Qualitative and Quantitative Methods in Libraries (QQML) 4: 505-–515, 2015

Formal ontologies and their role in the integration of library resources

Jolanta Szulc

Institute of Library and Information Science, University of Silesia, Poland

Abstract: The purpose of this article is to analyze and compare the resources including knowledge organization systems used in the Polish scientific libraries. The study was based on analysis of the systems available in the libraries of some universities in Poland and other scientific libraries. This paper presents the vocabulary control systems used in Polish academic libraries: subject heading language (KABA and MeSH), thesauri, classification and formal languages used to write these structures. In conclusion, the paper presents proposals for future research. The results of the study are addressed primarily to the specialists in the library and information science, librarians and professionals involved in the creation of the domain and developing a formal ontology.
Keywords: Subject heading language KABA, Subject heading language MeSH, Universal National Central Catalogue NUKAT, Mathematics Subject Classification (MSC), Library of Congress Classification (LCC), Universal Decimal Classification (UDC), thesaurus, classification, formal languages, formal ontologies.

1. Introduction

The purpose of this article is to analyze and compare the formal ontology used in the Polish scientific libraries. The study was based on analysis of the systems available in the libraries of some universities in Poland. The results of the study are addressed primarily to the specialists in the library and information science, librarians and professionals involved in the creation of a domain and developing a formal ontology.

Formal ontology definitions refer to the definition of Thomas R. Gruber: ontology is an explicit specification of a conceptualization" (Gruber, 1993). The purpose of the investigations is to determine a formal ontology understood as a "system of organization to ensure the codification of the concepts of knowledge stored in a variety of documents and the relationships between them, which are the basis for the logical calculation and automated reasoning" (Sosińska-Kalata, 2005).

Received: 21.10.2014 / Accepted: 1.3.2015 ISSN 2241-1925

Formal ontologies, to ensure the organization of digital resources, determine the direction of development of Knowledge Organization Systems (KOSs). These resources include knowledge organization systems such as systems for controlled vocabulary (KABA, MeSH), thesauri, classification and formal languages used to write these structures. The systems are presented in Table 1.

Table 1. Knowledge organization systems in selected university libraries in Poland

No	The directory name of the university library (Access)	Systems for controlled vocabulary	Classification	Formal languages
1	The Computer Catalogue of the Jagiellonian University Libraries Collections (access: https://www.bj.uj.edu.pl/uj/katalog?lng=en)	Subject heading languages: KABA, MeSH	-	Bibliographic description language, VTLS/VIRTUA
2	The Wroclaw University Library (Access: http://www.bu.uni.wroc.pl/katalog/index.html)	Subject heading language KABA	Mathematics Subject Classication (MSC), Library of Congress Classification Outline (LCC)	Bibliographic description language, System VIRTUA, Program MAK, format MARC
3	The Catalogue online of Poznań University Library (Access: http://lib.amu.edu.pl/index.php?option=com_content&task=view&id=536&Itemid=92)	Subject heading language KABA	-	Bibliographic description language, system Horizon, system Virtua, format MARC 21
4	The University of Warsaw libraries online catalogue (Access: https://opac.buw.uw.edu.pl/cgi-	Subject heading languages: KABA, MeSH	Library of Congress Classification Outline (LCC), Universal	Bibliographic description language, system VTLS/Virtu

	bin/katalog/chameleon?sessionid=20130427040110119 65&skin=default&inst=consortium&function=START&lng=en)		Decimal Classification (UDC)	a; Inne systemy: ACCESS, MAK, MicroISIS, MicroVTLS or program of their own; MARC 21
5	The Computer Catalogue of the MCS University Main Library in Lublin (Access: http://kameleon.umcs.lublin.pl/cgi-bin/gw_2010_1_3_1/chameleon?skin=umcs&lng=pl)	Subject heading language KABA	-	Bibliographic description language, system Virtua, MARC 21
6	The Main Catalogue of The Library of the University of Lodz (Access: https://elibrary.lib.uni.lodz.pl/uhtbin/cgisirsi.exe/?ps=MWSf4Pb1j3/BU/279660131/60/506/X)	Subject heading language KABA	-	Bibliographic description language, system Symphony, MARC 21
7	The NCU Library Catalogue of University Library in Toruń (Access: http://opac.bu.umk.pl/webpac-bin/B_horizonPL/wgbroker.exe?new+-access+top)	Subject heading language KABA	-	Bibliographic description language, system Horizon, MARC 21
8	The Catalogue of The Library of the University of Silesia and University of Economics in Katowice (Access:	Subject heading language KABA	-	Bibliographic description language, system PROLIB,

