számát, így a fenti eset-számokra leginkább alsó becslésként tekinthetünk.

² Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

³ A magyar fordításhoz Torzsa és munkatársai tanulmányában (2009) szereplő kifejezéseket használtam.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A depresszió keretezési módjai

A „depresszió keretezése” alatt a depresszió értelmezésének egyének (a szakértők és a laikusok) vagy intézmények általi kijelölését értem, azt a fogalmi keretet, amely mentén a depresszióra tekintenek.⁴ A tudományos diskurzusban (Busfield 2000, Comer 2013, Rogers – Pilgrim 2005) három keretezés különíthető el egymástól: (1) a biomedikális, amely a depressziót elsősorban szomatikus eredetűnek írja le, s azt gyógyszeres úton kezeli; (2) a pszichológiai, amikor lelki eredetű problémaként jelenik meg a depresszió és végül (3) a szociológiai, amely a két értelmezés mellett a társadalom szerepét hangsúlyozza a mentális zavarok kialakulásában. Annak, hogy milyen módon tekintünk a depresszióra, melyik a domináns keretezési mód, társadalmi szinten is nagy tétje van. A keretezés ugyanis kijelöli, milyen okoknak tulajdonítjuk a depressziót, így azt is, milyen kezelési, megküzdési módokat tartunk célravezetőknek. Mindez pedig hatással van arra, hogy a depresszióban szenvedő egyének felmerik-e vállalni problémájukat. A depresszió értelmezésének következményei társadalmi és egyéni szinten egyaránt megjelennek, például olyan kérdésekben, minthogy támogassa-e a társadalombiztosítás az antidepresszánsokat vagy a pszichológiai kezelést; hogy az érintett személy hatékony terápiát választ-e; de akár abban is, hogy a depressziós egyénre a társadalom miként tekint. Az alábbiakban a fenti három fő keretezési módot mutatom be röviden.

BIOMEDIKÁLIS KERETEZÉS

A szakemberek előtt még nem ismert, hogy pontosan milyen biológiai tényezők okozzák depressziót, csupán az, milyen különbségek figyelhetők meg a depresszióval érintettek testi folyamataiban az azzal nem érintettekhez képest. Egyes szerzők szerint (Busfield 2000, Comer 2013) ez a keretezés a jelenlegi uralkodó diskurzus a tudományos életben, főként amiatt, hogy a természettudományok jellegükből fakadóan objektíven mérhető eredményekkel tudják alátámasztani a hipotéziseiket.

A területen megjelenő különböző nézetek a genetika, illetve a biokémiai folyamatok szerepét hangsúlyozzák. A genetika szerepét támasztják alá azok a kutatások, amelyek szülők és gyerekek depresszióját vizsgálták olyan esetekben is, amikor a gyermek nem a vér szerinti szüleivel élt (Kamali – McInnis 2011). E szerint az iskola szerint tehát a depresszió, az arra való hajlam öröklődhet. A biokémiai nézőponthoz tartozik egyfelől az a megfigyelés, hogy két neurotranszmitter, a noradrenalin és a szerotonin aktivitása erősen kötődik a depresszióhoz, másfelől a depressziós

⁴ A használt keretezés fogalom részben eltér Lakoff (2004., o. xv) keretezés fogalmától, mely szerint a keretek azok a „mentális struktúrák, amelyek formálják azt, miként látjuk a világot”. Az általam használt keretezési fogalom ennél szűkebb, arra vonatkozik, a vizsgált jelenséget mely diszciplína szerint értelmezzük.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

személyekben megfigyelt magas kortizol szint (Gao – Bao 2011), amely a stressz kiváltásáért felelős hormon. A képpalkotó rendszerek segítségével napjainkban az is ismert, hogy az agy különböző részeiben (például az amygdalában vagy a prefrontális kéregben (Ressler – Mayberg 2007)) eltérő az aktivitás szintje a depresszióban szenvedő és az abban nem szenvedő egyéneknél. Azonban sem a biokémiai faktorok esetén, sem az agyi területek aktivitása esetén nem sikerült (még) kimutatni az okság irányát, azaz nem tudhatjuk, az agyi, biokémiai elváltozások okozzák-e a depressziót, vagy hogy éppen a depresszió hatására történnek változások ezekben.

A biomedikális keretezéshez kapcsolódó kezelési módok a megváltozott biokémiai folyamatokra hatnak, így jellemzően antidepresszánsok szedésével kezelik a depressziót (Goldstein et al. 2011).

PSZICHOLÓGIAI KERETEZÉS

A pszichológiai gyakorlatban a depresszió lelki eredetű problémaként definiált, azonban ezen belül is meg kell különböztetnünk többféle irányzatot: a pszichoanalitikus, a kognitív és a behaviorista nézetet. A pszichoanalízis irányzata szerint a depresszió a gyászhoz hasonló folyamat, a gyászhoz képest hosszabb távon fennálló, tényleges vagy szimbolikus veszteség váltja ki (Freud 1985, Sik 2018).

A behaviorista nézet szerint a depressziót a jutalmazás, megerősítés csökkenése, a negatív megerősítés növekedése idézi elő (Lewinsohn 1985). Súlyosbítja a helyzetet, hogy a depressziós személy társadalmi jutalmainak csökkenése a konstruktív viselkedést is csökkenti, amely a további jutalmak elmaradását okozhatja, így eredményezve depressziós spirált (Martell et al. 2013).

A kognitív nézet két jellemző, depressziót célzó magyarázata a negatív gondolkodás és a tanult tehetetlenség elmélete. A negatív gondolkodás nézete szerint a depresszió fő okozója az egyént körülvevő világ negatív interpretációja (Beck 1979). A tanult tehetetlenség (Seligman 1972) pedig az a jelenség, amikor az egyén korábbi negatív tapasztalatai alapján úgy gondolja, nem ő irányítja az életét, a jutalmakra és büntetésekre nincs hatása, így egy idő után nem is próbál javítani az életkörülményein (Comer 2013).

A depresszió kezelése a pszichológiában terápiával történik, de az egyes iskolák ebben is különböző terápiás módszereket alkalmaznak. A korábban domináló pszichoanalízisben a depressziót kiváltó veszteséget hívták elő és tudatosították a személyekben, de a kutatások nem minden esetben támasztották alá ennek a módszernek a hatékonyságát (Comer 2013). A behaviorista terápia arra helyezi a hangsúlyt, hogy az egyén olyan tevékenységeket végezzen, amelyek örömet okoznak számára, és amelyek pozitív megerősítést eredményeznek. A kognitív terápiában pedig tevékenyebbé teszik az egyént, napi célokat, feladatokat tűznek ki elé, majd az au-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

tomatikus negatív gondolatok automatikusságára és a gondolatok negatív mivoltára ébresztik rá, végül pedig az elsődleges, maladaptív attitűdjeit változtatják meg.

SZOCIOLÓGIAI VAGY SZOCIOKULTURÁLIS KERETEZÉS

A szociológiai keretezés az eddig tárgyalt két keretezési módot egészíti ki egyfelől azzal, hogy az egyént nem különálló tényezőként, hanem a társadalom szerves részeként tekinti, amely eredője lehet a depresszióhoz köthető pszichológiai problémáknak is; másfelől pedig abban az értelemben, hogy a biomedikális diskurzushoz képest, ahol az oksági kapcsolat iránya nem ismert, (a pszichológiai keretezéshez hasonlóan) a depresszió kiváltó okait keresi. Akárcsak a másik két értelmezési mód esetén, a szociológia különböző irányzatai is eltérő tényezőket emelnek ki a depresszió vagy általában a mentális zavarok vizsgálatakor.

A multikulturális megközelítés⁵ a depresszió társadalmon belüli eloszlását, szabályszerűségeit vizsgálja, magyarázatként akár olyan egyéni jellemzőket keresve, amelyek a társadalom struktúrájából fakadnak és így kapcsolódnak a depresszióhoz (Comer 2013).

Sik (2018) arra hívja fel a figyelmet a depresszió szociológiai aspektusát vizsgálva, hogy az örömtelenség, amely a depresszió egyik tünete, a társadalomból, annak működéséből fakadhat – például a különböző erőforrások hiányából vagy a társadalmi normák szigorú szabályaiból. Emellett a cselekvésre való képesség leépülése is összefügg a társadalom működésével, például a társadalmi elismerés hiánya vagy a munkanélküliség problémái kapcsán. Átfogó meghatározása (Sik 2018: 8) szerint ez alapján a depresszió tekinthető az eltorzult társadalom, a társadalmi zavarok interiorizálásaként is.

A depresszió szociológiai értelmezésének hermeneutikai vonatkozása is van, amelyben a szociológia szintén a két másik keretezési módot egészíti ki: eszerint a mentális zavarok, így ezen belül a depresszió meghatározása, vagyis a normális és patológikus elhatárolása társadalmi konstrukciók függvénye (Busfield 2000, Rogers – Pilgrim 2005), amelyek térben és időben is változnak.

A depresszió szociológiai értelmezése különösen fontos lehet a megfelelő kezeléshez: ha ugyanis csak medikális vagy terápiás úton történik a kezelés, azonban az egyén társadalmi helyzete nem változik meg (azaz a depressziót kiváltó probléma nem oldódik meg), a kezelés lényegében csak tüneti lesz. Természetesen a társadalom rendszerét, szabályszerűségeit nem lehet egyéni szinten megváltoztatni: ahhoz összehangolt, társadalmi szintű változások, (szociál)politika szükséges, egyéni szinten azonban hatékonyak lehetnek azok a terápiák, amelyek érzékenyek az egyén

⁵ Rogers és Pilgrim (2005) ezt a nézőpontot oksági magyarázatként nevezi, hiszen a diagnózist elfogadva ez a nézőpont az okokat keresi.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

társadalmi helyzetére és környezetére, a terápia során ezen tényezőkre is figyelnek (Comas-Díaz 2011).

Korábbi kutatások a témában

„HAGYOMÁNYOS” MÓDSZEREKET HASZNÁLÓ KUTATÁSOK

Az alábbiakban olyan, hagyományos eszköztárral készült kutatásokat mutatok be, amelyek a depresszió keretezési módjait vizsgálják laikus diskurzusokban. Kokanovic munkatársaival (2012) az érintettek diskurzusát interjúk alapján kutatta. Az interjúk fókuszában a diagnózis felállítása, annak esetleges tagadása és az életvitelre történő hatása állt. Az Ausztráliában 2008 és 2011 között felvett interjúk eredményei azt mutatják, hogy az érintettek (orvosi diagnózissal vagy öndiagnózissal rendelkezők) hol elfogadják, hol kiegészítik a medikális diskurzust, ingáznak a medikális és nem-medikális elemek között. Sokan gondolják például úgy, hogy az antidepresszánsok segíthetnek a tünetek enyhítésében, mindazonáltal nem elegendők a probléma kezeléséhez, hiszen a depresszió komplexebb annál, mintsem, hogy gyógyszerrel orvosolható legyen (amely rímél a szociológiai keretezésben megjelenő kezelés-problematikára). Érdeemes kiemelni, hogy az orvosi diagnosztikai gyakorlatot nem bírálták az interjúalanyok, miközben ennek bírálata a tudományos életben megjelenik (Busfield 2000, Kokanovic et al. 2012). Sőt elmondásuk szerint egyéb segítség híján az orvossal beszélnek meg a depresszióhoz köthető mindennapi, nem egészségügyi problémáikat is az interjúalanyok – akkor is, ha tudják, hogy azok nem az orvos kompetenciájába tartoznak. A tanulmány a szociológiai aspektust is említi, amikor arról ír, hogy világunkban a vidámság és elégedettség a cél, így patológikus viselkedésnek hat, ha valaki elégedetlen vagy lehangolt. Az interjúalanyok közül többen is úgy látták, a lehangoltságuk nem igényel orvosi beavatkozást, nem biológiai eredetű, hanem adott élethelyzetükből adódik (Kokanovic et al. 2012). E tanulmány nagyon jól mutatja, hogy az egyéni gondolkodásban nem feltétlenül van csak egyféle, egyértelmű értelmezési keret, sokkal inkább azok keveredése jellemző.

Az előbbieken bemutatott tanulmányban tehát kutatási célból készített interjúkat elemeztek. A következőkben bemutatott kutatások adatait már elemzésem szempontjából relevánsabb, online elérhető közösségi média hozzászólások képezik, feldolgozási módszerük azonban a jelen tanulmányéhoz képest eltérő, hagyományos, kutatói kvalitatív kódolás. A depresszió nyilvános értelmezését vizsgálja Pan munkatársaival (2018) egy kínai közösségi médium, a Twitterhez hasonló mikroblog, a Sina Weibo egyes hozzászólásainak tartalomelemzésével. Kvalitatív tartalomelemzéssel vizsgáltak 902 hozzászólást egy többszempontú kódrendszer alapján. Az egyik ilyen vizsgált szempont az volt, hogy a depresszió milyen tényezők miatt alakul ki, és milyen lehetőség van a megoldására. A szerzők négy kategóriát



TEMATIKUS TANULMÁNYOK – Digitális szociológia

különítenek el: a biológiai és medikális, a személyes (például önszabályozás), a szülőikkel vagy más közeli hozzátartozókkal való kapcsolathoz kötődő és végül a strukturális szintű okok és megoldások kategóriáját (Pan et al. 2018: 776). E négy kategória lényegében megfeleltethető a kutatásom alapját képező kategóriáknak – bár a kapcsolatokat mind a pszichológiai, mind a szociológiai kategóriába besorolhatjuk. E kategóriákkal kapcsolatban eredményeik azt mutatják, hogy ha strukturális vagy biológiai okok merülnek fel, akkor kisebb a hozzászólásokban megjelenő stigmatizáció, mintha a felhasználók egyéni tényezőket társítanak a depresszióhoz. Kutatásuk során a szerzők nem csak személyek, hanem intézményi fiókok hozzászólásait is elemezték, s összehasonlították, hogy a különböző típusú szereplők jellemzően milyen keretezési módokat használnak. Eredményeik szerint az állami média, a médiaszereplők és (az elsősorban abból tájékozódó) civil felhasználók hajlamosabbak az egyéni felelősséget hangsúlyozni; a tudományos élethez vagy az egészségügyhöz tartozó szereplők a biológiai tényezők fontosságát emelik ki; a strukturális tényezők pedig jellemzően a „civil újságírók” (citizen journalist) vagy különböző társaságok hozzászólásaiban jelennek meg. A korábban ismertetett tanulmányhoz hasonlóan e tanulmányban is megjelenik a biológiai és a nem-biológiai keretezés ellentéte.

