

10/2 (2012), 325–343

tmcs@math.klte.hu  
http://tmcs.math.klte.hu

**Teaching  
Mathematics and  
Computer Science**

## Metadata and education

ERZSÉBET FORCZEK

*Abstract.* This article is a (possible) conceptual educational model, which introduces data representation, information storage and retrieval possibilities on the Web in a way analogous to the levels of organization of metadata.

The model uses the traditional library and information systems as a starting point, referring to the levels and types of information organization, and describes directions of its development. General acquaintance with the dominant organizational levels and types helps to understand the information organization on the internet, the coexistence of both structured and unstructured elements, the closedness and deficiencies of the content of information, and also helps to find possible ways of correcting these deficiencies. One of the main advantages of model-driven approaches is that they, by using the well-known classical systems, make tangible the development of physical and content data organization types and levels of organization of information for medical students that usually do not possess informatics knowledge.

The conceptual model presented in details in the article can provide a basis for a general introduction to metadata and to develop curricula equally appropriate for traditional face to face classes, trainings and online courses.

*Key words and phrases:* metadata and classical library, internet, content retrieval, uses of metadata, semantics web.

*ZDM Subject Classification:* B10, B40, Q50, Q60.

### 1. Introduction

Adding semantics to medical documents, to make it easier for search engines to automatically process and integrate medical information available on the Web,

is a professional task of a specialists of a certain field (in our case a medical engineer informatics).

The practice of generating metadata schema requires, besides field related expertise, IT knowledge too, which includes acquaintance with different forms of storage possibilities (both technology and content-wise), with the methods of creating formalized knowledge, and with the working mechanisms of search engines.

However, the field of metadata is a diverse, wide-reaching and constantly changing area, so in traditional higher education curricula touch on it only sparsely and inasmuch as it can be connected to other areas. For graduate and undergraduate students dynamic and ad hoc metadata courses are organized. In these courses the curriculum usually includes only a special part of the current knowledge according to specific needs.

In Hungary library and information technology students learn in detail the development of local informatics systems, their feasibility criteria, the tools of feasibility, the relation of data, metadata and metadata systems, and the structure and functionality of system softwares and frame softwares. As an outcome of the current educational practice, after graduation both librarian and informatics trainees become excellent experts in metadata fields [21]. For others, such as trainees in medicine, nursing it is usually sufficient to be familiar with the main functions of systems involving file management, or data base management, code systems, image processing, and health systems, which form part of the classical curricula. Even the teaching of the usage of the web usually covers only the introduction to searching by formal criteria, the most important databases and web2.0 services. The emphasis is on the use, via medical examples.

Introduction into the theoretical foundations of metadata management is missing from the curriculum in Hungary. One reason for this is that working with metadata is obviously a complicated and complex task. Our goal is to reduce this complexity so that even non-IT students will be able to understand the essence of Internet information management.

In the following sections, this paper discusses the developmental stages of metadata, with focus on the relatively independently developing areas of libraries and informatics, (concentrating on the main stages of the evolution of fundamental classifications, the history of catalogues and bibliographic database systems, and the development of metadata in informatics) and the meeting-points of these two areas, relating to digital information storage and searches on the web, resulting

in strengthening of the role of metadata as information organizer and of the individuals as information publishers.

## 2. The role of metadata in classical library studies

Libraries have been using different local registration and classification systems for several hundreds of years. (see Figure 1 (a), (b)). The Dewey Decimal Classification, the DDC, was developed in 1876 for express library purposes [1]. It has subsequently been greatly modified and expanded through 22 major revisions [2]. At the end of the 19th century (bibliographers) it was transformed through international collaboration initiated by Henry la Fontaine and Paul Otlet. On the basis of the DDC, they set up the UDC (Universal Decimal Classification) system of library classification. Both the DDC and the UDC have since been greatly modified, extended and modernized several times [4]. As a result of these developments, the extent of the UDC had reached 66,000 records by the end of 2005 (see Figure 1 (d)) [3]. Efforts have recently been made to unify the two systems.

Thanks to their universality, unambiguity and extendability, these classifications are still in use today though not exclusively. Their inventiveness resides in the categories and the decimals denoting them, since the codes are entirely numeric and unlimitedly hierarchic [3]. The constant maintenance and development of these systems allows up-to date, precise and exact registration, and their structure permits fast and precise query.

Code systems are based on similar principles. Most of them apply content classification and are independent of digital storage, although the numeric codes built up hierarchically can also be of use in formalized digital processing and registering. The disadvantage for the user is that a conversion step is needed between the applied codes and the content.

A categorization of free text content that is favourable for the user is one of the greatest current informatics challenges; we will return to this question below.