	http://opac.ciniba.edu.pl/Scripts/cgiip.exe/wo2_search.p?R=1&IDBibl=19&ID1=DJJJQOHMQMGNFCGLLRNM&ln=pl)			MARC 21
9	The Computer Catalogue of the Library of the University of Gdańsk (Access: http://katalog.bg.ug.edu.pl/)	Subject heading language KABA	-	Bibliographic description language, system Virtua, MARC 21
10	The Main Catalogue Szczecin University Main Library (Access: http://katalog.bg.szczecin.pl/cgi-bin/koha/opac-main.pl)	Subject heading language KABA	Universal Decimal Classifica-tion (UDC)	Język opisu bibliograficznego, system Koha, MARC 21
11	The Main Catalogue of The Library of University of Opole (Access: http://aleph.uni.opole.pl/F?RN=245588909)	Subject heading language KABA, Library Faculty of Theology keywords	-	Bibliographic description language, system Aleph, MARC 21
12	The Central Catalogue of Library and Information System of University of Bialystok (Access: http://katalogi.uwb.edu.pl/F)	Subject heading language KABA	-	Bibliographic description language, system Aleph, MARC 21
13	The Catalogue of The Library of University of Warmia and Mazury in Olsztyn (Access: http://bu.uwm.edu.pl/F)	Subject heading language KABA, the symbol of the branch/filed	-	Bibliographic description language, system Aleph, MARC 21
14	The Catalogues of Cardinal Stefan Wyszyński University: books, multimedia,	Subject heading language, keywords	-	Bibliographic description language,

	maps (Access: http://baza.biblioteka.uksw.edu.pl/makwww/?BM=01)			system MAK, MARC 21
15	The Catalogue of Library of Rzeszów University (Access: https://opac.univ.rzeszow.pl/cgi-bin/wspd_cgi.sh/wo2_search.p?R=1&IDBibl=26&ID1=DGGGMHNLNMKEEDORKON&ln=pl)	Subject heading, subject matter	Internal classifica-tion	Bibliographic description language, system Prolib, MARC 21
16	The Catalogue of Library of University of Zielona Góra (Access: http://librarius.bu.uz.zgora.pl/cgi-bin/wspd_cgi.sh/wo2_search.p?R=1&IDBibl=33&ID1=HNFIRNJMQMGNFCJJLRNM&ln=pl)	Subject heading language	Universal Decimal Classification (UDC), internal classifica-tion	Bibliographic description language, system Virtua, MARC 21
17	The Library of Kazimierz Wielki University in Bydgoszcz (Access: http://82.146.234.5/)	Subject heading language	-	Bibliographic description language, system Horizon, MARC 21
18	The Main Catalogue of the The Library of the Jan Kochanowski University (Access: http://lib.ujk.edu.pl/F?RN=679047007)	Subject heading language KABA	-	Bibliographic description language, system Aleph, MARC 21

2. The systems for controlled vocabulary

To control the vocabulary used in the Polish academic libraries are applied primarily language subject headings KABA and Mesh (see Table 1). Work on the language subject headings KABA began in 1991 in the Library of the

University of Warsaw. Language KABA (Automatic Directory of Academic Libraries) is designed for use in large universal libraries. KABA Basic language features are defined in the general assumptions: unlimited vocabulary, vocabulary structure adapted to the filing of the detailing, vocabulary presented in the form of directory entries shown translatable into two languages RAMEAU and LCSH. It is co-created by our academic libraries during the cataloging of the collections (Głowacka, 2000). Currently language KABA is held at the NUKAT Center (Universal National Central Catalogue), which involved 129 libraries in the country (Access: http://centrum.nukat.edu.pl/)