A tanulmányban leírt vizsgálatnak azonban vannak korlátai. Csupán 902 hozzászólást elemeztek, holott hároméves intervallumban 2014 és 2016 között vizsgáldták, amikor (az adatgyűjtés leírásából következtethetően) nagyjából 36000 hozzászólás született az oldalon. A kis elemszám lehetővé teszi a kvalitatív kódolást, azonban a rendelkezésre álló adatoknak így csak kis részét elemzik. Ezen kutatás tehát viszonylag kevés adatra épül, amely mintázatok keresésére kevésbé alkalmas és jobban kitett az adatok esetlegességének. Továbbá az alacsony elemszám hátránya, hogy nem feltétlenül találkozunk minden, kutatási kérdésnek megfelelő kategóriával, ha annak számossága alacsony a vizsgált térben.

Ezzel szemben nagyobb mennyiségű szöveg gépi tanulási módszerekkel történő feldolgozása kevésbé hordozza magán a kutatói besorolás hátrányait, hiszen nagyobb mennyiségű adat feldolgozását teszi lehetővé kevésbé szubjektív, reprodukálható módon. Az alábbiakban ilyen, gépi tanulási módszert alkalmazó kutatások eredményeit mutatom be.

GÉPI TANULÁST ALKALMAZÓ KUTATÁSOK

A legtöbb természetesnyelv-feldolgozást használó depresszió témájú kutatás azt célozza, hogy a gép ismerje fel a depresszióval küzdők által írt szövegeket és ezáltal online szövegek alapján azonosítsa a depressziót, sőt, akár a potenciális öngyilkossági szándékot is.

Nguyen munkatársaival (2014) ennek érdekében egy általános, különböző témájú fórumoknak helyet adó oldal speciálisan depresszió, bipoláris zavar, szeparáció,



TEMATIKUS TANULMÁNYOK – Digitális szociológia

öngyilkosság és önsértés, önkárosítás témájú csoportjainak hozzászólásait vetette össze általános (például étel, állatok, divat) témájú csoportjainak hozzászólásaival. Előbbiből több mint 30 000, utóbbiból közel 230 000 hozzászólást gyűjtöttek le. A két csoportba sorolás nyelvi és tematikus jegyek alapján történt. A gépi tanulóval létrehozott tematikus csoportokon (topik modell eredményein) alapuló modell magas, 93 százalékos pontossággal működött. A témacsoportok, amelyek a besorolást segítették, olyan általános és egyértelműen a depresszióhoz sorolható kifejezések mellett, mint *suicide*, *mental health*, *cut*, *depression*, vagy *bipolar* olyan szavakat is tartalmaztak, mint a *meds*, *heart*, *pain*, *away*, *mother*, *father*, vagy *normal*, amelyek a biomedikális keretezésre és a társas kapcsolatokra, így szociológiai keretezés megjelenésére engednek következtetni a vizsgált oldalon zajló párbeszédekben is.

Thaul Lehrman és munkatársai a fórumhozzászólásokban megjelenő káros feszültség, vagyis distressz azonosíthatóságát vizsgálták (2012). Négy kategóriát határoztak meg: magas distressz, alacsony distressz, válaszhozzászólás, és boldog hozzászólás. A kutatás során kényelmi mintát alkalmaztak, a 200 hozzászólást különböző mentális egészséggel foglalkozó fórumokról úgy választották ki, hogy a hozzászólások fele distressz, másik fele pedig egyenlő arányban boldog és válaszhozzászólás legyen. A distressz csoportba tartozó hozzászólásokat ezt követően tovább bontották, elkülönítették az alacsony és a magas distressz szintű hozzászólásokat. A felderítő kutatás során különböző osztályozási módszereket (döntési fákat, Naiv Bayes, és maximum entrópia modelleket) használtak annak érdekében, hogy képet kapjanak a modellek teljesítményéről, amelyeket a későbbiekben nagy adatbázison terveztek alkalmazni. Eredményeik szerint a gépi modellek 20 százalékkal nyújtottak jobb teljesítményt, mint a véletlen allokáció.

O'Dea és munkatársai (2015) öngyilkosságra utaló tweeteket osztályoztak három kategóriába aszerint, hogy mennyire látszik öngyilkossági szándék a tweetek mögött. A három kategória a következő volt: (1) az öngyilkossági szándék erősen feltételezhető („*strongly concerning*”); az (2) öngyilkossági szándék előfordulhat („*possibly concerning*”); és (3) biztonsággal figyelmen kívül hagyható a hozzászólás („*safe to ignore*”). Összesen 14 701 tweetet töltöttek le 2014 februárja és áprilisa között, s egy abból vett 2000 elemű véletlen mintát kódolt be két kódoló a három kategória valamelyikébe. A modellezés során ennek a 2000 elemű bekódolt mintának egy részét biztosították a gép számára „tanulás” céljából és a maradékon tesztelték a tanulás utáni besorolás hatékonyságát. A klasszifikáció során logisztikus regressziót és support vector machine-t alkalmaztak különböző modellspecifikációkkal. Eredményeik szerint elkülöníthető az öngyilkossági fenyegetettség mértéke humán kódolók munkájára épített gépi tanulási modellekkel, bár a modellek teljesítménye még javítandó egy igazán megbízható kategorizáció érdekében. (A modellek 55-76 százalékos átlagos pontosságot mutattak attól függően, milyen algoritmust és milyen specifikációkat alkalmaztak.)



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A GÉPI TANULÁS ELŐNYEI ÉS HÁTRÁNYAI

A gépi tanulással történő elemzés elsődleges előnye a rendelkezésre álló rengeteg adat feldolgozhatósága. Az online elérhető adatok hagyományos kvalitatív szövegelemzéssel történő feldolgozása magán hordozza a kvalitatív eszköztár olyan hátrányait, mint a nehéz reprodukálhatóság vagy a korlátozott általánosíthatóság, amely a relatíve kis elemszámból adódik. A természetesnyelv-feldolgozás alkalmazásának segítségével jelentősen nagyobb mennyiségű szöveget reprodukálható módon lehet feldolgozni, amely humán erőforrással amúgy képtelenség lenne. A rendelkezésre álló teljes adatmennyiség (például egy fórum összes hozzászólása, vagy egy adott kulcsszót tartalmazó összes hozzászólás) így hiánytalanul feldolgozható, az elemzett szövegek egyértelműen körülhatárolhatók ezáltal jobban azonosíthatóak a kutatás megfigyelésének határai. Szintén a gépi feldolgozás előnye az adatok feldolgozásának és így az eredmények jobb reprodukálhatósága és visszakövethetősége, amely a programozott lépéseknek köszönhető.

A gépi tanulással végzett kutatások hátránya a magyar kutatóközösség számára, hogy egyes témákban, akár csak jelen vizsgálat témája, nem áll rendelkezésre elegendő adat ahhoz, hogy a gépi tanuló algoritmusok megfelelően tudjanak működni. Amiatt, hogy a szövegelemzés nyelvspecifikus, nagy mértékben függ az osztályozás sikeressége attól, az adott nyelvre alkalmazható, rendelkezésre álló programok milyen minőségűek, mennyire megbízhatók. Szintén megnehezíti a szociológiai kutatásokban való alkalmazhatóságát, hogy jelentős programozási ismereteket igényel, emiatt kevésbé gyorsan tanulható, mint egy felhasználóbarát elemzőprogram (például az ATLAS.ti vagy az NVivo). Ugyan a gépi tanulás elsődlegesen ott válik a szociológia előnyére, hogy a nagy mennyiségű adat válik általa feldolgozhatóvá, ez egyben hátránya is, ugyanis a szövegek osztályozását rábízuk a gépre, így nem ellenőrizhető minden szöveg esetén, hogy valóban megfelelő osztályba sorolta-e a gép.

Az online elérhető, felhasználók által írt szövegek elemzésének hátránya, hogy széles közönség számára elérhetőek a honlapok, fórumok, és a kutatóknak nincs pontos ismerete arról, kik írják ezeket a szövegeket, tehát nincs egy jól definiálható populáció, emiatt az adatokból nem is vonhatunk le általánosítható következtetéseket, azokat csak a megfigyelt honlapok diskurzusaira (és nem a felhasználókra) tudjuk megtenni.

Témám érzékeny aspektusa szempontjából kiemelendő a szövegek automatikus kategorizálásának egy másik előnye is, miszerint nem embereknek kell feldolgoznia több tízezer, depressziós személyek által írt szöveget. A gépi feldolgozáshoz elengedhetetlen bizonyos mennyiségű szöveg elolvasása, de ezen szám alacsonyan tartásával a lelki teher is kisebb lehet, amely a kutatásban közreműködőket érheti.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Adatok és módszerek

A kutatás során használt, angol nyelvű adatbázist az RC²S² kutatócsoport⁶ keretében végzett kutatás során állítottuk elő. A kutatásban az angol mellett magyar nyelvű szövegek elemzését is célul tűztük ki, magyar nyelven azonban nem volt elérhető megfelelő mennyiségű fórumhozzászólás. A depresszióval foglalkozó online fórumokat a „depression forum”, „depression online” kifejezésekre történő kereséssel találtuk meg. A keresés célja annak modellezése volt, ahogyan egy érintett a gyakorlatban rátalálhat ezekre a fórumokra. A keresés eredményeit a következő elvek alapján szűkítettük: csak olyan fórumok hozzászólásait gyűjtöttük le, amelyek az utóbbi három évben is aktívak voltak és amelyek hozzászólásait regisztráció nélkül lehet olvasni. Az így előálló 19 fórum között szerepelnek szűken a depresszióval foglalkozó honlapok fórumai,⁷ általánosan (mentális) egészséggel foglalkozó honlapok fórumai,⁸ illetve a felhasználók által akármilyen témában indított fórumokat tartalmazó honlap depresszió témájú fórumai is.⁹ Ez utóbbiak annak érdekében kerültek az adatbázisba, hogy ne csak tematikus, orvosi vagy pszichológiai oldalak fórumai, hanem általános fórum is szerepeljen, így lehetőség szerint csökkentve a honlapok tematikájából eredő torzító hatást.

A honlapokról ezt követően legyűjtöttük a 2016. februárja és 2019. februárja közötti összes hozzászólást a SentiOne webes szöveganalitikai platform adatgyűjtési szolgáltatásával. Ezt követően a nyers adatbázison az alábbi, kétlépcsős szűrést alkalmaztuk a kutatócsoportban hozott döntés alapján:

(1) Csak azokat a beszédflowamokat¹⁰ (úgynevezett threadeket) választottuk ki, amelyekben előfordul a „depression” vagy „depressed” szó.

(2) A kiválasztott beszédflowamokból pedig csak azokat a hozzászólásokat tartottuk meg, amelynél a beszédflowam címében, a beszédflowam linkjében vagy magában a hozzászólásban szerepel a depresszió szó, vagy annak valamilyen szinonimája.¹¹

⁶ A tanulmány alapját képező, kiinduló kutatást Németh Renátával, Sik Domonkossal és Katona Eszterrel hajtottuk végre. Az itt bemutatott eredmények a kutatás annotált adatbázisának általam történő újbóli előfeldolgozására és a szintén általam készített új modellekre alapulnak.

⁷ például depression-understood.com

⁸ például mentalhealth.com, healtunlocked.com

⁹ például reddit.com

¹⁰ Beszédflowamok alatt azokat a fórumtopikokat értem, amelyeket egy-egy felhasználó, bizonyos témában, kérdésben indít el.

¹¹ A szinonimák: depression, depressed, bummer, desolation, desperation, moody, upset, gloom, hopelessness, depressant, melancholia, sorrow, unhappiness, feeling blue, depressive, depressive disorder, unipolar depression, bipolar, bipolar depression, major depression, mdd, persistent depressive disorder, pdd, cyclothymia, mood disorder, adjustment disorder, chronic fatigue syndrome, cfs, premenstrual dysphoric disorder. A szinonimák kiválasztásánál ügyeltünk arra, hogy amennyire lehetséges, a találatok keretezési módjait ne befolyásolják a kiválasztott szavak, így kőznapi kifejezéseket és szakszavakat is használtunk. Az esetleges torzításokat enyhítik a beszédflowam címében vagy a linkjében előforduló szinonimák miatt bekerülő teljes beszélgetések.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

A szűrést több tényező is indokolta. Technikai ok volt, hogy bizonyos honlapokról (például reddit.com) a honlapstruktúra miatt nem lehetett csak a depresszióval foglalkozó fórumokat leválogatni az adatgyűjtés során, így a teljes fórum hozzászólásai-ból a letöltés után kellett kiválogatni a kutatás szempontjából relevánsakat. Továbbá a legújított hozzászólások olvasása során azt tapasztaltuk, hogy egy általánosan depresszióval foglalkozó fórumon belül sem csak ilyen témájú beszédanyagok fordulnak elő (például gyakori a fórumszabályokkal kapcsolatos beszédanyag, de előfordultak filmekkel, zenéssel kapcsolatos beszélgetések is). Emellett azt találtuk, hogy a depresszióval foglalkozó beszédanyagokon belüli hozzászólások is szerteágazóak lehetnek, így nagyon magas lehet a téma szempontjából irreleváns hozzászólások aránya. Ezen megfigyelésünk indukálta a szinonimák előfordulásának kritériumát. Ha azonban csak olyan beszélgetésanyagokat elemeztünk volna, amelyek címében előfordul a depresszió szó, kizártunk volna olyan beszélgetésanyagokat, amelyek például gyógyszerek vagy kezelések köré szerveződnek, de magát a depresszió szót nem tartalmazzák nevükben. Éppen ezért azon hozzászólásokat is beemeltük az elemzésbe, amelyek bár olyan beszélgetésanyagokban fordultak elő, amelyeknek címében nem szerepelt a „depresszió” szó, azonban a hozzászólásban magában igen.