In the fields of health care and medical science, the Bertillon Classification of Causes of Death was adopted by the International Statistical Institute at a meeting in Chicago in 1893 [13]. The diagnostic code system in use today, the ICD (The International Statistical Classification of Diseases and Related Health Problems), was created from this system through several revisions (at present the 11th revision is under way) [14]. Countless further classification code systems and nomenclatures have been created in health care to describe diseases and

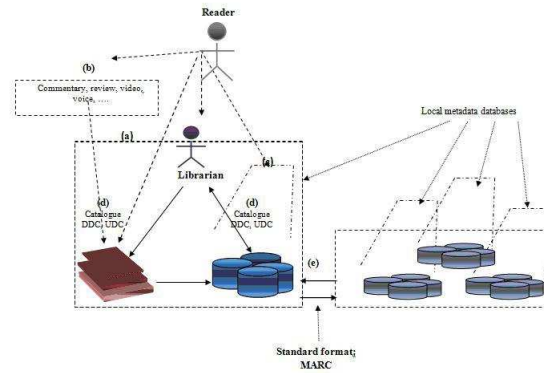


Figure 1. Bibliographic database

their treatment. With the emergence of electronic patient records, their role has increased in research and everyday therapy.

Larger libraries and documental institutions introduced digital processing and the storage of bibliographical descriptions in the 1960s. The demand for the sharing of these bibliographical records among institutions soon appeared. For such data exchange, an internationally accepted logical record format was created, which necessitated only conversion between the local digital format and the logical system MARC (Machine Readable Cataloguing) (see Figure 1 (e)). MARC has several international and national versions (LCMARC, USMARC, HUNMARC, etc.). The disadvantage of the MARC format is that it is primarily built up on the data units of the bibliographical description and categorization, i.e. on the formal characteristics, and it is connected to the form of the initial logical database management [11].

Rather than enhancing MARC and MARC-based systems, their aim is to give priority to interoperability with other encoding schemes and systems. There is a need to meet the demands that have arisen from the rest of the information universe.

Large research projects have been devoted to the development of MARC even in recent years. Series of patents and concepts clearly demonstrate that a format that can provide a universal solution for data exchange has not yet been found. Developers have to face not only the diversity of tasks and aspects of processing, but the new challenges and problems resulting from the constant changes in the

transmission and storage environment. Among the present research and development projects, researches within the framework of “OCLC research” are the most significant, aimed at finding transnational and transdisciplinary solutions for the joint use of MARC and non-MARC data [7], [8], [9]. Various institutions have developed different formats for their own bibliographical databases and for data exchanges through networks [12].

The emergence of the internet resulted in the appearance of new possibilities, expectations and problems in the dissemination of information [10]. It became necessary to reorganize access to information with regard to new principles not only in library science and health care, but additionally in all fields dealing with information organization [5]. This will be discussed below.

*The other main stream is digital information storage and representation, where metadata has played the main role, and provided the basis of up-to-date information management.*

### 3. The function of metadata in digital information storage

With the development of file managers and data-base managers, metadata has gradually gained importance and become a significant factor in the development of operating systems. Accordingly, this section discusses two important elements of digital data storage: file systems and database management; mention will be made of several stages that were crucial for the development of information systems and applications as we know them today.

A file system is a method for naming and locating files in order to store and retrieve them and the data within them. Use of the name of the tree structure with the file name is an appropriate way to create a name and a location that refer to the content of the file, and make retrieval easier. This identification system is based on the information attached to the file or located within the file, i.e. metadata. The field names of the records were essentially metadata (see Figure 2), which have since expanded to contain several dozen parameters of the file. The NTFS (New Technology File System), one of the most popular file systems nowadays, represents this development of metadata.

The file systems also undergo continuous improvement. Among the numerous significant developments, the results of the project conducted by the researchers at the Jack Baskin School of Engineering, at the University of California, are worth mentioning, in which the possibilities of traditional file systems are integrated with the rich hyperlink structure of the web. They created a Linking File System,

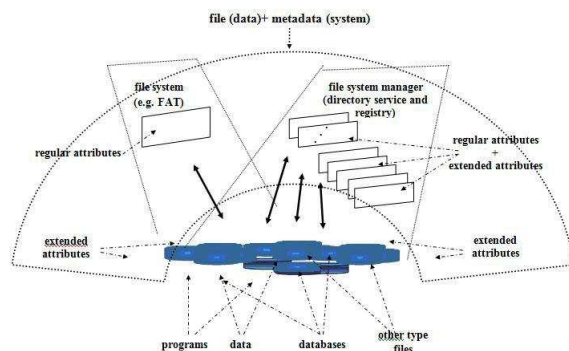


Figure 2. The file systems

“a file system design in which files may have both arbitrary user- or application-specified attributes, and attributed links between files” [15]. The joint application of traditional file systems and rich hyperlink structure can bring us closer to unfolding the deep web.

It is clear that the development of file systems (and the development of operating systems) is closely related to the development of databases. The demand for storage of the content parameters of the data together with the data, beyond the technical storage provided by the file system, emerged together with the development of storage applications.