Language MeSH Medical Subject Headings (MeSH) is the primary language of subject headings in the field of medicine, emerging since 1960, the National Library of Medicine, USA (NLM). Work on the Polish parallel version took place in 1997 in the Central Library of the then University of Medical Sciences and Jagiellonian University Medical Library. In 2001 the medical library directors signed an agreement which resulted in the consolidation of the work of these two centers and collaboration around the files of all other medical academic libraries. Currently "Card file MeSH subject headings 2012 - Polish version" is supplemented by the Polish equivalent of English terms. Some of the terms on the card come from the Medical Polish-English Thesaurus, issued by the Chief Medical Library in Warsaw in 1995-1996, the other were created on the basis of Polish medical terminology contained in the relevant medical dictionaries, language and domain-specific and other scientific sources. Some suggestions included in the original directory do not yet have equivalents in Polish, as work on the translation files are still ongoing (Access: http://www.mesh.pl/).

The National Library also uses another language: National Library of Poland Subject Headings. All these languages are controlled by the authority file.

3. The thesauri

The choice of a thesaurus as a knowledge organization system is affected by many factors, including: the existence of a relatively extensive literature on methods to create a thesaurus (including the Polish standard PN-N-09018: 1992. Monolingual thesaurus: the principle of creation, form and structure), flexibility thesaurus for automated information retrieval systems, the opportunity to present a paper descriptor lexical symbols and numerical (Nahotko, 2001). Proposed use of UDC to develop a thesaurus appeared in the preface to the third edition of the UDC tables 1982, FID 616). In Poland, have already been attempts to implement similar tasks (eg. in the Main Library in Warsaw Communication), but in a limited way. Some centers in Poland are using thesauri created outside the country, such as:

1. Central Agricultural Library of Michael Oczapowski uses AGROVOC, multilingual agricultural thesaurus maintained by FAO;

2. Main Medical Library in Warsaw creates PolMeSH (English Medical Thesaurus);
3. The Polish Institute of International Affairs, describes the collection of the Library Institute of International Affairs by the terms of the European multilingual thesaurus of International Relations and Regional Studies;
4. The Publications Office of the European Union (Publications Office) manages EuroVoc thesaurus based on ontology and semantic web technologies in accordance with the guidelines of the W3C, as well as recent developments
5. in standards for thesauri;
   (5) English Eurydice Unit provides the latest editions Thesaurus for Education Systems in Europe (TESE);
6. (6) Institute of Environmental Protection provides access to the Polish version GEMET (The General Multilingual Environmental Thesaurus);
7. (7) Central Institute for Labour Protection - National Research Institute developed The "Occupational Safety and Ergonomics" thesaurus, which is the adaptation of the Thesaurus of International Information Centre for Health and Safety at Work (CIS / ILO).

There exist projects for construction of new thesauri, such as "Thesaurus of Cultural Heritage" developed at the Institute of Art History, University of Wrocław (Access: http://www.historiasztuki.uni.wroc.pl/tezaurus.html) or "Thesaurus of physical culture" Tadeusz Szubry created in the Library of the Academy of Physical Education in Krakow (Access: http://www.awf.krakow.pl/index.php?option=com_wrapper&view=wrapper&Itemid=162). The following are examples of domain thesaurus created in the main libraries in Poland:

- Thezaurus of Theology Sciences is produced by The Federation of the Church Libraries FIDES (Access: http://www.fides.org.pl/index.php?option=com_content&view=article&id=111:tezaurus-nauk-kocielnych&catid=23:bazy&Itemid=73)
- Thesaurus of Geographic Names and International Organizations (Chmielewska-Gorczyca, 1997);
- Thesauri Parliamentary Library (STEBIS) (Access: http://bs.sejm.gov.pl/F/5M27AAPQC4M339TM3FCIS3KR56Q93I1P5LDICHPYTKQDY2RFAP-87227?func=find-b-0&local_base=tez10).

4. The classifications

The libraries surveyed use the following classification systems: Mathematics Subject Classication (MSC), Library of Congress Classification (LCC), Universal Decimal Classification (UDC) and internal classifications.