Az adatbázis tisztítása során elsőként az adatgyűjtés fázisában keletkező hibákat kezeltük: megszüntettük a duplikátumokat és olyan esetekben, amikor a válaszként érkező hozzászólásban szerepelt az eredeti hozzászólás szövege is, ez utóbbit kitöröltük. Eltávolítottuk a 20-nál kevesebb szót tartalmazó hozzászólásokat, hiszen tapasztalataink alapján azok kevésbé hordoznak elemezhető információt: inkább csak üdvözlések, megerősítések, így a depresszió értelmezése, keretezése szempontjából kevésbé nyújtanak információt, ellenben nagy mértékben megnövelik az adatbázis méretét, amely hosszabb futásidőt eredményez. A hozzászólásokból ezt követően el kellett távolítani a különböző linkeket, e-mail címeket is, amelyek szintén irrelevánsak a kutatás tárgyát illetően, ám a későbbi elemzési fázisban hátráltatták volna a modellezési eljárást.¹² A fentiek alapján leválogatott és kitisztított adatbázisban, vagyis a végleges szövegkorpuszban 67 857 hozzászólás szerepelt.

A gépi tanulás során a számítógép számára elsőként egy olyan adatbázist kellett biztosítani, amin „megtanulhatja” az általunk elvárt kategorizációt. Ehhez egy, a teljes adatbázishoz képest kisebb adatmennyiségen kézi, humán erővel be kell sorolnunk a hozzászólásokat és utána ezen kisebb adatbázis egy részét átadni a számítógépnek, hogy a besorolt hozzászólásokban mintázatokat keressen, amikről a kategóriák felismerhetővé válnak. (Ezen „tanult” mintázatok alapján fogja a gép elvégezni a teljes, nagy adatbázisban meglévő hozzászólások kategorizációját később.) A kisebb adatbázison végrehajtott, emberi erőforrással történő besorolást nevezzük annotálásnak.

¹² Annak ellenére, hogy egy link hordozhat információt arról, a hozzászólás írója miként értelmezi a depressziót, ezen linkek tartalmának felkeresése túlmutat a jelenleg használt modellek működésén.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Az annotálás folyamata

Ahhoz tehát, hogy a számítógép képes legyen felismerni azokat a mintázatokat, amelyek az egyes hozzászólásokat adott csoportba tartozóvá teszik, a kategóriákat egy kisebb adatbázison először emberi erővel szükséges megadni. Ezt a feladatot gyakran különböző online platformokon toborzott munkavállalók végzik (például Amazon Mechanical Turk, Figure Eight), azonban a hozzászólások osztályozásához szükséges elméleti ismeretek miatt kutatásunkban szociológiai tudással rendelkező, (egy fő kivételével) mesterszakos hallgatók végezték a hozzászólások osztályozását. A tíz egyetemi hallgató 4500, az adatbázisból egyszerű véletlen módszerrel kiválasztott hozzászólást osztályozott. (Az egyszerű véletlen kiválasztás nagy valószínűséggel adja a teljes adatbázis egy reprezentatív mintáját.) A besoroláshoz ötféle címke állt rendelkezésre: a három, már említett keretezés mellett bevezettük a (1) transzcendentális kategóriát,¹³ azt, hogy (2) a hozzászólás nem besorolható egyik keretezési módba sem, de a depresszióról szól, továbbá azt, hogy (3) a hozzászólás nem a depresszióról szól. Mivel nem feltételezhettük, hogy minden hozzászólás csak egy értelmezési módban íródott, ezért lehetősége volt az annotátoroknak a hozzászóláshoz második címkét (kategóriát) is adni, ha úgy látták, második keretezési mód is megjelenik a hozzászólásban.

Az annotálást követően egységesítettük a hozzászólásokhoz tartozó címkéket (vagyis a kategóriát jelölő értéket). Ezt úgy tettük meg, hogyha csak egy olyan címke volt, amely mindkét annotátornál előfordult, akkor az vált az elsődleges címkévé. Ha csak az egyik annotátor adott a három főkategóriába tartozó címkét, akkor az vált a címkévé. Ha mind az első, mind a második címke megegyezett, akkor azok, helyzetüknek megfelelően kerültek az első és a második címke helyére. Ha a két kategória sorrendje nem volt azonos, akkor a szakértői kódolás alapján történt a besorolás, akárcsak abban az esetben, ha csak eltérő kategóriába tartozó címkék szerepeltek.

Az egységesítés után előálló kategóriák eloszlásából (1. táblázat) látható, hogy az esetek túlnyomó többségében nem szerepelt második címke és az általunk vizsgált hozzászólásokban a pszichológiai keretezési mód a leggyakoribb első címke (32,7%), majd pedig a biomedikális (28,6%), a szociológiai keretezési mód pedig viszonylag ritka első kategóriaként (8,6%). Azokban az esetekben, ahol két címkét is kaptak a hozzászólások, elsősorban a pszichológiai első címkéhez tartozik szociológiai második címke, illetve a biomedikális első címkéhez pszichológiai második. Tehát az adatbázisban leghangsúlyosabb pszichológiai értelmezés mellett feltűnik a szociológiai keretezési mód is, ahogyan az elsődlegesen biomedikális nézőpont mellett is megjelenik a pszichológiai az általunk vizsgált hozzászólásokban. A modellezés során csak az első címkét vettem figyelembe, mivel a két címkével rendelkező

¹³ A transzcendentális kategóriába tartoztak olyan hozzászólások, amelyek a depresszió okaként vagy kezelési módjaként is természetfeletti, vallási vagy esetleg áltudományos tényezőket említettek.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

hozzászólások esetén is az első címke a hangsúlyos, annak megfelelősége megbízhatóbb.

1. táblázat. Az első és második címkék kategóriánkénti együttes előfordulása

Első címke	Második címke					
	Nincs második címke	Bio-medikális	Pszichológiai	Szociológiai	Egyéb	Összesen
Biomedikális	858	0	327	88	11	1284
Pszichológiai	915	196	0	333	23	1467
Szociológiai	209	25	150	0	2	386
Egyéb	172	8	18	6	0	204
Nem depr.	1144	0	0	0	0	1144
Összesen	3298	229	495	427	36	4485

Forrás: Saját szerkesztés

Az adatbázis előkészítése

A szövegeket bizonyos előfeldolgozási lépéseknek kell alávetni annak érdekében, hogy egységesebb, a számítógép által könnyebben értelmezhető szöveget kapjunk, majd ezeket speciális formátumú adatbázisba kell rendezni annak érdekében, hogy az algoritmusok fel tudják azt dolgozni. Az előfeldolgozásba tartozik például a számok és a betűformátumok kezelése, a különböző módon ragozott szavak azonos alakra hozása (lemmatizálás), a több szóból álló kifejezések (szignifikáns bigramok) együtt kezelése és a valódi jelentéssel nem bíró elsősorban mondatösszetartó szerepű szavak eltávolítása (stopszavazás).

Ahhoz pedig, hogy a szöveget algoritmikus, számítógép általi feldolgozásnak vessük alá, számokká kell alakítani azt. Egy ilyen lehetséges folyamat a vektorizálás, amelynek során egy dokumentumon belül a mondat szerkezeti sajátosságokat nem vesszük figyelembe, csak azt, hogy milyen szavak szerepelnek az adott szövegben. Ennek során egy hatalmas, dokumentum-szó mátrix keletkezik, amelynek minden sora egy dokumentumot jelöl, minden oszlopa pedig egy adott szót vagy más dokumentum-jellemzőt (például azt, milyen honlapról származik a hozzászólás, vagy milyen hosszú). A mátrix cellájában pedig olyan számok szerepelnek, amelyek a szavak adott dokumentumbeli fontosságát jelzik.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Felügyelt osztályozó algoritmusok

Ahogy arra már korábban utaltam, egy ilyen, általam is alkalmazott felügyelt szöveg-osztályozási feladat esetén a modellnek először egy emberek által felcímkezett kisebb adatbázisból kiindulva meg kell „tanulnia”, fel kell ismernie azokat a hozzászólásokban lévő mintázatokat, amelyek az egyes kategóriákhoz tartoznak, s amelyek alapján meg lehet különböztetni azokat. Ezen kisebb adatbázison tanulva, a megismert mintázatok segítségével a gépnek új, a kisebb, tanuló adatbázisban nem szereplő, „még nem látott” szövegeket az előre definiált osztályok valamelyikébe kell sorolnia. Így téve lehetővé nagymennyiségű adatok automatizált osztályba sorolását. Az ilyen algoritmusok hétköznapi használata például, ha egy ügyfélszolgálatnak írt levelet annak szövege alapján az algoritmus az adott területtel foglalkozó osztályra irányít. A szövegosztályozó algoritmusok tehát a szöveget alkotó, előfeldolgozás utáni szavak, illetve egyéb, kutatók által megadott jellemzőkből kiindulva döntenek el, hogy az egyes dokumentumok melyik osztályba sorolhatók.

Az osztályozó algoritmusok új adatok besorolásában nyújtott teljesítményének megítéléséhez szükségünk van legalább még egy olyan, emberek által már felcímkezett adatbázisra, amelynek adatai (tehát az általa tartalmazott hozzászólások) elkülönülnek azoktól az adatoktól, amelyeken „tanult” a modell, tehát amelyben már megismerte a hozzászólások kategóriáját. Ennek érdekében az emberek által már bekódolt, ismert kategóriákkal rendelkező kisebb adatbázist véletlenszerűen két részre: tanító és teszt adatokra bontjuk. Előbbiekben tanítjuk a modellt, utóbbiakon pedig mérjük az alkalmazott modell osztályozási teljesítményét. A tanító adatokon tehát mind a kutatók, mind a gép ismeri a helyes besorolást, a teszt adatoknál azonban csak a kutatók ismerik a helyes besorolást, és nem bocsátják azt a gép rendelkezésére. A teszt adatokon éppen azt vizsgálják ugyanis, hogy a gép mennyire tudja eltalálni a helyes kategóriákat. A tanulási folyamat során olyan modell kifejlesztése a célunk, amely nem csak a tanító adatokra illeszkedik jól, de a teszt adatokon is jó teljesítményt nyújt, hiszen az jelzi a modell általánosíthatóságát.

Annak érdekében, hogy a nem ismert adatok besorolását a legjobb modellen hajtsuk végre, a különböző modellspecifikációkat validálni szükséges. A kutatás során ennek érdekében keresztvalidálást alkalmaztam, mely során a tanító adatokat öt egyenlő, diszjunkt részre bontottam, ahol négy rész szolgált tanító adatbázisként, egy rész pedig tesztadatként. A tanítást és tesztelést ebből következően ötször végeztem el, mindig egy egységet hagyva tesztalumnak. Ekkor a modell teljesítménye az ötszöri futtatás átlagos eredményeként írható le.

Alkalmazott modell: Logisztikus regresszió

A társadalomkutatásban jól ismert logisztikus regressziós modell a szöveganalitikában is használatos a jelen esetben is fennálló klasszifikációs problémák esetén. Szö-



TEMATIKUS TANULMÁNYOK – Digitális szociológia

vegelemzésben osztályozási problémára alkalmazott logisztikus regressziónál az egyes szavak vagy egyéb, osztályozáshoz használt tényezők értékei veszik fel a „független változók” szerepét, míg a kategóriába tartozás lesz a „függő változó”.

Logisztikus regresszió során a becslés tehát arra vonatkozik, hogy az adott dokumentum melyik kategóriába tartozik a benne szereplő szavak és hozzá tartozó egyéb információk alapján. Akárcsak a szociológiában elterjedt használata esetén, szövegelemzésnél is megkülönböztetünk kétféle modellt attól függően, hogy csak két vagy több osztályba kell az eseteket besorolni. Bináris logisztikus regresszió esetén csak két kategória van, tehát azt kell eldönteni, adott kategóriába tartozik-e a dokumentum vagy sem. Multinomiális logisztikus regresszió esetén, amikor a kimeneti változónk nem csak két értéket vehet fel, egy referenciakategóriához viszonyítva mindegyik kategóriára elvégezzük a számításokat, majd abba a kategóriába soroljuk az esetet, amelyik kategóriának legnagyobb a valószínűsége (a referenciakategóriához képest).

A logisztikus regresszió előnye, hogy a szavakhoz és egyéb jellemzőkhöz tartozó együttthatók segítik a modellek és a kategóriák interpretálhatóságát, hiszen az értékeken keresztül megérthetjük, hogy mely szavak (vagy egyéb, a modellbe beépített jellemzők) befolyásolták jelentősen az adott kategóriába sorolást (magas pozitív együtttható) és mely szavak előfordulása hatott az adott csoportba sorolás ellen (nagy negatív együtttható). Így tehát „beleláthatunk” a modell működésébe, megkapjuk azokat a szavakat vagy egyéb használt jellemzőket, amelyek a besorolás során fontosnak bizonyultak.

Fontos kiemelni, hogy a szövegelemzésben klasszifikációs problémáknál használt logisztikus regresszió esetén a modell jóságának megítélését nem a társadalomtudományból ismert magyarázómodellek esetén használatos mérőszámokkal (például AIC, vagy pszeudo R^2) mérik, hanem a tesztadatokra a modell által becsült és a valós kategória közötti eltéréssel. Tehát az alapján, hogy a tesztadatokon a modell mennyire találta el a tényleges, helyes kategóriát. Az alábbiakban bemutatom ezen mérőszámokat, mivel eredményeim elemzése során is részben ezekre fogok hivatkozni. Ilyen mérőszám az *accuracy* (pontosság), a *precision* (precizitás)¹⁴, a *recall* (visszaidézés) vagy az F1-score. A pontosság a jól besorolt esetek arányát mutatja az összes eseten belül, míg a precizitás osztályonként adja meg azon esetek arányát, amelyek az osztályba sorolt esetek közül valóban jó helyre kerültek. A visszaidézés számításánál szintén az egyes kategóriáknál számítjuk ki, hogy a kategóriába tartozó esetek mekkora arányát sorolta be valóban abba az osztályba a modell. Az F1-score a precizitás és visszaidézés harmonikus átlagaként számolandó.