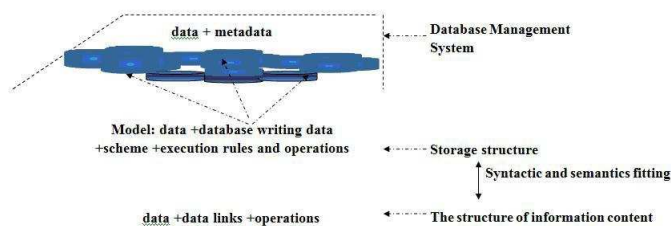


Figure 3. Database management

This demand led to the creation of databases. With the appearance of databases, information began to be stored with regard to the parameters and relations of the data tools. With the advent of database management system (DBMS), the use of metadata increased tremendously. This metadata includes

relation and attribute names, key and domain and other information (see Figure 3).

In the early 1990s, designs and software codes, knowledge and methodologies, subroutines, classes, or an encapsulated component: patterns and algorithm, software system architectures, project plans, design and user documents and other relevant knowledge sources were already known and used. However, in the course of technological development, new database paradigms also came to the fore, the interpretation/meaning of metadata became wider, and it already included the Complex Object Model, Nested Relation Data Model and Object-Oriented Data Model.

In parallel, several other fields, not expressly focused on data storage and retrieval, gained significance in science and medicine. Among the diverse application areas mention will be made of only a few that operate with metadata. One of these is CAD/CAM (Computer-Aided Design/Computer-Aided Manufacturing). Another special application that is worth mentioning is Virtual Reality (VR), which creates the virtual environment by using images and simulations. An example of an interesting application of VR is the Austrian Science and Research<sup>1</sup> project, carried out in the framework of international collaboration [35] where the virtual environment has been developed from 2000 images.

DSS (Decision Support Systems) is a possibility with great promise in numerous fields. One of its most important application areas is health care. Making diagnoses on the basis of clinical data is only a subset of the spectrum in which DSS can be used in a clinical setting. It may be based on either expert systems or artificial neural networks or both.

The use of *very large spatial databases* became a problem in the 1990s. The initial version of *GIS* (Geographic Information System) technology proved to be limited for the management of *large spatial databases* [19]. The concept of metadata was modified as follows: “Metadata is the data required to describe, locate process and environment entities, and control data in a database. It is the information that allows data identification and selection based on properties of data such as content, sources and quality”. One of the most well-developed element sets at present is the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) [16], [17].

In the past 20 years, the really significant concept of data warehouse appeared in data storage and retrieval. A *data warehouse* is a repository of the

<sup>1</sup>ASO-N/4/5 2007 (Austrian Science and Research) project: “Künstliche ,Intelligenz eines Therapieraumes für Gruppentherapie für AphasikerInnen”.

electronically stored data of an organization, designed to facilitate reporting and analysis (see Figure 4). Metadata in the data warehouse context guides the extraction, cleaning and loading processes and it includes usage and access maps, joins specifications, network security user privilege profiles, network security usage statistics and others. In the metadata layer, there are dictionaries for the entire warehouse and sometimes dictionaries for the data that can be accessed by a particular reporting and analysis tool.

A further interesting solution is the “metadata warehouse”, which was developed to manage metadata on the web [6].

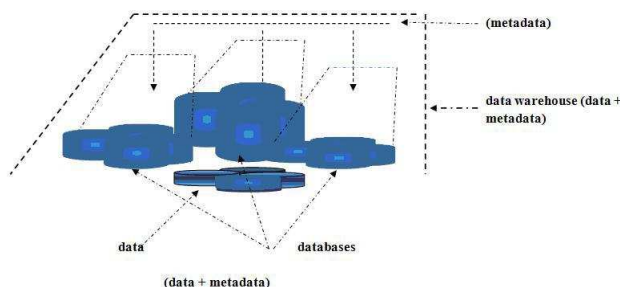


Figure 4. Data warehouse

The data warehouse concept has quickly gained ground with the expansion of the internet in every area of life. Data warehouses provide the information in the background of numerous internet portals [28], [18].

#### 4. Identification on the internet

End-users can manage the administration of files and search the contents of databases in local networks and on desktops with minimal effort thanks to file managing softwares. Running these special softwares usually does not require deeper informatics knowledge. On local computers, some parts of the information are well organized, while others, such as the loads of documents, images, videos and music, are stored in at most hierarchically organized folders.

The same structure, however, requires digital processing on the internet, due to its size. For search engines, the internet is an aggregate of information built up of semi-structured or unstructured elements, with free text, multimedia or



database documents on the surface, interrelated by links, which meets only loose formal and semantic requirements. For search engines, the content of multimedia elements or databases remains hidden; the storage structure (representation) of the information rarely reflects its content. Besides the storage structure, the individual term usages and formal solutions also make the managing of content more complicated on the web. Theoretical and practical researches strive to make the information on the web recognizable for search engines [27].

Below, the possibilities of identification and content-based search of data elements on the web will be discussed.