Mathematics Subject Classification (MSC) is a numerical classification; the scheme was established in the American Mathematical Society. Founded in 1959 as a result of the agreement of the editors of two journals bibliographies: American Mathematical Reviews and Zentralblatt für Mathematik German. The

classification has three levels and covers all branches of mathematics and their applications, and the links between them. Polish Mathematical Library of universities: University of Warsaw, Jagiellonian University, Nicolaus Copernicus University, University of Wroclaw, Adam Mickiewicz University, University of Gdansk, University of Lodz and others, and the Central Library of Mathematical create home directories with the classification in kind by MSC or MSC-based classification (Czarnocka-Cieciura, 2006).

Library of Congress Classification (LCC) is a system of alphanumeric symbols (consists of strings of letters and numbers). Library of Congress classification uses Latin letters and Arabic numerals. The first letter is the main section and the next subdivision. Additional numbering serves to clarify the content of the book. In 1998, only the Warsaw University Library was interested in the use of the online catalogue. LCC symbol stored in the 050 MARC bibliographic record has ensured proper indexing by the system. Currently, the Library is using or planning to use this symbol as a signature to set the open access collections bring KBK symbols to bibliographic records after they are copied from NUKAT to their local directories (Burchard, 2006).

The first Polish version of the summary tables Universal Decimal Classification (UDC) was published in 1923. From 1959 to 1972 the Institute of Scientific and Technical Information prepared and published a complete edition of the Universal Decimal Classification - FID 327. It was the first post-war full edition of UDC tables in the world. In the years 1974-1979 a four-volume version of the intermediate - FID 502 has been issued. Other issue summary tables UDC in Poland, were as follows: FID 424, 1969, FID 546, 1978, FID 667, 1988, P022 UDC Consortium in 1997 and P058 UDC Consortium, 2006. Interest in the use of the central directory NUKAT UDC classification exhibited a library of technical universities and the National Library (Access: http://www.bn.org.pl/dla-bibliotekarzy/ukd/historia-ukd).

In the libraries are also created internal classifications. These classifications organize collections of the Library and the Library of the University of Rzeszów and University of Zielona Góra.

5. The formal languages

Bibliographic description is a characteristic that allows to clearly identify the document, classified as a formal description language. As used herein, refers to the formal qualities of the publication. The Polish libraries are using the International Standard Bibliographic Description (ISBD), developed by the International Federation of Library Associations (IFLA). According to this standard bibliographic description form elements grouped into the following areas: content form and media type area, title and statement of responsibility area, edition area, material or type of resource specific area, publication, production, distribution, etc., area, material description area, series and multipart monographic resource area, note area, resource identifier and terms of

availability area. ISBD standardized elements and punctuation used in the bibliographic description. Consolidated edition published in 2011, includes the W3C standard RDF and assumptions ISBD/XML Study Group. It is the basis of universal bibliographic control, allows its use in an open online community and development of tools for the Semantic Web.

Formal languages also consist of a set of symbols and rules for combining them, but unlike natural languages are created by people in a conscious way by formally defined all the rules by which they should be applied. Formal languages, being the mediator between the man and machine (computer), are used to perform communication between them. They are also a form of communication and content instruction (Access: http://www.inzynieriawiedzy.pl/start). In these libraries, these languages are combined with computer systems (Aleph, Horizon, Mak, Prolib, Symphony, Virtua) and assist with the man-machine communication.

Formal languages are written using symbols, letters and words in a very precise way, which allows its automatic processing by machine. For example, the catalogue of the Library of the University of Lodz allows you to search by keyword, advanced search and search pattern. This search includes collections of the Central Library of the University of Lodz and a network of 91 departmental libraries.

Search by keyword includes all the words in the description of bibliographic items. The word or phrase should be typed in the box. After selecting the appropriate button (a word or search phrase, author, title, etc.), we obtain the results of the search. Search results get release dates in order from the newest to the oldest, and the string can be visible after opening the bibliographic description. If the number of results is very high due to typing too broad a search term, you can limit your search to: a specific library to the category selected from the set visible on the right side of the screen, the location of a copy, the document language or the year.

Advanced search allows precise selection criteria for retrieval and further filter of the search results. When you select an advanced search, on the right side of the screen also shows the additional possibility of narrowing down the search results by thematic categories, as in the case of keyword search. If the reader is offline it is possible to search by signature only the Main Library University of Lodz.