¹⁴ A magyar szakirodalomban a *precision*t jellemzően pontosságnak fordítják (Sebők et al. 2016, Tikk 2007), én azonban az *accuracy* szótól való megkülönböztetés miatt a precizitás fogalmát használom (ld még: (Tan et al. 2012). Ez utóbbinak nincs bevett magyar fordítása, hol hitelességnek (Sebők et al. 2016), hol szabatoságnak (Tikk 2007) fordítják.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Eredmények

KÜLÖNBÖZŐ MODELLSPECIFIKÁCIÓK HATÁSA AZ OSZTÁLYOZÁSRA

A modellezés során különböző modellspecifikációkat alkalmaztam annak érdekében, hogy megtaláljam a legalkalmasabb modellt, továbbá azért, hogy a modellek robusztusságát is teszteljem. A különböző specifikációk eredményeit a 2. táblázat tartalmazza. A fentiekben bemutatott mérőszámok mellett minden modell esetén közöltem a Cohen's kappa értékét is, amely azt mutatja meg, a tökéletes besorolás és a véletlen besorolás különbségének mekkora részét képezi az, amennyivel a modell jobban teljesített a véletlenszerű besoroláshoz képest. A mérőszám értéke 1, ha a modell tökéletesen sorolja be az eseteket és 0, ha a véletlenszerű egyezésnél nem jobb a modell besorolása. A köztes értékek interpretálásához a Landis és Koch (1977) által leírt értelmezés szolgál ökölszabályként.¹⁵

Kezdetben a legalapvetőbb modellt használtam, amelyben csak a hozzászólások szövege szerepel egyszer stopszavazva (lásd az előfeldolgozási lépéseknél), másszor pedig stopszavazás nélkül. A két modell között minimális különbség van a stopszavazott szöveg javára (1-2. modell). Ez a különbség minden egyéb modellspecifikációnál is megfigyelhető volt, így a továbbiakban csak a stopszavazott modell eredményeit közlöm.

Mivel a modell besorolásának jósága csak mérsékelten volt jobb, mintha véletlenszerűen soroltuk volna be a hozzászólásokat (a Cohen-kappa értéke 0,45), a hozzászólásokban szereplő szavakon túl olyan változókat szándékoztam bevonni a modellbe, amelyek a szöveget kiegészítő információt hordoznak, így segíthetik a kategóriákba sorolás döntési folyamatait. Az első információ, amit beépítettem a modellbe, az a hozzászólást tartalmazó beszélgetésfolyam címe volt (3. modell). Ezt először szöveggként (bináris vektorizálással) építettem be a modellbe – így kívántam megragadni azt, hogy a kategóriába sorolásánál a beszélgetésfolyam témáját is figyelembe vegye a modell. Az eredmények ellentmondtak várakozásaimnak, ugyanis a modell rosszabb teljesítményt mutatott, mint a pusztán szövegeket tartalmazó modell. Ha azonban a beszélgetésfolyamoknak nem a címét, hanem csak azt emeltem be a modellbe, hogy melyik beszélgetésfolyamhoz tartozik a hozzászólás (4. modell), számmal jelölve az egyes beszélgetésfolyamokat, akkor a csak szöveget tartalmazó (1.) modellhez hasonló, bár szintén nem jobb eredmények tapasztalhatók.

A korábbiakhoz képest jobb besorolási teljesítmény tapasztalható azonban, ha a modell az osztályozás során figyelembe veszi azt, hogy az adott hozzászólás beszélgetésfolyama pszichológiai, egészségügyi vagy általános tematikájú fórumról származik (5. modell).

¹⁵ Eszerint 0,00-0,20 kis mértékű a javulás; 0,21-0,40-ig elfogadható; 0,41-0,60-ig mérsékelt; 0,61-től 0,80-ig jelentős és 0,81-től 1,00-ig közel tökéletes az osztályozás (Landis – Koch 1977: 165).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Nem csak a hozzászólások lelőhelyére (fórumára), hanem a hozzászólások tartalmára vonatkozó információkat is építettem a modellbe, nevezetesen azt, előfordul-e gyógyszernev a hozzászólásban. Ez az információ elsősorban a biomedikális kategória felismerését szolgálja, hiszen a gyógyszerekről zajló párbeszéd a biomedikális nézőpontot jelzi. Beépítésével a modell teljesítménye a csak szöveget tartalmazó modellhez képest néhány ezreddel jobb eredményeket mutat minden mérőszámban (6. modell).

A fent írt jellemzők kombinációjával hasonló teljesítményt nyújt a modell (7–8. modell). A gyógyszernev és beszélgetésfolyam száma, és ezek mellett a fórum csoportját tartalmazó modell is 61 százalékos pontosság és 0,45-ös Cohen-kappa értéket mutatott.

2. táblázat. Különböző modellspecifikációk eredményei

	Modell	Accuracy/ Pontosság	Precision/ Precizitás	F1-score	Cohen- kappa
1.	Stopszavazott szöveg	0,608	0,613	0,579	0,445
2.	Stopszavazás nélküli szöveg	0,605	0,611	0,575	0,441
3.	Stopszavazott szöveg, és a beszélgetésfolyam neve	0,559	0,548	0,539	0,381
4.	Stopszavazott szöveg, és a beszélgetésfolyam száma	0,605	0,607	0,576	0,441
5.	Stopszavazott szöveg, és a fórum csoportja	0,609	0,611	0,581	0,447
6.	Stopszavazott szöveg és előfordul-e gyógyszernev a szövegben	0,609	0,617	0,582	0,448
7.	Stopszavazott szöveg, és előfordul-e gyógyszernev a szövegben és beszélgetésfolyam száma	0,610	0,610	0,582	0,449
8.	Stopszavazott szöveg, és fórum csoportja, beszélgetésfolyam száma, előfordul-e gyógyszernev a szövegben	0,609	0,615	0,582	0,448

Forrás: Saját szerkesztés

A fenti eredmények alapján látható, hogy a gyógyszernev előfordulásának és a fórum csoportjának meghatározása minimálisan javította csak a modell teljesítményt a pusztán a szöveget tartalmazó modellhez képest, de egy kivételével mindegyik modell 0,61-es pontosságot és 0,45-ös kappa értéket mutatott.

A csak stopszavazott, alapvető modell tévesztési mátrixát (*confusion matrix*) vizsgálva (lásd 3. táblázat) látható, hogy a rossz besorolás elsődleges oka, hogy a modell nem ismeri fel sem a szociológiai, sem az egyéb kategóriákat, ezen belül is



TEMATIKUS TANULMÁNYOK – Digitális szociológia

a szociológiai kategóriát legnagyobb arányban a pszichológiai kategóriába sorolja. Nagyobb arányú tévesztés látható továbbá a nem depressziós kategória pszichológiába történő besorolásban, illetve az egyéb kategóriába tartozó hozzászólások közel negyedét nem depressziósnak címkézi a modell.

3. táblázat. Tévesztési mátrix a legalapvetőbb modellre

Becsült	Nem depr.	Biomedikális	Pszichológiai	Szociológiai	Egyéb	Összesen
Tényleges						
Nem depr.	126	27	66	1	1	221
	46%	10%	24%	0%	0%	
Bio-medikális	26	194	41	0	0	261
	8%	60%	13%	0%	0%	
Pszichológiai	50	30	224	2	0	306
	13%	8%	60%	1%	0%	
Szociológiai	20	3	36	5	0	64
	23%	3%	42%	6%	0%	
Egyéb	14	4	22	0	5	45
	23%	6%	35%	0%	8%	
Összesen	236	258	389	8	6	897

Forrás: Saját szerkesztés

Megjegyzés: A mátrix a tényleges és a modell által becsült kategóriák eloszlását mutatja be. A cellákban található százalék a tényleges kategóriákon belüli, vagyis sorszázalék

KATEGÓRIÁK CSÖKKENTÉSÉNEK HATÁSA AZ OSZTÁLYOZÁSRA

Már az annotálás során felmerült, hogy a pszichológiai és szociológiai keretezési mód nehezen megkülönböztethető, a társadalom és egyén közti határok a hozzászólásokban gyakran elmosódnak. Az annotált adatokban is kis számú a szociológiai kategóriába besorolt eset, amely eredményeként, mivel csak kevés eseten tanulhatja meg a számítógép a kategóriához tartozó mintázatokat, nehezebb feladat a szociológiai szövegek gyakori jellemzőit megtalálnia. Korábbi eredményeink alapján az a felvetés is megfogalmazódott kutatótársaimban (Németh et al. 2020), hogy a szociológiai keretezés kiforratlansága, intézményrendszerének hiánya nem teszi lehetővé, hogy jól felismerhető fogalomkészlet tartozzon hozzá.

Ezen hipotézis alapján megfigyeltem az osztályozó modell teljesítményét akkor is, amikor a szociológiai és a depresszió témájú, de nem besorolható (egyéb) hozzászólásokat egy csoporttá vontam össze. Vagyis a szakirodalomban megfigyelhető



TEMATIKUS TANULMÁNYOK – Digitális szociológia

biomedikális – pszichológiai – szociológiai bontás helyett elsősorban a biomedikális és a pszichológiai keretezési módot kellett elkülönítenie az osztályozónak. Összesen tehát négy kategória közül kellett a modellnek döntenie: nem depresszióról szól a hozzászólás, biomedikális keretezés, pszichológiai keretezés jelenik meg benne vagy pedig depresszió témájú, de egyéb keretezésű a szöveg. Ezt a kategorizációt már csak a korábban jobb teljesítményt nyújtó modelleken teszteltem. Az eredményeket a 4. táblázat tartalmazza, melyből látható, hogy a fentebb ismertetett modellekhez képest mindegyik új modell jobb teljesítményű, egymáshoz képest azonban továbbra is minimális eltéréseket találunk csak a modellek teljesítményében.

4. táblázat. A különböző modellek specifikációi a szociológiai és egyéb kategóriákat összevontan tartalmazó osztályokra

	Modell	Accuracy/ Pontosság	Precision/ Precizitás	F1-score	Cohen-kappa
1.	Stopszavazott szöveg	0,613	0,617	0,593	0,454
2.	Stopszavazott szöveg, és előfordul-e gyógyszernév a szövegben és beszélgetés- folyam száma	0,616	0,621	0,596	0,458
3.	Stopszavazott szöveg, és fórum csoportja, beszélgetés- folyam száma, előfordul-e gyógyszernév a szövegben	0,613	0,618	0,595	0,455

Forrás: Saját szerkesztés

Mivel a biomedikális kategória a korábbi eredmények alapján jól felismerhető és az ahhoz tartozó szókészlet is specifikus (például gyógyszernevek vagy különböző szakszók), tovább egyszerűsítettem a kategorizációs problémát három kategóriára: nem depresszió témájú hozzászólás, biomedikális keretezésű és nem biomedikális keretezésű. Ekkor a modell teljesítménye számottevően javul (5. táblázat): mindegyik modell 69 százalékos pontosságot és precizitást mutat. A Cohen-kappa értékek is 5 századot javultak. Mindazonáltal a különböző jellemzőket tartalmazó modellek között itt sem figyelhető meg nagy eltérés.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

5. táblázat. A különböző modellek specifikációi a nem depressziós, biomedikális és nem biomedikális osztályokra

	Modell	Accuracy/ Pontosság	Precision/ Precizitás	F1-score	Cohen-kappa
1.	Stopszavazott szöveg	0,690	0,688	0,679	0,500
2.	Stopszavazott szöveg, és előfordul-e gyógyszernev a szövegben és beszélgetés-folyam száma	0,690	0,689	0,679	0,499
3.	Stopszavazott szöveg, és fórum csoportja, beszélgetés-folyam száma, előfordul-e gyógyszernev a szövegben	0,692	0,692	0,681	0,503

Forrás: Saját szerkesztés

Ahogy korábban is említettem, a logisztikus regresszió „független változóiként” szereplő szavak és egyéb tényezők együttthatói megmutatják, hogy melyek azok a szavak és tényezők, amelyek jelentős szerepet játszottak a kategorizáció során. A magas pozitív együttthatók azokat a szavakat jellemzik, amelyek jelenléte egy hozzászólást az adott kategóriába tartozóvá tesz, a nagy negatív értékek pedig éppen a kategóriába sorolás ellen szólnak. Az 5. táblázatban található utolsó, mindhárom tényezőt tartalmazó modellből a *nem-depressziós* kategória esetén azokat a szavakat érdemes megfigyelni, amelyek a kategória *ellen* szólnak, hiszen ezek azok, amelyekkel a modell igazából a depresszió témájú hozzászólásokat ismeri fel. A nem-depressziós kategóriánál a *depression*, *anxiety*, *depressed* és *help* szavakat találjuk legnagyobb negatív (tehát a kategória ellen szóló) együttthatóval. A biomedikális kategóriához kerülnek jellemzően azok a hozzászólások, amelyek a *medication*, *med*, *antidepressant*, *doctor*, *drug* szavakat tartalmazzák, s a kategória ellen szólnak a *friend*, *want*, *love*, *therapist*, *like*, *write* kifejezések. A nem biomedikális osztálynál magas együttthatókat találunk a *depression*, *therapist*, *psychologist*, *anxiety*, *thought* és *step* szavak esetén, míg alacsonyokat az egészségügyi kifejezéseknél, mint a *med*, *medication*, *doctor*, *antidepressant*, *drug* és *sleep*. A fontos tényezőknél tehát jól elkülönülnek a fizikai és a lelki egészséghez kötődő szavak és kifejezések.

Ha megfigyeljük ezen, legjobb modell tévesztési mátrixát (lásd 6. táblázat), akkor látható, hogy elsősorban a nem depressziós kategóriában figyelhető meg a bizonytalanság, méghozzá a nem depressziós hozzászólások nagy arányának nem biomedikális kategóriába sorolása miatt.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

6. táblázat. Tévesztési mátrix az 5. táblázatban található 3. modellre

Becsült	Nem depr.	Biomedikális	Nem biomedikális	Összesen
Tényleges				
Nem depr.	94	23	104	221
	43%	10%	47%	
Biomedikális	23	185	53	261
	9%	71%	20%	
Nem biomedikális	43	23	349	415
	10%	6%	84%	
Összesen	160	231	506	897

Forrás: Saját szerkesztés

Megjegyzés: A mátrix a tényleges és a modell által becsült kategóriák eloszlását mutatja be. A cellákban található százalék a tényleges kategóriákon belüli, vagyis sorszázalék

A szociológiai és nem besorolható kategória alacsony előfordulása kézenfekvővé tette e két kategória összevonását és a biomedikális és pszichológiai típus mellett harmadik, egyéb kategóriaként történő szerepeltetését (lásd a 4. táblázat modelljeit). Azonban az irodalmak (Billings – Moos 1982) és az annotálás tapasztalatai alapján is két lehetséges kritikus pont merült fel. Az egyik, hogy a depressziós, de nem besorolható és a nem depressziós kategóriák között nem minden esetben volt egyértelmű a választás. A másik pedig, hogy a pszichológiai és szociológiai keretezés közötti határvonal sokszor nem volt egyértelmű. Éppen ezért úgy véltem, hogy megalapozott lehet egy olyan kategorizáció is, amelyben a nem besorolható kategóriát a nem-depressziós kategóriával, a szociológiai kategóriát pedig a pszichológiaival vontam egybe. Így a modellnek végül három kategóriába kellett sorolnia a hozzászólásokat: nem kategorizálható, biomedikális, pszichoszociális. A modellek eredményei a 7. táblázatban láthatók. Az eredeti (1. táblázatban látható) modellekhez képest ezek a modellek is jelentősen jobb teljesítményre voltak képesek. Ugyan a hatékonyságuk elmaradt a korábbi háromkategóriás (nem depressziós, biomedikális, nem biomedikális) modellek teljesítményétől, azonban a jelenlegi kategorizáció nem egyszerűsíti túl a feladatot és jobban megfelel a depresszió keretezési módjainak foglalkozó elméleteknek is.