#### 4.1. Uniform Resource Identifier

As the registration of files in local computers and networks is very important (see above), the identification and availability of resources on the internet are equally essential. Using the URI, we can assign precisely defined meaning to internet resources, which means either designation of the location (URL - Uniform Resource Locator), or solely reference by name (URN - Uniform Resource Name) (see Figure 5). The URL contains the communication protocol required for reference (e.g. FTP, MAILTO or NEWS), the name (or IP dress) of the computer or the domain, the port number where the certain service is available, and the path within the destination computer where the resource is available (other parts of the name do not belong strictly to the name). The URN furnishes identification with a unique name, independent of the location and protocol, assigned to an object. The syntaxes of the URL are defined by standards; besides these, the denomination is optional. Terms in the name referring to content and address can help to achieve a better hit list.

#### 4.2. Content retrieval

The URI is suitable for identifying sites unambiguously. We can easily get anywhere we know or where links lead us during surfing. However, this is an infinitesimally small section of the information on the web. We attempt to reach the greater part of it by using search engines, with the help of keywords. The exact functioning of search engines is usually not public, though some of the general working principles, processing methods, algorithms and theoretical considerations are known (see Figure 6).

Most search engines are known to be built up from three main parts: collecting, indexing and searching.

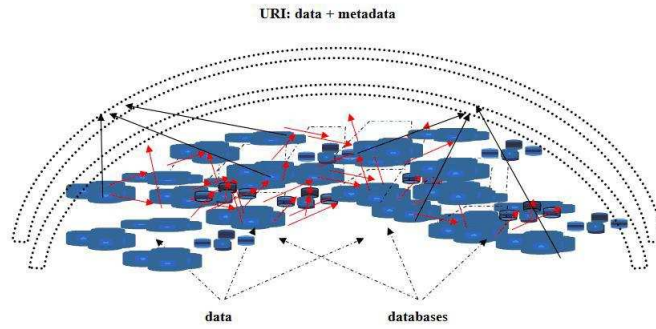


Figure 5. URI

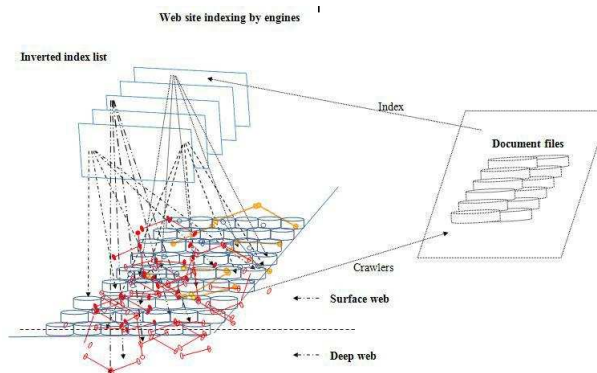


Figure 6. Search engine

Starting with URL lists and hyperlinks within them, the agents of search engines, the web crawlers (spiders or bots), constantly scan and map the web, using different algorithms. URL addresses are taken by the crawler control module, and the respective files by the page repository. Crawlers work according to the method described in robot.txt, their most important characteristics being the volume of the area they visit, and how often and how deeply they visit it. (Web pages can be registered in such lists so that crawlers become aware of them.) [22].

The next important element of the working of search engines is indexing, when the indexing program creates different index databases (such as link or text indices, and others). The precise process of indexing and the exact structure of the metadatabase developed are usually not known. The text index generally contains

an inverted index list, made up of the words of the documents, characteristics of index terms (e.g. PageRank value), and usually a short summary. Inverted lists are today the most common indexing technique; there are other types of search, but they are either too slow or too expensive.

The usefulness of a search engine depends on the third decisive element: searching. The most common way to search is searching by keywords taken out of context. Depending on the occurrence of keywords, the search engine provides a listing of web pages which best match the searching criteria. Search engines do not monitor content relations of free text information; at best, they perform some linguistic check. With the testing of different keywords, it is partly up to the endurance and intelligence of the user to find the relevant results. However, besides the quality of the web page, the intelligence and goals (business strategy) of the search engine also determine the relevance of the result set [32], [41].

Search engines crawl the web themselves to build listings by different principles and methods; the result sets of different search engines therefore differ. Instead of crawling the web themselves, metasearch engines send keywords to various search engines, utilizing relevant results sets, and after processing them deliver a more relevant spectrum of results than a single search engine (see Figure 7).

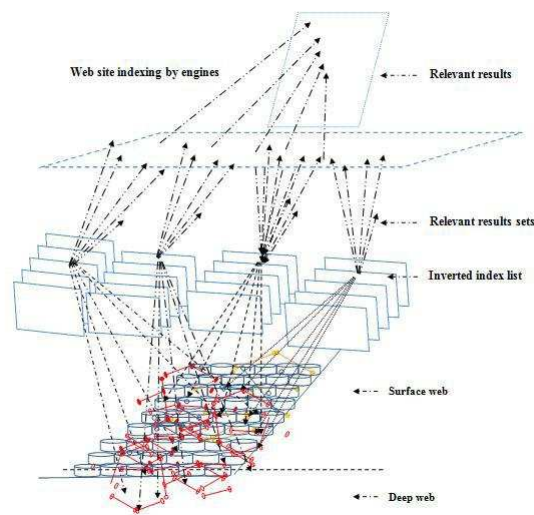


Figure 7. Metasearch engine

Search engines work with an enormous amount of information; however, compared to the volume of the web, it is still an order of magnitude smaller. The invisible part is called the deep web, which primarily consists of databases, multimedia elements and limited or banned information. Traditional search engines cannot find content in the deep web (See Figure 8) [39].