6. Conclusions

Modern knowledge management systems do not yet meet all the requirements addressed to them. One of these requirements is to create a metadata scheme. According to some authors, these schemes should use universal formal language designed to represent different in a data-structured way. Such language may be Extensible Markup Language (XML). Researchers have defined four stages of

the development of metadata schema: (1) text and databases (pre-XML), (2) XML documents for a single domain, (3) taxonomies and documents with mixed vocabularies, (4) ontologies and rules (Daconta, Obrst, Smith, 2003). There are studies on: how to use the ontology to formally describe DL (Digital libraries) architectures (Gonçalves, 2008), how to test and evaluate the Resource Description and Access (RDA) (Bothmann, 2011). The specification of the metadata structures and the implementation of its processing is informed by relevant principles and knowledge representation with formal ontology (Greene, 1997).

The Polish academic libraries use systems for controlled vocabulary (KABA, MeSH), thesauri, classification and formal languages applied to write these structures. All of these systems can be used to build a formal ontology, understood as a formal description of the field. Ontology understood as a generic class model of the field consists of three groups of elements: a set of objects of knowledge, a set of associated relations, a set of axioms that allow to build rules based on defined relationships. Most ontologies are built based on the first two elements. Incorporation of a rules engine is a more advanced intelligent ontological model (Rokicka-Broniatowska, 2004).

Ontology can be understood in terms of overall knowledge used in the computer system. The surveyed academic libraries use such systems as Aleph, Horinzon, Mak, Prolib, Symphony, Virtua. They allow modeling bibliographic relationships, the ability to formulate precise queries, facilitating users needs clarification by more accurate inquiries, directory integration, service charges for licensed access to the documents. The analysis of these issues will be the subject of future research.

References

Burchard, M., (2006). *Języki informacyjno-wyszukiwawcze w katalogu centralnym NUKAT* [Subject indexing languages in NUKAT union catalog of Polish research libraries] [Online]. [Cit. 2013-04-15]. Available at:
< http://www.nowyebib.info/publikacje/matkonf/kaba/burchard.php>.

Bothmann, R. (2011). Cataloging News. *Cataloging & Classification Quarterly*, Vol. 49, Issue 6, 542-566.

Chmielewska-Gorczyca, E., (1997). *Tezaurus nazw geograficznych i organizacji międzynarodowych*. Wydaw. Sejmowe, Warszawa.

Czarnocka-Cieciura, D., (2006). Język haseł przedmiotowych KABA a Mathematics Subject Classification. In: *Opracowanie przedmiotowe dokumentów z zakresu nauk ścisłych: matematyczno-przyrodniczych i technicznych. Język haseł przedmiotowych KABA: teoria, praktyka, przyszłość* [Online]. [Cit. 2013-04-15]. Available at: <http://www.nowyebib.info/publikacje/matkonf/kaba/czarnocka.php>.

Greene, S., (1997). Metadata for units of measure in social science databases. *International Journal on Digital Libraries*, Vol. 1, Issue 2, 161-175.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition,* Vol. 5, Issue 2, 199-220.

Daconta, M. C., Obrst, L. J., Smith, K. T., (2003). *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Wiley Publishing, Inc., Indianapolis.

Gonçalves, M. A., Fox, E. A., Watson, L. T., (2008). Towards a digital library theory: a formal digital library ontology. *International Journal on Digital Libraries*, Vol. 8, Issue 2, 91-114.

Nahotko, M., (2001). Ogólnopolska współpraca bibliotek technicznych w tworzeniu klasyfikacji opartej na UKD [Online]. *Biuletyn EBIB* Nr 6 (24) czerwiec [Cit. 2013-04-15]. Available at:
<http://www.nowyebib.info/biuletyn-ebib/24/a.php?nahotko>.

Rokicka-Broniatowska, A., (2004). Wybrane aspekty zastosowań ontologii w zarządzaniu wiedzą webowych systemów uczenia. In: *Multimedialne i sieciowe systemy informacyjne*. Vol. 1. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 433-443.

Sosińska-Kalata, B., (2005). Systemy organizacji wiedzy w środowisku sieciowym. In: *Od informacji naukowej do technologii społeczeństwa informacyjnego*. Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, Warszawa, 141-162