Ennek a kategorizációnak a modellezés szempontjából az is előnye, hogy ez felosztás eredményezte a legkiegyensúlyozottabb tanító adatbázist: a nem depressziós esetek az adatok 29 százalékát, a biomedikális esetek a 30 százalékát, a pszichoszociális esetek pedig az adatok 41 százalékát adják, így nincs olyan kategória, amely túlzottan domináns lenne.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

7. táblázat. A különböző modellek specifikációi a nem depressziós vagy nem azonosítható, biológiai és pszichoszociális kategóriákra

	Modell	Accuracy/ Pontosság	Precision/ Precizitás	F1-score	Cohen- kappa
1	Stopszavazott szöveg	0,673	0,674	0,669	0,494
2	Stopszavazott szöveg, előfordul-e gyógyszernev a szövegben és beszélgetésfolyam száma	0,676	0,678	0,672	0,499
3	Stopszavazott szöveg, és fórum csoportja, beszélgetésfolyam száma, előfordul-e gyógyszernev a szövegben	0,675	0,677	0,672	0,498

Forrás: Saját szerkesztés

Ezen koncepció 3. modelljének alkalmasságát alátámasztják az osztályba sorolás során magas pozitív vagy alacsony negatív együtthatóval rendelkező szavak, amelyek részben természetesen azonosak a korábbiakban bemutatottakkal. A gyógyszernev előfordulását és a beszélgetésfolyam számát tartalmazó modell esetén a nem-depresszió kategóriához magas együtthatókkal tartoznak például a *god, friend, guy, jesus* szavak, míg alacsonnyal a *depression, anxiety, help, mental_health*. A vallásos szavak megjelenésének az az oka, hogy a transzcendentális kategóriát, vagyis amikor a depresszióhoz vallásos, természetfeletti keretezés társult, kezdetben külön kategóriaként kezeltük, de alacsony előfordulása miatt az egyéb kategóriába került. A biomedikális diskurzust a *medication, med, doctor, antidepressant, drug, depression, dose, side_effect, 00mg,¹⁶ take* szavak segítettek felismerni, a pszichoszociálist pedig a *therapist, psychologist, depression, anxiety, step, therapy, thought, hard, write*. Ez utóbbi szavak alapján is látható, hogy a pszichológiai aspektus a domináns. A biomedikális keretezés ellen hatottak a *friend, want, love, therapist, write, like, life* vagy *others* szavak, melyek között az érzelmekhez kapcsolódó kifejezések mellett a *write* is megjelenik, amely a naplórészhez, blogíráshoz is kapcsolódhat, amely az egyik, nem medikális jellegű megküzdési módszer és a *life* és *others*, ami inkább a szociológiai keretezéshez kapcsolható. A pszichoszociális kategória esetén nagy negatív együtthatókkal bír a *medication, med, god, antidepressnat, drug, doctor* szavak, de a tíz legalacsonyabb együttható között az *eat* is megtalálható.

A részletesebben ismertetett modell tévesztési mátrixa (8. táblázat) az együtthatókhöz hasonlóan azt mutatja, hogy a modell könnyebben felismerte a nem-depressziós hozzászólásokat a korábbi legjobb modellhez képest. Ennek oka vélhetően az, hogy a transzcendentális témák fogódzót nyújtottak ehhez. Megjegyzendő azon-

¹⁶ Az előfeldolgozás során a különböző számokat mind 0-ra cseréltem.



TEMATIKUS TANULMÁNYOK – Digitális szociológia

ban, hogy a nem depressziós kategóriában továbbra is bizonytalan a modell és nagy arányban sorolja ezeket a hozzászólásokat a pszichoszociális kategóriába.

8. táblázat. Tévesztési mátrix a 2. modellre

Becsült Tényleges	Nem depr.	Biomedikális	Pszichoszoc.	Összesen
Nem depr.	145 55%	26 10%	95 36%	266
Biomedikális	33 13%	187 72%	41 16%	261
Pszichoszoc.	69 19%	22 6%	279 75%	370
Összesen	247	235	415	897

Forrás: Saját szerkesztés

Megjegyzés: A mátrix a tényleges és a modell által adott kategóriák eloszlását mutatja be. A cellákban található százalék a tényleges kategóriákon belüli, vagyis sorszázalék

Összegzés

Kutatásomban azt vizsgáltam, alkalmazható-e a természetesnyelv-feldolgozás automatikus osztályozása annak felmérésére, hogy a depresszió szakértői keretezése megjelenik-e laikus diskurzusokban. A dolgozatban először a depresszió három, szakirodalomban előforduló keretezési módjának automatikus osztályozását végeztem el logisztikus regresszió segítségével. A kapott eredmények alapján az eredeti kategorizálást többféleképpen módosítottam, amely eredményeként a modell teljesítménye jelentősen javult. Ezek a módosítások nem csak a modellek hatékonyságának növelését szolgálták, hanem azt is, hogy megértsük, egy ilyen komplex osztályozási feladat során milyen attribútumok azok, amelyek segítik és melyek azok, amelyek nehezítik a besorolási feladatot. Mindezen eredmények pedig a kutatott téma és a gépi tanuláson alapuló automatikus osztályozás mélyebb megértéséhez is hozzájárulnak.

Az eredmények azt mutatják, az alkalmazott modellspecifikációktól függetlenül, az osztályozási feladat egyszerűsítésével is 60 és 70 százalék közötti pontosság és 40-50 százalék közötti Cohen-kappa értéket érhetünk el. Annak megítélésében, hogy ez az eredmény megfelelően megbízható-e vagy sem, többféle tényezőt kell figyelembe vennünk. Az annotátorok besorolását alapul véve azt mondhatjuk, az automatikus osztályozás is hasonló teljesítményt nyújtott, mint az emberi besorolás, vagyis az annotátorok (0,45 és 0,42 az eredeti osztályozási feladatban).



TEMATIKUS TANULMÁNYOK – Digitális szociológia

Ennek egyik oka az eredeti modellben származhat abból, hogy az osztályok eloszlása nagyon egyenlőtlen volt a tanuló adatbázisban: az egyéb és a szociológiai kategória mind számában, mind arányában annyira alacsony volt, hogy a gépnek nem volt elég lehetősége mintázatokat találni az ezekhez tartozó hozzászólásokban. Amely ilyen, többféle témát tartalmazó, ezért szerteágazó szókészlettel bíró kategóriák esetén még nehezebb feladat, hiszen nincsen sok hívószó, amelyek jelenléte segítené a kategóriába sorolást.

Annak megítéléséhez, hogy az általam alkalmazott modellek teljesítménye jó-e, érdemes megvizsgálni a témához köthető, depresszióval, öngyilkossággal kapcsolatos, laikusok által írt online szövegek automatikus felismerésével foglalkozó kutatások eredményeit. A fent ismertetett depresszió témájú kutatások pontossága 60-93 százalék közötti, (O’Dea et al. 2015, Thaul Lehrman et al. 2012) Cohen-kappa értékek pedig 20 és 55 százalék közöttiek. Ezen modellek azonban nagyobb különbségekkel bíró nyelvi struktúrákon állnak: korábbi kutatások alapján ismert, hogy eltérő szóhasználat jellemzi a depresszióval küzdőket (Losada – Crestani 2016). Az általam vizsgált kérdés nem a depressziós-nem depressziós elkülönítést kívánja vizsgálni, hanem azt, a különböző értelmezési módok felismerhetők-e az automatikus osztályozó számára. A kutatások leírása alapján a jelen kutatásban használt adatbázis variábilisabb volt, ugyanis a depresszió különböző szinonimái voltak maguk a kulcsszavak, amelyeknek a hozzászólásokban elő kellett fordulnia – s nem az a depresszióhoz, öngyilkossághoz szorosabban kapcsolódó kifejezések. Ez alapján a kapott eredményeim nem mutatnak jóval alacsonyabb teljesítményű modelleket, mint a hasonló kutatások, egyszersmind az idézett kutatások is kiemelik a feladat bonyolultságát, mind az emberi kódolás, mind a gépi besorolás során. Az alkalmazott modellek érvényességét támasztja alá az is, hogy sem a szövegek részben eltérő feldolgozása, sem az osztályozást segítő egyéb jellemzők bevonása nem eredményezett jelentős változást a modellek teljesítményében.

Az eredeti osztályozási problémához képest a szociológia kategória egyéb kategóriával történő összevonása némiképpen javította a modell teljesítményét, de az eredmények nagyobb mértékű javulása igazán a szociológiai és pszichológiai kategória összevonásával történt meg. biomedikális biomedikális Ezt támasztják alá az annotálás eredményei és az irodalmak (Kokanovic et al. 2012) is a szociológiai és a pszichológiai keretezési mód a laikus beszélgetésekben nem különül el egymástól, ugyanaz az egyén mindkettőt egyszerre tarthatja érvényesnek azokat saját vagy mások helyzetére.

Az eredmények általánosíthatósága korlátozott, nem állítható, hogy a depresszió keretezésétől eltérő kérdések esetén vagy magyar nyelven is ilyen megbízhatósággal működne a modellek, ez azonban nem is célja a kutatásnak. A vizsgálat az új technológiák nyújtotta lehetőségek, a gépi tanulás szociológiai alkalmazhatóságának felméréséhez kíván hozzájárulni azáltal, hogy megvizsgálja, egy hermeneutika-ilag bonyolultabb feladat esetén hogyan használható, miként teljesít egy gépi tanuló



TEMATIKUS TANULMÁNYOK – Digitális szociológia

algoritmus. Kutatásom eredményei alapján azonban elmondható, hogy az általam vizsgált fórumon az alkalmazott gépi tanuló algoritmus (logisztikus regresszió) az emberi annotáláshoz hasonló teljesítményt nyújtott. A modell teljesítményét a modellspecifikációk módosítása érdemben nem befolyásolta, így robusztus modellről beszélhetünk.

Irodalom

- Aggarwal, C. C. (2018): *Machine Learning for Text*. Springer Publishing Company, Incorporated. Yorktown Heights USA
<https://doi.org/10.1007/978-3-319-73531-3>
- Beck, A. T. (1979): *Cognitive therapy and the emotional disorders*. International Universities Press. Boston <https://doi.org/10.1192/S0007125000214918>
- Billings, A. C. – Moos, R. H. (1982): Psychosocial theory and research on depression: An integrative framework and review. *Clinical Psychology Review*, 2(2): 213–237. [https://doi.org/10.1016/0272-7358\(82\)90013-7](https://doi.org/10.1016/0272-7358(82)90013-7)
- Busfield, J. (2000): Introduction: Rethinking the sociology of mental health. *Sociology of Health & Illness*, 22(5): 543–558.
- Comas-Díaz, L. (2011). Multicultural approaches to psychotherapy. In: Norcross, J. C. – VandenBos, G. R. – Freedheim, D. K. (eds.): *History of psychotherapy: Continuity and change*. American Psychological Association, 243–267.
<https://doi.org/10.1037/12353-008>
- Comer, R. J. (2013): *Abnormal Psychology*. Worth Publishers, New York
- Dessewffy T. – Láng L. (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika (92-93)*: 157–170.
- Freud, S. (1985): Mourning an Melancholia. In: Coyne, J. C. (ed.): *Essential Papers on Depression*. New York University Press, New York, 48–63.
- Gao, S.-F. – Bao, A.-M. (2011): Corticotropin-Releasing Hormone, Glutamate, and γ -Aminobutyric Acid in Depression. *The Neuroscientist*, 1(17): 124–144.
- Goldstein, D. J. – Potter, W. Z. – Ciraulo, D. A. – Shader, R. I. (2011): Biological theories of depression and implications for current and new treatments. In: Ciraulo, D. A. – Shader, R. I. (eds.): *Pharmacotherapy of depression*. Humana Press, Totowa, NJ, 1–32. https://doi.org/10.1007/978-1-60327-435-7_1
- Kamali, M. – McInnis, M. G. (2011): Genetics of mood disorders: General principles and potential applications for treatment resistant depression. In: Greden, J. F. – Riba, M. B. – McInnis, M. G. (eds.): *Treatment resistant depression: A roadmap for effective care*. American Psychiatric Publishing, Washington, DC



TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Kokanovic, R. – Bendelow, G. – Philip, B. (2012): Depression: The ambivalence of diagnosis. *Sociology of health & illness*, 35(3): 377-390
<https://doi.org/10.1111/j.1467-9566.2012.01486.x>
- Labov, W. (1972): *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia
- Lakoff, G. (2004): *Don't think of an elephant!: know your values and frame the debate: the essential guide for progressives*. Chelsea Green Publishing
- Landis, J. R. – Koch, G. G. (1977): The measurement of observer agreement for categorical data. *Biometrics*, 33(1): 159–174. <https://doi.org/10.2307/2529310>
- Lewinsohn, P. M. (1985): *A Behavioral Approach to Depression*. In: Coyne, J. C. (ed.): *Essential Papers on Depression*. New York University Press, 150–180.
<https://doi.org/10.1016/j.jbtep.2010.10.001>
- Losada, D. E. – Crestani, F. (2016): A test collection for research on depression and language use. *International Conference of the Cross-Language Evaluation Forum for European Languages*: 28–39.
https://doi.org/10.1007/978-3-319-44564-9_3
- Martell, C. R. – Dimidjian, S. – Herman-Dunn, R. (2013): *Behavioral activation for depression: A clinician's guide*. Guilford Press, New York, NY
- Németh R. – Sik D. – Máté F. (2020): Machine Learning of Concepts Hard Even for Humans: The Case of Online Depression Forums. *International Journal of Qualitative Methods*, (19): <https://doi.org/10.1177/1609406920949338>
- Nguyen, T. – Phung, D. – Dao, B. – Venkatesh, S. – Berk, M. (2014): Affective and Content Analysis of Online Depression Communities. *IEEE Transactions on Affective Computing*, (5): 217–226. <https://doi.org/10.1109/TAFFC.2014.2315623>
- O'Dea, B. – Wan, S. – Batterham, P. J. – Cleave, A. L. – Paris, C. – Christensen, H. (2015): Detecting suicidality on Twitter. *Internet Interventions*. 2(2): 183–188.
<https://doi.org/10.1016/j.invent.2015.03.005>
- Pan, J. – Liu, B. – Kreps, G. L. (2018): A content analysis of depression-related discourses on Sina Weibo: Attribution, efficacy, and information sources. *BMC Public Health*, 18(1): 772–782. <https://doi.org/10.1186/s12889-018-5701-5>
- Ressler, K. J. – Mayberg, H. S. (2007): Targeting abnormal neural circuits in mood and anxiety disorders: From the laboratory to the clinic. *Nature neuroscience*, 10(9): 1116–1124.
- Rogers, A. – Pilgrim, D. (2005): *A sociology of mental health and illness 5th Edition*. Open University Press, Berkshire
- Sebők M. (Szerk.) (2016): *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*. L'Harmattan, Budapest