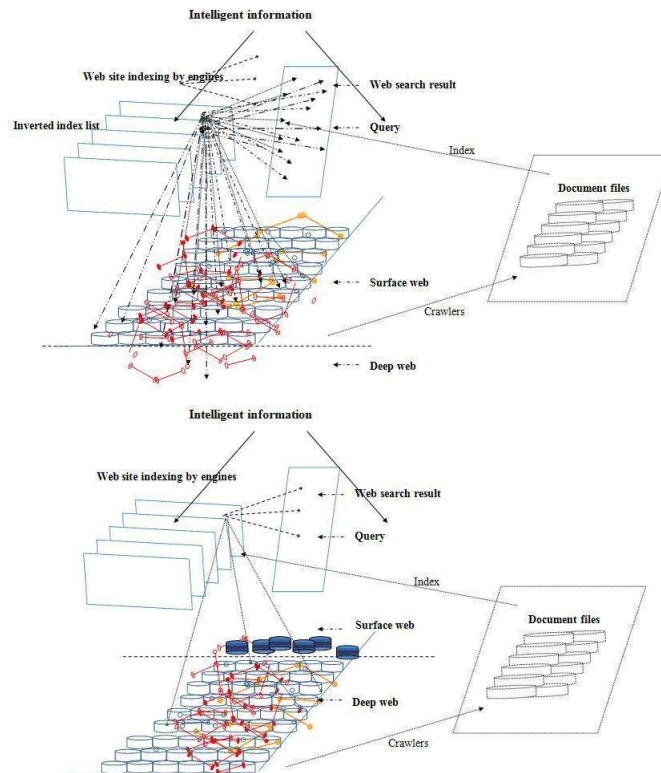


Figure 8. Searching on surface and deep web

It seems an obvious solution, similarly to the principle of databases, to create a thin layer, where unstructured data of the web or data not available to search engines are able to cooperate and communicate in a formalized way. Figure 9 shows an ideal concept created after the model of a database and a data warehouse. With the help of metadata, search engines try to determine the meaning of the text or other data and then create connections for the user. The availability

of machine-readable metadata would enable automated agents to access the web more intelligently.

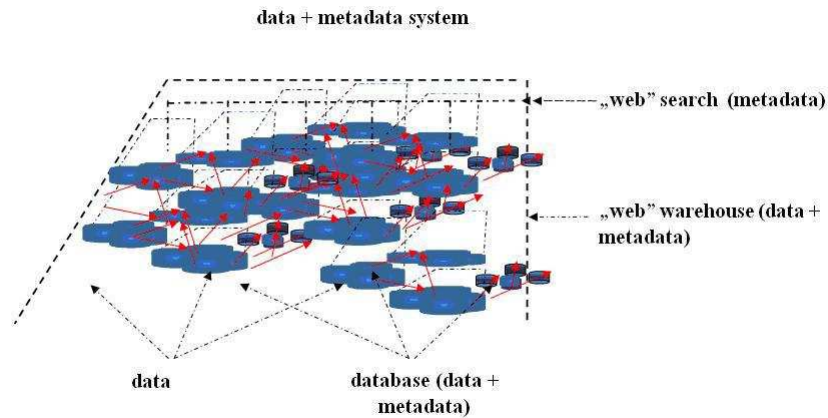


Figure 9. “Web” warehouse

Let us next investigate the possibilities of content coupling. The information organization methods described below are not new; they already existed independently from the web, but gained importance in content organization on the web [37], [38].

#### 4.3. Adding semantics to the data

Some search engines give priority to metadata. With a formalized description, we can make the content of databases and multimedia elements known, and emphasize the main points in the documents (see Figure 9). The degree of organization of formalized description can range from simple labelling to general ontology organization on a rather wide spectrum [29].

The easiest way to represent content is to use keywords, a method well known in library practice. Keywords can be scientific terms or everyday words (folksonomy). The use of keywords is of low efficiency and they do not give information about contexts [20].

Clinical codings have acquired an essential role in medical informatics since the introduction of clinical information systems. Medical coding is the process of transforming descriptions of diagnoses and procedures into universal medical code

numbers. These coding systems create classifications that cover most of the field of health care. Each is designed to help professionals communicate information about patients as concerns their treatment condition and medical history.

Thesauri contain more information than code systems; relationships between concepts can be represented explicitly (in the thesaurus). Hierarchies play a key role in allowing expanded retrievals. The hierarchical relationships (parent-child) thought of as broader or narrower relationships are better understood as representing these retrieval sets. The most significant thesaurus in medicine is the MeSH, a comprehensive, controlled vocabulary for subject indexing and the searching of journal articles, in primarily MEDLINE. This is one of the most sophisticated thesauri in existence nowadays. Descriptors or subject headings are arranged in a hierarchy; a given descriptor may appear at several places in the hierarchical tree. The tree locations carry systematic labels known as tree numbers.

Originally in English, MeSH has been translated into numerous other languages and allows the retrieval of documents; it can be browsed free of charge on the Internet through PubMed [23]. In Hungary, its translation has been started, but never completed.

Keywords, medical codes and thesauri are useful demonstrating aids for introduction of the different types of information organization. If these aids are used in education, students can acquire a better insight into the concepts of health care, and into information and metadata management on the web [24].

#### 4.4. Ontology organization

The most effective form of information addition is ontological engineering [40]. In information and computer science, ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It is the structural framework for organizing information, defines a set of representational primitives (classes, attributes and relationships), and includes information about their meaning and the possibilities and constraints of their logically consistent application.

The domain ontology model represents particular meanings of terms. Medicine, for example, has produced standardized, large, structured vocabularies such as SNOMED and semantic networks such as the Unified Medical Language System, which uses SNOMED [24], [25]. Many disciplines are now developing standardized ontologies that domain experts can use to share and annotate information in their fields. In medical informatics classes, ontologies are rarely built

together with students, and even if they are, only a very small part of medical ontology is surveyed through simple examples for demonstration purposes.

#### 4.5. Uses of metadata

Metadata is either technical, when it serves to identify objects and define their parameters, or descriptive, when it supports the identification of content. Descriptive data can be free texts, structured texts or texts with inner logical relations, i.e. ontologies. Metadata can ensure the consistency and the consequence of the content, and also the easy relation between different well-organized objects. This is essential in the efficient processing of multimedia and textual documents [34].

Due to the many different kinds of file formats [26], and to the fact that metadata is usually to be found within the object (.jpeg, DICOM, .TIFF, .pdf, .doc, etc.), present-day search engines can see at best only a part of the information stored on the web, and it is probable that the full content of only some of these objects will be accessible in the future. Metainformation can also be part of the web environment (e.g. the HTML called META), and free text annotations can be added to files. The former requires descriptive metadata, and the latter semantic metadata (see Figure 10).

For a content search, the Semantic web provides a standardized solution. Among the different options, the use of ontology is the most thorough, but since ontology development is rather time and energy-consuming, it is rarely applied. Several simple standards have been developed recently, but the Dublin Core (DC) metadata scheme, which consists of only 15 items, has become the most popular [36]. In the DC, we can work with universally interpretable elements (e.g. the author of an academic article and the writer of a script are equally a “creator”), and in item 13 annotations can be added. Formalization of the free text in the annotation (in a manner discussed above) [30], [31], [33] leads to better search results, i.e. if the annotation is organized into ontology (Protege, OWL), or standardized parameters are added, determined by thesauri or code systems [26] content relations too can be assigned to a document.

### 5. Summary

Metadata was already used some 150 years ago in early classification, catalogue and code systems, primarily in the field of library science and health care.

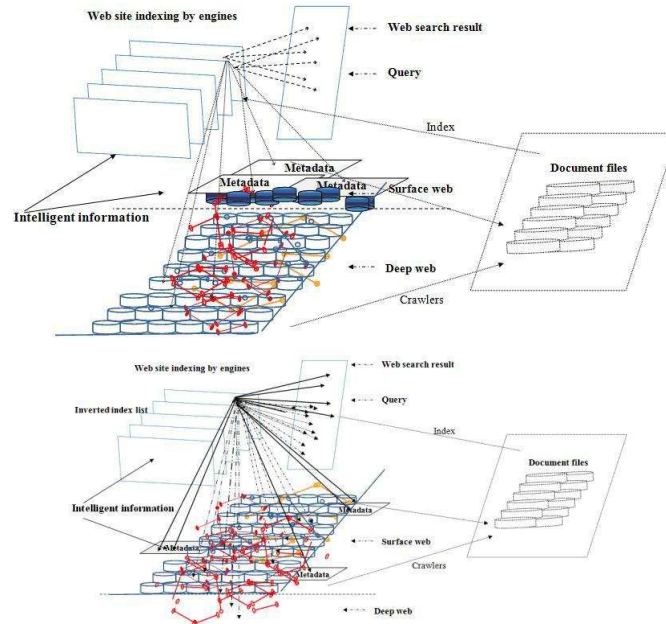


Figure 10. Search engine with metadata

The history of digital data storage and procession coincides with that of metadata, understood in terms of informatics as the additional information about information, necessary for digital storage and processing. While possible uses of the web seems endless, information about information is still largely missing. Therefore adding metadata to Web content is one of the most urgent research and educational directions and practical problems.