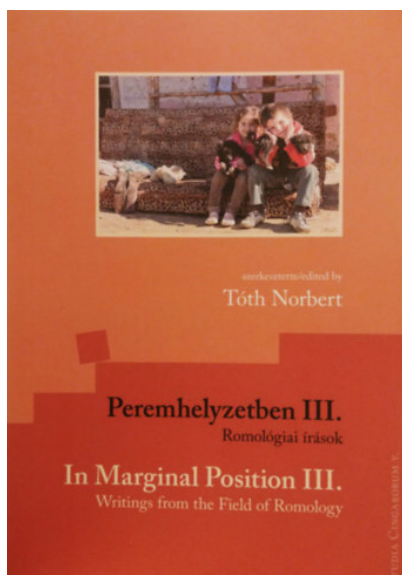
TEMATIKUS TANULMÁNYOK – Digitális szociológia

- Seligman, M. E. (1972): Learned helplessness. *Annual review of medicine*, 23(1): 407–412. <https://doi.org/10.1146/annurev.me.23.020172.002203>
- Sik D. (2018): From mental disorders to social suffering: Making sense of depression for critical theories. *European Journal of Social Theory*, 22(4): 477–496. <https://doi.org/10.1177/1368431018760947>
- Tan, P.-N. – Kumar, V. – Steinbach, M. (2012): Bevezetés az adatbányászatba – Alapvetés. Panem Kft., Budapest
- Thaul Lehrman, M. – Ovesdotter Alm, C. – Proaño, R. A. (2012): Detecting Distressed and Non-distressed Affect States in Short Forum Texts. *Proceedings of the Second Workshop on Language in Social Media*: 9–18. <https://www.aclweb.org/anthology/W12-2102>
- Tikk D. (Szerk.) (2007): Szövegbányászat. Typotex, Budapest
- Torzsa P. – Szeifert L. – Dunai K. – Kalabay L. – Novák M. (2009): A depresszió diagnosztikája és kezelése a családorvosi gyakorlatban. *Orvosi Hetilap*, 150(36): 1684–1693. <https://doi.org/10.1556/OH.2009.28675>
- World Health Organization (2017): Depression and other common mental disorders: Global health estimates. World Health Organization

RECENZIO

Studia Cingarorum – romológia a gyakorlati tudásforgalmazásban

A. GERGELY ANDRÁS¹



Egy kevésbé közismert, mégis alaposabb figyelemre érdemes kötetéről kell szólni az elismerés – és legalább annyira az elmélyült érdeklődés – hangján: Tóth Norbert szerkesztésében a Debreceni Egyetem adott ki tanulmánykötetet *Peremhelyzetben III./In Marginal Position III.*² címen, romológiai írásokból. A kétnyelvű kötet (ez esetben nem ugyanazok az írások szerepelnek két nyelven is, hanem a magyar dolgozatok angolokkal egészülnek ki) az egyetem Gyermeknevelési és Gyógypedagógiai Karán (a Lippai Balázs Roma Szakkollégium keretei között) indult kutatások utóbbi anyagaival, melyek a tizenöt éve folyó „romológiai, a kisebbségi cigány közösségek kultúráját, társadalmi viszonyait és együttélési kapcsolatrendszerét elemző” tevékenység eredményei, s ezeket (mint az előző két kötetet is) a *Studia Cingarorum* sorozatban jelentetik meg. E kötetek a ro-

mológiai kutatások tapasztalatait kívánják „a gyakorlati társadalomtudományi alkalmazásának lehetőségére” figyelemmel „a szélesebb közönség számára is hozzáférhetővé tenni”, ezzel egyszerre kínálva a szakembereknek, „hogy munkájukat bemutassák és a tudományos kutató és elemző tevékenység konkrét társadalmi jelentőséget az integráció, illetve a társadalmi felzárkózás összefüggésében értelmezzék”. (7. oldal)

Tudatos szerkesztési cél, hogy a debreceni karon „folyó képzések (cigány-roma nemzetiségi óvodapedagógus, szociálpedagógus) szakirodalmi kínálatát bővítse”, együtt más, például a „határon túli magyar tannyelvű felsőoktatási intézmények

¹ ELTE TáTK Kulturális Antropológia Tanszék, e-mail: andrasgergelya@gmail.com

² Tóth Norbert (szerk.) (2018): *Peremhelyzetben III. Romológiai írások*. In *Marginal Position III. Writings from the Field of Romology*. Didakt Kft., Debrecen, 279 oldal.



RECENZÍÓ

társadalomtudományi szakjain” folyó oktatási programokhoz is valamelyest segítő, kutatási és megértési modelleket népszerűsítő háttéranyagot kínáljon a kötet és a sorozat. Nem mellékes tehát a cél, hogy a könyvsorozat „a magyarországi és a határon túli magyar többségi környezetben élő cigány-roma kisebbségi közösségek társadalmi integrációjának ügyét, valamint az ezzel kapcsolatos hazai kutatásokat, esettanulmányokat” segítsen elő, s a kétnyelvűség része annak is, hogy „lévén a téma az Európai Unió kiemelt társadalompolitikai célterülete – a nemzetközi közvélemény számára is láthatóvá tegye”. E kötetben a Karon megrendezett nemzetközi konferencia (Kihívás és/vagy esély – A romák integrációja Európában, 2017) előadásai szerepelnek egymás után, nem tagolva tematikus vagy más rendszerbe. Ez azonban a hasonló köteteknél megszokott megoldás – a kutatói és elemzői státusz, a közlések címei már magukban is körvonalazzák tartalmukat.

A tizenhat előadás (huszonegy szerző) a legeltérőbb aspektusokat is egymás közelébe engedő megközelítési módjai a recenzensnek sem kínálnak többet, mint egyenkénti ismertetést. Erre azonban itt a tér is, az idő is csekély, nem utolsósorban pedig a szerzői tónus, a vizuális háttér, a szövegszerű idézetek, az ábrák és táblák, alapos bibliográfiák sokszínűsége csak érdemtelen „kivonat” elkészítésére készíthetne. Fontosabbnak látom az írásközhangjának, a szemléletmódnak megfogalmazását, mely egyúttal több dolgozatot is egyberendel, mintegy „összeolvasásuk” vagy párhuzamos tárgyalásuk válik így lehetővé.

Az „összeolvasás” persze nem előzmények nélküli. Két apropóból sem az. Egyfelől a sorozatban korábban megjelent kötetek már részben „előkészítették” a tematika helyénvalóságát: a sorozatban eddig megjelent Bocsi Veronika szerkesztette *Peremhelyzetben. Romológiai írások* kötet (2015), valamint a Biczó Gábor és Bálint Péter összeállította *Peremhelyzetben II. Romológiai írások* (2017) már tematikusan hangolták a mostani kiadvány tudományos horizontján elfoglalt megismerői pozíciókat, továbbá a Láczy Magdolna szerkesztette *Az esélyteremtés missziója. Tanulmányok a roma közösségi kultúra hagyományairól és a tehetséggondozásról* (2015) és a Pálfi Sándor összeállította *Roma gyermekek nevelése és segítése* (2015) ugyancsak állapot-ábrát kínált a romológia tudásterében akuttá váló kérdések részválaszaihoz. De a másik szempont sem mondható mellékesnek (túl a pályázati vonzáskörön és a résztvevők számára is Európa vonzásköre, az emberi jogi, kisebbségjogi és állampolgári partnerségi lehetőségek közvetlen körén). Ez pedig a lehetséges és szükséges összehasonlítási horizontok találkozása. Az összevetés idői és térbeli viszonyrendszert is meghatároz: egyszerre van jelen a térbeliség az időben Jenei Teréz: *Ciganológiai kutatások Oroszországban* (27.), Nagy Pál: *A cigányosztályok indításának politikai és pedagógiai dilemmái az 1950-es években* (83.), Maríá José Andrade: *The weight of history on the Roma situation in Europe: the case of the Spanish Gitanos and their legal status* (241.), Bálint Zsuzsa: *„Folktales with interspersed songs” – comparative analysis of the Singing Bone tale type in some Spanish and Hungarian Gypsy variants* (265.) című tanulmányában, és meghatározó momentum az időkomponens a térben



RECENZÍÓ

Anikó Vida: *Benefit recipients or citizens? Changes in the notion of unemployment in Hungarian parliamentary debates in two decades: 1991–2011* (139.), Draženko Tomić – Vladimir Legac: *Dealing with the Roma Issue in the Magazine Kršćanska obitelj (Christian Family) (Mostar, 1900–1920 and 1938–1944)* (181.), Eugenia I. Ivanova – Velcho Krustev: *The tinkering – the handicraft, that integrate gypsies toward neighboring in Bulgaria* (209.) című elemzésekben.

A téri és időtagoltság mint konstrukciós elem ugyanakkor nem hiányzik a strukturális, avagy intézményi és kulturális közösségi dimenziók körére épülő írásokból sem. Ezekben a pedagógiai, adaptációs, esélyegyenlőségi, kompetenciateremtési, értékszemléleti szempontok kapnak méltó hangot és elemző aspektust, ilyenek Benedek Katalin: *Esélyteremtés és vizuális kompetencia* (9.), Lehmann Miklós: *Integráció és világháló* (39.), Máté Dezső: *Generációs változások a Roma diplomások életútjában, a társadalmi reziliencia hatása* (49.), Mező Ferenc: *Alulteljesítő tehetséges roma tanulók azonosításának és ellátásának pszichológiai, pszichopedagógiai aspektusai* (71.), Oláh Dezső: *Cigányok a római katolikus egyház és a kisebbségek célkeresztjében* (107.), Orsós Anna: *Romológia a felsőoktatásban* (125.), Andras L. Pap: *Conceptual paradigms and policy discourses on the Roma in Hungary – and beyond* (149.), Edina Berta Héderné: *Who is served by child welfare services? Power as an agent in the changes in the content of personal help* (193.), Mihály Fónai – Lajos Hüse – Erzsébet Balogh – Andrea Toldi: *School integration, inclusion and exclusion: experiences of empirical researches* (221.) című írások.

Jóllehet, az ilyenfajta „elnagyolt” vagy elnagyoló áttekintés nem hoz túl sokat – sem a Szerzőknek, sem az Olvasóknak – azon a téren, mit is keressenek a könyvben, miért is kell figyelemmel követni azt, ami a roma kutatások látszólag háttérterülete, s a velük foglalkozásnak is csak keveseket érintő marginális tere. Meglátásom szerint (nem nagy felfedezés ez, csak megerősítő...) főképpen azért, mert a cigánykutatások megannyi évtizede sem volt elegendő ahhoz, hogy érvényes tudást szerezzünk a roma kisebbség – akár bármely létkérdésének, de mindenekelőtt – integrációs helyzetének, adaptációs megoldásainak, partnerségi útválasztásának vagy az ezt elutasító távolságtartásának alapkérdéseiről. Amikor az iskolai integráció is egyszerre kicsuk és bezár (Fónai, Hüse, Balogh, Toldi), amikor a gyermekjóléti alapkérdések legjava cigányoktatási problémakör (Héderné), vagy a mindennapi életvezetés alkalmazkodást elősegítő „politikáin”, pedagógiai „rejtett tantervein” múlik (Pap, Benedek, Orsós) a „felzárkóztatást” elősegítő tevékenységek strukturális elgondolása is – akkor már lassan fontosabb kérdéssé válik: létezik-e még az a („többségi társadalom” vagy „integrált” és „nemzeti” eszmény, amelyhez egyáltalán adaptációs modellek dolgozhatók ki. S ez esetekben lehet kulcskérdés, alapkompenciák szempontja az egyházak funkcióvállalása (Oláh, Ivanova – Krustev), a tehetséggondozás felelőssége (Máté, Mező, Lehmann, Pap, Héderné), továbbá a kortárs világjelenségekhez alkalmazkodás kihívásainak megtanítása (Benedek, Lehmann, Mező, Orsós, Vida, Héderné), s hasonlóképpen fölerősödik a kötet egészének egyik meghatározó aspektusa:



RECENZÍÓ

mindezek ismerete és kutatása mint kulcstevékenység. S miként Jenei Teréz az orosz ciganológia kutatásának felelősségét a 20. századi csekélyke eredmények és „népi kultúra” vagy „szubetnikum”-kérdés felől a komparatiztika és a szociálanropológia felé megtett lépések előnyeivel világítja meg, vagy ahogyan Bálint Zsuzsa a mesei hősök és maga a történetmesélés kultúrák közötti átíveléseit bemutatja a népmesei motívumok történeti változói révén, az is jelzi, mennyiféle párhuzamosan is releváns kérdés rejtőzik a ma még nem kellőképpen ismeretes kutatások háttérvilágában.

A kötet ezzel is, a romológiai kutatás számos új témakörének, oktatásának, megismerési folyamatának újabb szaktudományi problematikáját földolgozó szándékokkal is (továbbá nem feledendő: a kutatás horizontjának és az önértelmező szereptudatnak akár indirekt leíró szándékával is) oly kísérlet, melyre egyetlen kötet nem is lehet teljes válasz, csak egy hosszadalmasabb, kitartó, támogatott kutatás-sorozat egyik állomása. Még nem „Nagyállomás”, de már nem is csupán feltételes megálló. Folytatása, további szakkönyvek kiadása ezért is lehetne kulcsfontosságú. Meg hát nem utolsósorban azért, mert egyetlen kulcs sosem nyit minden megismerési kiskaput, értő szándékot, partnerségi horizontot. Ezeknek jövője viszont mindannyiunk belátási szándékától, érzékenységétől, tudományos igényességétől, méltóság-tudatától lesz biztosítható.