In this paper I presented a model for teaching metadata for medical students. The first two sections focused on the classical data storage and organization methods. In the third section I described the method we use to introduce our students to the theory and practice of efficient management of information on the web. Through these simple tools they are provided with information to increase their intelligence of creating internet data-sources, search information effectively on the Internet and to develop critical appraisal of information. In the fourth section, applying our method, I demonstrated the ideal and the actual organization of content on the internet, pointed out the emerging semantic gap, and ways to narrow it.



## References

- [1] Dewey Decimal Classification, Centennial 1876-1976, Facsimile reprinted by Forest Press Division Lake Placid Educational Foundation, Printed and Bound Kingsport Press, Inc. Kingsport, Tennessee,  
<http://www.gutenberg.org/files/12513/12513-h/12513-h.htm>.
- [2] Summaries DDC Dewey Decimal Classification OCLC Online Computer Library Center, Inc. Dublin, Ohio, 2003,  
<http://www.oclc.org/dewey/resources/summaries/deweysummaries.pdf>.
- [3] A. G. Taylor, *Introduction to Cataloging and Classification*, 10th ed., Westport, 2005, Conn.: Libraries Unlimited (Chapter 15: Decimal Classification).
- [4] I. C. McIlwaine, *The Universal Decimal Classification: A Guide to Its Use*, The Hague: UDC Consortium, 2000, <http://www.udcc.org/about.htm>.
- [5] A. Hajdu Barat, *Knowledge Organization in the Cross-cultural and Multicultural Society, Culture and Identity in Knowledge Organization*, (Clément Arsenault and Joseph T. Tennis, eds.), Würzburg: Ergon Verlag, 2008.
- [6] A. Sen, Metadata Management: past, present and future, *Decision Support Systems* **37**, no. 1 (2004).
- [7] K. Smith-Yoshimura, C. Argus, T. J. Dickey, C. C. Naun, L. Rowlinson de Ortiz and H. Taylor, Implications of MARC Tag Usage on Library Metadata Practices, *Report produced by OCLC Research in support of the RLG Partnership*, 2010, [www.oclc.org/research/publications/library/2010/2010-06.pdf](http://www.oclc.org/research/publications/library/2010/2010-06.pdf).
- [8] Key Perspectives Ltd. A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process, *Report commissioned by OCLC Research*, 2009,  
<http://www.oclc.org/research/publications/library/2009/2009-09.pdf>.
- [9] L. Ayers, C. B. Picknally, L. German, P. Johnson, C. Miller and K. Smith-Yoshimura, What We've Learned from the RLG Partners Metadata Creation Workflows Survey, *Report produced by OCLC Research*, 2009,  
<http://www.oclc.org/programs/publications/reports/2009-04.pdf>.
- [10] R. Erway, Impact of the Google Book Settlement on Libraries (Revised Version), *Report produced by OCLC Research*, 2009,  
[www.oclc.org/programs/publications/reports/2009-01.pdf](http://www.oclc.org/programs/publications/reports/2009-01.pdf).
- [11] MARC 21 Format for Holdings Data, OCLC Online Computer Library Center Product Code REF1111 0608/2000282W. 05M, 2006.
- [12] H. Dierickx and A. Hopkinson, *UNISIST, Reference manual for machine readable bibliographic description*, 3rd. ed. Paris: UNESCO, 1986, 2nd ed. (1981). 1st ed. (1974), <http://unesdoc.unesco.org/images/0000/000062/006279EB.pdf>.
- [13] History of the development of the ICD World Health Organization,  
<http://www.who.int/classifications/icd/en/HistoryOfICD.pdf>.