OLVASS FELESLEGESET!

– SIK ENDRE ÁLLANDÓ ROVATA –

Amióta a szakmában vagyok, harcolok a túlzásba vitt specializáció, a semmiről való mindentudás irányába mutató szakszerűség eluralkodása ellen. Természetesen a kudarc elkerülhetetlen, de kedvenc fegyverem ennek ellenére hasznossá válhat mások kezében is. Arról van szó, hogy világéletemben vigyáztam arra, hogy ha nyaralni mentem vagy feleslegesnek tűnő konferencián üldögéltem, akkor olyan tanulmányokat olvassak, amelyeknek semmi köze nem volt ahhoz, amivel akkor éppen foglalkoztam. Gondosan mindig valamilyen feleslegeset olvastam. Ezek között sok volt olyan, amibe beleszerettem, s félretettem, hogy valaha talán valakinek a kezébe adhatom, hogy foglalkozzon vele behatóbban, mert a mű valamilyen szempontból zseniális.

A következő elemzés egy 2015-ben íródott doktori értekezés alapján készült¹ a Glasgowban letelepedett szlovák és cseh bevándorlókról, akik történetesen romák.

¹ Guma, Taulant (2015): Everyday negotiations of in/securities and risks: an ethnographic study amongst Czech- and Slovak-speaking migrants in Glasgow. PhD thesis. <http://theses.gla.ac.uk/6315>



OLVASS FELESLEGESET!

Halmozódó hátrányok? – avagy cseh és szlovák roma bevándorlók mindennapjai Glasgowban¹

JUHÁSZNÉ FAZEKAS ZSUZSANNA²

A migráció kérdéskörével foglalkozó kutatások önmagában is érdekes olvasnivalót jelentenek a téma iránt érdeklődők számára. Taulant Guma disszertációja azonban még érdekesebb aspektusból közelít a migráció témájához, történetesen a kelet-európai cigányok migrációja felől.

A kutatásban azt vizsgálta, hogy milyen helyzetekkel, nehézségekkel kell szembenéznie a cseh és szlovák roma bevándorlóknak Skóciában a 2004-es Európai Unióhoz csatlakozást követően. Az adatgyűjtés időszaka 2012 januárja és decembere közé datálódik, helyszíne Glasgow, módszere etnográfiai terepmunka, megfigyelés.

A kutatás egyik fókuszpontjában az a kérdés állt, hogy az EU 8-ak migrációja pozitív vagy negatív hatást gyakorolt-e a brit szociális rendszerre és gazdaságra. A korai kutatásokban arról olvashatunk, hogy nem volt negatív hatása a bevándorlásnak (Gilpin et al. 2006, Blanchflower et al. 2007, Pollard et al. 2008), sőt Dustmann (2008) egyenesen pozitívnak bélyegzi a migráció gazdaságra gyakorolt hatását. Mások azonban a munkaerőpiac negatív irányú változásaira, valamint a fiatalok munkanélküliségének veszélyére figyelmeztettek (Scott 2008), noha a bevándorlóknak fizetett alacsony bér és a magas színvonalú szakképzettséget, hozzáértést, szaktudást nélkülöző munkaterületeken tevékenységek dominanciája jellemezte a bevándorlók munkaerőpiaci helyzetét (Anderson 2010).

A nemzetállamok jogokat szavatolnak az állampolgáraik számára, így ennek a kérdéskörnek a vizsgálata, a társadalombiztosítás és a migráció összefüggései a témával foglalkozó kutatók érdeklődésének középpontjában állnak (Beck 2000, Wimmer – Glick-Schiller 2002). Adódik a kérdés, hogy a 2004-es csatlakozást követően a hazájukat elhagyó migránsok milyen jogokkal bírnak abban az országban, ahová emigráltak. Ez a kérdés felmerül a kelet-európai roma bevándorlók kapcsolatban is, akik a csatlakozást követően Skóciába költöztek. Jelentős számuk hatást gyako-

¹ Guma, Taulant (2015): Everyday negotiations of in/securities and risks: an ethnographic study amongst Czech- and Slovak-speaking migrants in Glasgow. PhD thesis. <http://theses.gla.ac.uk/6315>

² Debreceni Egyetem Humán Tudományok Doktori Iskola Szociológia és Társadalompolitika Doktori Program, e-mail: fazekzsu@gmail.com



OLVASS FELESLEGESET!

rol a brit jóléti, közszolgáltatási, illetve társadalombiztosítási rendszerre, valamint a társadalom működésére egyaránt.

A disszertációban külön elemzés tárgyát képezi annak vizsgálata, hogy a roma bevándorlók megjelenése hogyan hatott Glasgow városára: milyen új kihívásokkal kellett szembenéznie a városi hivataloknak, hogyan változott a város képe, lakosságának összetétele, valamint mindennapi élete. Fontos kérdés, hogy ezek a kockázatok és kihívások mennyiben voltak a bevándorláshoz és mennyiben a bevándorló romák származásához kapcsolhatók, (Poole – Adamson 2008, Paterson 2011, Sime 2014, Clark 2014) akik Skóciában szegények és kiszolgáltatottak, kedvezőtlen az egészségi állapotúak és angolul alig beszélők voltak. Elgondolkodtató Poole és Adamson (2008) problémafelvetése, miszerint a roma gyermekek körében még mindig jellemző, hogy alacsony iskolai végzettséggel rendelkeznek, valamint a roma kultúrában kisebb hangsúlyt kap az oktatás fontossága. A nevelés és oktatás területén feltérképezhető hiányok hosszútávú következményei a későbbi munkavállalás területére, a munkaerőpiacon elfoglalt pozícióikra is hatást gyakorolnak.

Grill (2012) résztvevő megfigyelésen alapuló kutatása során arra jutott, hogy a kelet-európai migrációval együtt Glasgow városában hamar megjelentek a cigánysággal kapcsolatos negatív asszociációk (antiszociálisak, bűnözők, a többségi társadalomtól meglehetősen különböznek). Problémás csoportként tekintettek rájuk, akik potenciális veszélyt jelentenek a lokális közösségre.

A disszertációban vázolt kutatás másik fókuszpontja annak a jelenségnek a vizsgálatára irányult, hogy a roma közösségek szociokulturális környezetében milyen hatást gyakorol a szociális és gyermekvédelemmel foglalkozó szervezetek tevékenysége. A kliensek és a hatósági szakemberek dilemmáit, nehézségeit, a sztereotípiákat egy példán keresztül hozza közelebb a szerző, a roma gyermekek családból történő kiemelése felől közelít.

A glasgowi roma közösség vizsgálata során 28 emberrel állt szoros kapcsolatban a kutató. A családok Glasgow különböző részein éltek, családi állapotukat tekintve voltak közöttük házasságban élők, özvegyek, elváltak, gyermekesek és gyermektelenek egyaránt. Közülük tízen csehek és tizennyolcan szlovák nemzetiségűek voltak. A tanulmány szerzője a családok mellett kilenc szakembert is bevont a vizsgálatába, akik munkájuk révén kapcsolatban álltak a bevándorlókkal, úgy mint munkaügyi tanácsadókat, közösségfejlesztőket, és a jog területén jártas szakembereket.

Disszertációjában a szerző megjeleníti a bevándorlók és a hivatalnokok interakciójában elhangzottakat, valamint azt is, hogy az említett szakemberek milyen megjegyzéseket fűztek a bevándorlók életvitelére, életmódjukra a velük történő együttműködés során: „nem értem miért jönnek ide emberek alapvető készségek nélkül”, „tanultak volna angolul, mielőtt idejönnek”, vagy éppen a „fogalmuk sincs milyen egy állásinterjú”.

A roma bevándorlók helyzetét személyes példákon keresztül mutatja be a szerző és enged betekintést mindennapjaikba. Így közel hozza az olvasó számára a beván-



OLVASS FELESLEGESET!

dorló lét minden problémáját. Találkozunk munkájukat elvesztő és súlyos anyagi nehézséggel küzdő családokkal, etnikai hovatartozása miatt megbélyegzett roma férfival és olyan diszfunkcionálisan működő családokkal is, akik a gyermekvédelmi rendszer látókörébe kerültek.

A szerző a Simkos család élettörténetén keresztül fontos összefüggésekre világít rá.

A család 2005 decemberében érkezett Skóciába egy nyugat-csehországi kisvárosból, a jobb élet reményében. A házaspár férfitagja az építőiparban dolgozott, míg a feleség háztartásbeli volt. Nagy reményekkel és tervekkel vágtak az ismeretlennek, és a kezdeti időszak igen reménykeltő volt mindkettejük számára. A Nyugat-Glasgowtól nem messze működő JVC-gyárban dolgoztak mindketten, mígnem 2008-ban a gyár bezárása miatt elvesztették munkájukat. Kisebb alkalmi munkákból tartották fenn ezután magukat, de életükben állandósult a tartós munkanélküliség, a pénztelenség, az egészségügyi állapotuk hanyatlásnak indult, elszigetelődtek környezetüktől.

A használható nyelvtudás, valamint a szakképzettség hiánya nagyban befolyásolta a család életesélyeit, munkaerőpiaci lehetőségeit. Egy idegen országban nyelvtudás és alapvető készségek használata nélkül nehezen képzelhető el más út a bevándorlók számára, mint a marginalizáció.

Azt, hogy milyen romának és bevándorlónak lenni egy idegen földön, jól szemlélte Mr. Búrik, a roma bevándorló és egy támogató szakember találkozásának bemutatása. Mr. Búrik egy adózással kapcsolatos nyomtatvány kitöltése miatt kereste fel a segítő szakembert, aki az adatfelvétel során megkérdezte tőle szlovákul, hogy pontosan mikor érkezett az Egyesült Királyságba és milyen jövedelmi viszonyok között él.

Az ügyintézését követően a szakember hirtelen egy kérdést intézett az éppen távozni készülő ügyfele részére. „Mr. Búrik, mielőtt távozna, lenne még egy utolsó kérdésem. Ön roma származású, ugye?” A kliens némi hezitálás után, elpirulva és zavartan azt mondja, hogy „Igen. Félig. Miért?” A segítő szakember tájékoztatásul annyit közöl, hogy a nyomtatványon megjelölendő lehetőségek között nincs olyan, hogy félig. Csak az, hogy roma. Teljesen az. Az ügyfél nem tudta leplezni meglepődöttségét és zavarát, esetleges szégyenét, miközben a nemzeti és etnikai hovatartozását kérdezték tőle. A szerző kiemeli, hogy a történelemből ismert etnikai hovatartozást középpontba állító kérdések feltevése önmagában stigmatizáló és távolságot teremt, jelen esetben Mr. Búrik és a segítő szakember között.

A tanulmány másik nagyon fontos aspektusa a gyermekek családból kiemelésének kérdését helyezi a fókuszba.

A szerző a roma gyermekek családból kiemelésének problémájával egy közösségi térként funkcionáló helyen szembesült, ahol a környékbeli lakosok, romák és nem romák, segítő szakemberek és önkéntesek összejönnek, beszélgetnek egymás-



OLVASS FELESLEGESET!

sal, megosztják a problémáikat. Egy negyvenes éveiben járó szlovák asszony arról érdeklődött a szerzótől, hogy hallott-e már arról, hogy „elviszik a gyerekeket”. A nő egy szlovák anyanyelvű dokumentumfilmben hallott először arról a szlovák anyáról, aki Londonban él. A helyi ottani szociális szolgálat elvette a gyermekeit, és két éve sikertelenül próbálja őket visszaszerezni. A nő azt állította, hogy a *Sociálka* (szociális szolgáltatások, családvédelem, gyermekvédelem rendszere) pénzért viszi el a gyermekeket.

A szerző igen alapos és körültekintő, minden tényezőre kiterjedő kutatásba kezdett a témával kapcsolatosan. Feltárta, hogy az említett dokumentumfilmet a szlovák televízió sugározta 2012 szeptemberében, és mivel a szlovák adások és a közösségi média minden felülete elérhető a glasgowiak számára is, így a közösség értesült a televízióban bemutatottakról. Az említett londoni szlovák anya gyermekeit az őket ellátó orvos jelzésére emelték ki a családból, mivel fizikai bántalmazás nyomait vélte felfedezni a testükön. Az eljárásban a családból kiemelt gyermekek ilyenkor egy gondozó személyhez kerülnek elhelyezésre, majd egy hónap letelte és az eset felülvizsgálata után – amennyiben a problémák még fennállnak – a gyermekek nevelőotthonba kerülnek. A szóban forgó esetet bemutató műsor oknyomozó riport jellegű, amely több figyelmeztető riportot jelentetett már meg az Egyesült Királyságban tapasztalható szociális szolgáltatásokkal kapcsolatosan. Habár a riport szakértőket is megszólaltat, szociális szakembereket, jogászokat, a szlovák kormány kompetens szakembereit, mégis kedvezőtlen színben tünteti fel a brit gyermekvédelem rendszerét. A feltűnéskeltés közben a szociális szolgáltatókat egyenesen náci szervezetekként aposztrofálja, akiknek a gyermekek örökbe adása a célja, mindezt egy jól jövedelmező iparág köntösébe bújtatva, és ennek célpontjai a külföldi roma gyermekek. Mrs. Benaková, aki elsőként mesélt az esetről, egyenesen azt állítja, hogy fizetnek is a családoknak a gyermekekért, akár ötezer fontot is. A szlovák riportfilm állításai igen erősen hatottak a helyi roma közösségre. A közösségi média lehetőségeit kihasználva a szlovák fővárosban még tüntetést is szerveztek a brit nagykövetség elé a helyiek, a problémára történő figyelemfelkeltés céljából.

A nemzetközi jogszabályok és egyezmények, a gyermekek jogait szavatoló törvények igen szigorúak abban a tekintetben, hogy szavatolják minden gyermek biztonságát és biztosítsák jogaiknak érvényesülését. Szigorú szabályai vannak annak, hogy mit tehetnek a gyermekvédelemmel foglalkozó szervek és szolgálatok akkor, ha a gyermekek jólléte sérül, és veszélyeztető tényezők alakulnak ki körülöttük. Mivel az említett eset kapcsán a brit hatóságok nem kívántak állást foglalni, ez még inkább erősítette a szlovák média azon álláspontját, hogy a hatóságok hallgatása az ügy kapcsán egyenesen a bűnösségük elismerése.