- [14] R. C. Madden, C. Sykes and B. Ustun, *World Health Organization Family of International Classifications: definition, scope and purpose*, 2007, [www.who.int/classifications/en/FamilyDocument2007.pdf](http://www.who.int/classifications/en/FamilyDocument2007.pdf).
- [15] A. Ames, N. Bobb, A. Brandt Scott, A. Hiatt, C. Maltzahn, E. L. Miller, A. Neeman and D. Tuteja, Richer File System Metadata Using Links and Attributes, *Proceedings of the 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2005)*, Monterey, CA., 2005, April.
- [16] Content Standard for Digital Geospatial Metadata, Metadata Ad Hoc Working Group, FGDC-STD-001-1998, Federal, <http://www.fgdc.gov/metadata/geospatial-metadata-standards#csdgm> Geographic Data Committee, <http://nationalmap.gov/gio/standards/>, [http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2\\_0698.pdf](http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf).
- [17] Geospatial Metadata: What are Metadata?, The Federal Geographic Data Committee, 2005, <http://www.fgdc.gov/metadata>.
- [18] P. Russom, *Wave of the future*, TDWI reports on what's next in the evolution of data warehouse platforms, 2010, <http://www.teradatamagazine.com/Article.aspx?id=13386>.
- [19] N. Trivedi and T. R. Smith, A Conceptual Framework for Integrated Metadata Management in Very Large Spatial Databases, *Technical Report 91*, no. 2 (1991), NCGIA National Center for Geographic Information and Analysis.
- [20] A. K. Majumdar and J. F. Sowa, Two paradigms are better than one, and multiple paradigms are even better, *Proceeding ICCS '09*, 2009.
- [21] I. Nahalka, Science Education and Theories of Science, *Magyar Pedagógia* **95**, no. 3–4 (1995), 229–250.
- [22] A. Gulli and A. Signorini, The indexable web is more than 11.5 billion pages, ACM Press, 2005, 902–903, <http://doi.acm.org/10.1145/1062745.1062789>.
- [23] S. J. Nelson and J. L. Schulman, *A Multilingual Vocabulary Project - Managing the Maintenance Environment*, European Association for Health Information & Libraries (EAHL) Workshop, 2007.
- [24] S. J. Nelson, S. H. Brown, M. S. Erlbaum, N. Olson, T. Powell, B. Carlsen, J. Carter, M. S. Tuttle and W. T. A. Hole, Semantic Normal Form for Clinical Drugs in the UMLS: Early Experiences with the VANDE, *Bio\*medical Informatics: One Discipline*, Proceedings of the Annual Symposium of the American Medical Informatics Association, 2002.
- [25] National Institutes of Health, About the UMLS, U.S. National Library of Medicine, [http://www.nlm.nih.gov/research/umls/about\\_umls.html](http://www.nlm.nih.gov/research/umls/about_umls.html).
- [26] E. Forczek and A. Szanyi, Barriers of using images in education, *Applied multimedia JAMPAPER* (2007), [www.jampaper.eu](http://www.jampaper.eu) [accessed 21/4/2010].
- [27] E. Forczek, On usability of scientific publications in education, *Applied multimedia JAMPAPER* (2007), [www.jampaper.eu](http://www.jampaper.eu) [accessed 21/4/2010].
- [28] H. Garcia-Molina, WebBase: Building a Web Warehouse, *Proceedings of the Fifth Mexican International Conference in Computer Science (ENC'04)*, 2004.

- [29] P. Auillans, O. Mendez, P. Rosenstiehl and B. Vatant, *A Formal Model for Topic Maps*, The Semantic Web, Springer, Berlin-Heidelberg, 2002.
- [30] B. Haslhofer, W. Jochum, R. King, C. Sadilek and K. Schellner, The LEMO annotation framework: weaving multimedia annotations with the web, *International Journal on Digital Libraries*, Springer, Berlin-Heidelberg (2009).
- [31] N. Manouselisa, J. Najjarc, K. Kastrantasa, G. Salokheb, C. Stracked and E. Duval, Metadata interoperability in agricultural learning repositories: An analysis, *Computer and Electronics in Agriculture* (2009, 2010).
- [32] E. Forczek, Metadata and information structure design on web sites - towards a web for all, *Int. J. Knowledge and Web Intelligence* (2011).
- [33] B. Popov, A. Kiryakov, I. Kitchukov, K. Angelov and D. Kozhuharov, Co-occurrence and ranking of entities based on semantic annotation, *International Journal of Metadata, Semantics and Ontologies* (2008), 21–36.
- [34] C. Taylor, *An Introduction to Metadata*, University of Queensland Library, 2003, <http://www.library.uq.edu.au/iad/ctmeta4.html>.
- [35] M. Horváth, Cs. Dániel, J. Stark and C. Sik Lányi, Virtual Reality House of Aphasic Clients, *Transaction on Edutainment III*, LNCS 5940, Springer Verlag, Berlin-Heidelberg (2009), 231–239, ISSN 1867-7207.
- [36] Dublin Core, 2004 Dublin Core Metadata Element Set Version 1.1, Reference Description, 2004, Retrieved 25 November 2008 from, <http://dublincore.org/documents/2004/12/20/dces/>.
- [37] K. Petridis, I. Kompatsiaris, M. G. Strintzis, S. Bloehdorn, S. Handschuh, S. Staab and N. Simou, Knowledge Representation for Semantic Multimedia Content Analysis and Reasoning, *European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, Royal Statistical Society, London., 2004, 25-26 Nov. 2004.
- [38] W. Yorick, The Semantic Web: Apotheosis of Annotation, but What Are Its Semantics?, *IEEE Intelligent Systems* **23**, no. 3 (2008), 41–49.
- [39] Ko. D. Madhavan, Kot, V. Ganapathy, A. Rasmussen and A. Halevy, *Google’s DeepWeb Crawl*, ACM, VLDB Endowment, 2008, ISBN 978-1-60558-306-8.
- [40] W3C, Metadata Activity, 2001, <http://www.w3.org/Metadata/Activity.html>.
- [41] W. Lucas and H. Topi, *Learning and training to search, E-LIS*, [Review of:] *New Directions in Human Information Behavior*, (A. Spink and C. Cole, eds.), Dordrecht: Springer, 2006, 209–228.

ERZSÉBET FORCZEK  
DEPARTMENT OF MEDICAL INFORMATICS  
UNIVERSITY OF SZEGED

*E-mail:* [forczek.ertzsebet@gmail.com](mailto:forczek.ertzsebet@gmail.com)

(Received September, 2011)