Ennek tükrében nem meglepő, hogy a glasgowi romák között vannak családok, akik félnek a gyermekvédelemmel foglalkozó szervezetektől és a szociális szolgáltatásoktól is. Ezt a félelmet a gyermekek elvitelének hamis képzete táplálja, amely megnehezíti a velük foglalkozó segítő szakemberek feladatát. Gyakran a családok



OLVASS FELESLEGESET!

annyira tartanak tőlük, hogy a közösségi programokra sem viszik el a gyermekeket, ahol pedig ingyenes programok, étkezés, érdekes és kreatív foglalkozások várnák őket.

Ezek a hiedelmek, rosszul berögzült félelmek mindkét oldalt akadályok elé állítják. Azokat is, akik segítségre szorulnak, és azokat is, akik a segítségre hivatottak.

A romákkal folytatott segítő tevékenység, legyen az az oktatás, a szociális ellátások területe vagy akár a lakhatási és anyagi nehézségeket célzó szolgáltatások rendszere, nem nélkülözheti azt az elvet, hogy a helyi hivatalok, szolgáltatók és szélesebb kontextusban az az ország, ahová a bevándorlók érkeztek, megismerje a roma bevándorlók szociokulturális sajátosságait. Enélkül nehezen képzelhető el a bevándorlókat célzó, lokális szinten biztosított közszolgáltatások célba juttatása. Ez ugyanakkor kétirányú folyamat is, hiszen nemcsak a szolgáltatói oldalnak vannak „feladatai” a megismerés és elfogadás terén, hanem a szolgáltatásokat igénybe vevő oldalnak is, jelen esetben a kelet-európai roma migránsoknak is, hogy megismerjék és elfogadják annak az országnak a sajátosságait és szabályrendszerét, ahová érkeztek.

A szerző történeti kitekintésében a romák stigmatizálásával kapcsolatban arról ír, hogy ő maga a terepmunka során nem találkozott olyan bevándorlókkal, akiket érintettek a cseh(szlovák) születésszabályozás vagy szegénygondozás negatív intézkedései, vagy a családjuktól való elszakítás. De úgy véli, hogy a korábbi generációk által megélt sérelmek és igazságtalanságok a roma identitás részévé váltak.

Összegzés

A szerző terepen szerzett tapasztalatait olvasva, az interjúalanyok nehézségeit és problémáit megismerve azt is gondolhatnánk, hogy ez a történet akár egy magyarországi szegregátumban is játszódhatna, talán egy borsodi zsákfaluban, talán egy Szolnok megyei nagyváros peremkerületében. A glasgowi tapasztalatok könnyen párhuzamban állíthatók a hazai viszonyokkal, már ami a romákkal való szociális munka sajátosságait illeti. Számtalan hasonlóság és kapcsolódási pont található, és mind a roma identitáshoz kapcsolható. Általánosságban elmondható, hogy azok a szociális munkások, akik velük dolgoznak és terepmunkát végeznek, hasonló módon élik meg a nehézségeket itt is. Hiszen egy borsodi cigányasszony ugyanúgy fél a gyermek családból való kiemelésétől, mint egy szlovák anyanyelvű bevándorló cigány, aki Glasgow külterületén lakik. Így a szociális szolgáltatásoktól való félelem szintén egy közös pont. Gondolkodásra készítenek, hogy mit tehet a szociális szakma annak érdekében, hogy ne ezek az alaptalan félelmek, hiedelmek határozzák meg a kliensek és a szociális munkások kapcsolatát. Ugyanakkor empatikus módon azon is tündöndhetünk, hogy milyen lehet manapság cigánynak, kirekesztettnek, szegénynek lenni, és mi mint állampolgárok, szakemberek, segítők, hogyan tudunk előítéletektől men-



OLVASS FELESLEGESET!

tesen, címkézés és megbélyegzés, hibáztatás és bűnbakok keresése nélkül odafor-
dulni egy olyan közösség felé, amelynek problémáit csak közös társadalmi felelős-
ségvállalással kezelhetjük. Tudunk egyáltalán?

Irodalom

- Beck, U. (2000): What is globalization? Cambridge: Polity Press.
- Blanchflower, D. et al. (2007): The impact of recent migration from eastern Europe on the UK economy [Online]. London: Bank of England. The Impact of the recent Migration from Eastern Europe on the UK Economy, speech by David Blanchflower (bankofengland.co.uk) (Utolsó letöltés: 2021. 10. 25.)
- Clark, C.R. (2014): Glasgow's Ellis Island? The integration and stigmatisation of Govanhill's Roma population. *People, Place and Policy* 8(1), pp. 34–50.
- Dustmann, C. (2008): The effect of immigration along the distribution of wages. CREAM Discussion Paper No. 03/08. London, University College London. The effect of immigration along the distribution of wages (ucl.ac.uk) (Utolsó letöltés: 2021. 10. 25.)
- Gilpin, N. (2006): The impact of free movement of workers from Central and Eastern Europe on the UK labour market. Department for Work & Pensions, Corporate Document Services Working Paper No 29.
- Grill, J. (2012): It's building up to something and it won't be nice when it erupts: The making of Roma/Gypsy migrants in post-industrial Scotland. *Focaal - Journal of Global and Historical Anthropology* 62, pp. 42–54.
- McCullum, D. and Findlay, A. (2011): Trends in A8 migration to the UK during the recession. *Population Trends* 145, p. 77–89.
- Paterson, L. (2011): Unequal and unlawful treatment: barriers faced by the Roma community in Govanhill when accessing welfare benefits and the implication of Article 49 for the Equality Act 2010. Glasgow, Govanhill Law Centre
- Pollard, N. (2008): Floodgates or turnstiles? Post EU-enlargement migration flows to (and from) the UK. London: Institute for Public Policy Research
- Poole, L. – Adamson, K. (2008): Report on the situation of the Roma community in Govanhill, Glasgow.Paisley,University of West of Scotland. <http://www.bemis.org.uk/resources/gt/scotland/report%20on%20the%20situation%20of%20the%20roma%20community%20in%20govanhill,%20Glasgow.pdf> (Utolsó letöltés: 2021. 10. 25.)
- Scott, S. (2008): Staff shortages and immigration in agriculture. Croydon, Migration Advisory Committee



OLVASS FELESLEGESET!

- Sime, D. (2014): Roma families' engagement with education and other services in Glasgow. Glasgow, University of Strathclyde, Glasgow City Council
https://pure.strath.ac.uk/portal/files/39688102/Sime_etal_roma_families_engagement_with_education.pdf (utolsó letöltés: 2021. 10. 25.)
- Ulc, O. (1988): Gypsies in Czechoslovakia: a case of unfinished integration. East European Politics and Societies 2, pp. 306–332



Abstracts

New ways in exporting Society The potential of donation.based digital data collection

KMETTY, ZOLTÁN

More and more digital data is being generated every day, and more and more social science analyses are using Twitter, Instagram, or Facebook data. Many international and national studies have already explored the social science opportunities and dilemmas raised by the phenomenon of ‚big data‘ – but the issue of ‚access to data‘ has only been touched upon tangentially. And access to data is becoming increasingly difficult. What can we do if market players close the access to their data, and, if we find data available, the Research Ethics Board tells us to stop? The answer is simple: go to the users and ask them for the data. This approach is what the literature calls data donation. This paper will describe the data donation approach in detail, focusing on how researchers can access data through users on the current major Western platforms. The practical feasibility of data donation access will be illustrated based on a domestic pilot study.

KEYWORDS: data collection, data donation, social media, big data, Facebook

The challenges of supervised machine learning in sociological applications

NÉMETH, RENÁTA

The sociological applications of supervised machine learning, already well proven in industrial/business applications, raise specific questions. The reason for this specificity is that in these applications, the algorithm is tasked with learning complex concepts (e.g. whether a tweet contains hate speech). Supervised learning consists of learning to classify previously annotated (hate speech/non-hate speech) texts by the algorithm, looking for characteristic text patterns. The questions that arise are: how to prepare annotation? How can a hermeneutic challenge such as hate speech recognition be performed by annotators? Are routinely applied, detailed annotation guidelines helpful? The article also discusses how large companies perform coding on crowdsourcing platforms, and describes AI bias, which in this case means that annotators themselves introduce bias into the data. I illustrate these issues with our own research experiences.

KEYWORDS: supervised machine learning, annotation, crowdsourcing, AI bias



Hungarian Videoblogger Networks Online

HEGEDŰS DÁNIEL

The web 2.0 phenomenon and social media – without question – not only reshaped our everyday experiences, but they have established an environment for new types of social practices and social actors. The demotization (Turner 2010) effect of such technologies has created entirely new fields where celebrities might emerge from: one of them is videoblogging. Many video bloggers gained great reputation through peculiar micro-celebrity practices (Marwick 2015, Senft 2012), and, as a result, became key figures in distributing ideas, values and knowledge in today's society. These cognitive patterns are disseminated with a discursive apparatus that is largely based on social media activity, including posts, tweets, self-imagery and the videos themselves, which are tied to a certain logic according to environmental affordances, creating the possibility for fans to interact (share, comment, like, retweet etc.) with artifacts of the celebrity. This mechanism puts the celebrity in a so-called expert system (Giddens 1990) position as they provide adequate schemas of attitude, mentality or behavior. Most importantly, all of these public interactions are accessible for scholars to conduct scientific research. With the help of the SentiOne application this research attempts to reconstruct online networks of video bloggers based on mentions, which either occurred in an artifact (post, video description etc.) or in a fan comment. Apart from the network itself, SentiOne enables us to get insights regarding each individual connection established in it with different types of aggregated data.

KEYWORDS: digital sociology, expert systems, networks, discourse analysis

Empirical analysis of the judgment of unconditional basic income through YouTube comments

OLÁH, ESZTER

One of the world's largest video-sharing platforms is YouTube, where viewers can comment on the videos and their topics. The aim of this study is to examine the values and opinions about unconditional basic income according to the comment sections of several Youtube's videos which topic is the previously mentioned UBI which is receiving increasing attention in parallel with today's economic and social changes. Our research works with a mixed method, data collection, storage, sentiment analysis and the bag of words method which were implemented using IT procedures, while categorization was done through manual coding. The results of the sentiment analysis show that positive arguments appear to a lesser extent in the comments. Positive arguments have value characteristics such as inclusion, the principle of the right to exist, justice and freedom. Among the positive arguments feasibility enjoys the highest support. Negative category values arise more frequently, so the emphasis on the values of injustice, exclusion, unaffordability, and performance-orientation is dominant in the analyzed comments.

KEYWORDS: text mining, computer text analysis, sentiment analysis, unconditional basic income, work-based society



What's the matter? A text mining analysis of political topics and user engagement on politicians' Facebook pages during the 2018 Hungarian general election campaign

BENE, MÁRTON

The research investigates the way users interact with leading topics of the 2018 Hungarian general election campaign on candidates' Facebook pages. It expects that the prominent (immigration, corruption) and campaign-related topics generate more user engagement, while policy topics and mobilization content are less interacted. It also tests the theory of issue ownership in relation with user engagement. These expectations are tested on a dataset that includes all posts (38030 posts) posted by all candidates during the campaign (511 candidates). Topics are identified by text mining methods. The study demonstrates that corruption, development policy and campaign are highly engaged topics, while immigration was more interacted only on opposition politicians' pages since the followers of pro-government candidates engage less with immigration-related content. The most surprising result is that a reversed issue ownership effect can be detected since politicians are generally less successful with their own topics.

KEYWORDS: social media, issue ownership, immigration, user engagement, text mining

Where to go, net generation? Lifestyle-based segments of the Hungarian youth

SÜTŐ, ANNA

My study attempts to explore the lifestyle-based segments of the Hungarian youth through an innovative methodology based on social media data, incorporating the dimension of digitization into the creation of lifestyle groups. The examination of the segments' lifestyle attitudes is assisted by a review of the related theoretical milieu approaches, international and Hungarian empirical milieu researches.

KEYWORDS: lifestyle, milieu, youth research, Facebook

Social media communication in the digital medical space

SIMON, SÁRA

In the environment of 21st century technology, the transformation of information acquisition of health care and patients has had an increasing emphasis. Despite the earlier authoritative doctor-patient relationship, a need for an equal, cooperation-based communication has emerged and there are so many digital healthcare projects to achieve this (Koskova 2015).

Information acquisition on the internet has allowed patients that based on the increasingly available medical information they acquire information about their condition, become part of patient communities, ask for second opinions, and become committed helpers of their doctors in their disease (Meskó et. al 2017).

This can be especially true for patients with rare diseases, where a diagnosis might take even a decade, the patient needs lifelong condition maintenance and treatment, if it is available. While the proportion of patients with rare diseases is low compared to the whole of society, the number of such patients is approximately 30 million in Europe (EURORDIS), which means patients and



their relatives need not only a harmonized health care system, but extensive information so that they can live with the rare disease with less difficulty.

The aim of our study was to present the options of information acquisition in the social media, focusing on Twitter, via an interdisciplinary and social approach. In this study therefore we carried out a Big Data based social media analysis based on #Asthma and #CysticFibrosis databases of the Symplur corporation. This study results contain the complete online communication of 7 years (2012–2019) regarding these hashtags. The analysis has few levels including semantic research, stakeholder and hashtag review, engagement, and the whole tweet activity exploration.

KEYWORDS: *digital healthcare, e-patients, Big Data, cystic fibrosis, asthma*

Classification of depression-related online forums using Natural Language Processing

MÁTÉ, FANNI

The study of the phenomenon of depression is not new in sociology, but since the depression is becoming a wider social problem, it is still a relevant issue today. In addition to the bio-medical and psychological aspects of depression, the sociological perspective is becoming more noteworthy in the discourse about the causes of depression. In the research of the discourse on depression, the online texts offer many new possibilities, as the forum's anonymity and accessibility make the online seeking for help popular. In this research, natural language processing (logistic regression) was applied to find patterns in the definition of depression in lay discourses. These methods make it possible to analyze a large amount of text – which would have been difficult to process with human resources. During the analysis, 67 857 posts of English-speaking online forums were categorized along the categories of the scientific discourse about depression. This study presents the first results, which shows logistic regression classifier performs like the annotators. . Although the research has analyzed English-speaking forums, my findings may be useful to anyone observing abstract sociological concepts in online texts written by users.

KEYWORDS: *depression, Natural Language Processing mental health, online forums*