



Topic Map Library = Better Library: an Introduction to the “National Library of Poland” Project

Bartłomiej Włodarczyk
National Library of Poland
Bibliographic Institute
Warsaw, Poland
E-mail: b.wlodarczyk@bn.org.pl

Session:

**117 — *Subject access now: inspiring, surprising, empowering* —
Classification and Indexing**

Abstract

The National Library of Poland Subject Headings (JHP BN) is used in the Polish National Bibliography, the national library catalogue, as well as many other Polish libraries: public, pedagogical, church and academic. Subject searches in our library catalogue are still comprised of a significant number of all searches, but understanding and exploration of subject headings cause many problems, not only for the end-users, but also for many librarians. The solution could be a properly designed web application based on a topic map using appropriate visualisation that supports indexing and retrieval in the National Library of Poland. The paper presents the main stages of a planned project.

Introduction

Since 1957, the National Library of Poland has been developing its own indexing language, the National Library of Poland Subject Headings (JHP BN), used in the Polish National Bibliography, the national library catalogue as well as many other Polish libraries: public, pedagogical, church and academic (Klenczon & Stolarczyk, 2007; Klenczon, 2011). It is a pre-

coordinated system built according to existing standards. The subject authority file, maintained in the MARC21 format, contains approximately 69000 headings (preferred terms), 87000 variants (non-preferred terms), and 117000 relationships between the terms. Subject searches in our library catalogue still comprise a significant number of all searches – statistics of transaction logs show that it is almost 20% -- and the proportion of failures in finding information is not high. However research has demonstrated that understanding and exploration of subject headings cause many problems, not only for the end-users, but also for many librarians (Franz, Powell, Jude & Drabenstott, 1994; Drabenstott, Simcox & Fenton, 1999; Antell & Huang, 2008). One of the important issues is to improve the quality and speed of indexing. We do not want to compete with search engines, but the speed of providing information is important for end-users. We also want to improve the time spent in teaching new indexers. This is connected with the complication of the pre-coordinated grammar. Another important issue is to help indexers in finding the best term in the complex network of relationships to index specific book or other resources¹ (Library of Congress Subject Headings Pre- vs. Post-Coordination, 2007). Our observations and statements of librarians from libraries that use JHP BN confirm these insights. We are aware that a more efficient way of information retrieval is required, but we do not want to waste high quality controlled vocabulary.

The first step to improve subject information discovery was an attempt to simplify the syntax and to shorten the subject headings: form subdivisions have been taken out of the subject headings strings and later transformed in genre/form index terms. This was the first step toward facetisation. Facetisation is thought to have many advantages (Condit Fagan, 2010), but it does not bring a solution to one of the crucial issues facing libraries, today and in the future, which is the way of presenting library resources in the online environment. The library still offers a textual representation of catalogues, although appropriate visualised information is easier to understand and navigate. We will continue the work on simplification of JHP BN. Our intention is to develop JHP BN vocabulary and to simplify its grammar and application rules. We also want to carry out research that will improve in the future process of indexing in the National Library of Poland and other Polish libraries. We

¹ These problems are observed in our everyday work.

also want to try mapping Polish subject headings to the notations of the Universal Decimal Classification.

The importance of the visualisation has been identified in cognitive systems theory (Ware, 2004)]. Already at the beginning of the 19th century, Scottish engineer and political economist William Playfair said that, “regarding numbers and proportions, the best way to catch the imagination is to speak to the eyes” (Palsky, 1999). His opinion concerned only numbers, but this statement is also true regarding concepts and words. This could be the clue for libraries to highlight information visualisation. It is a common phrase – “a picture is worth a thousand words”. A traditionally organised library OPAC that presents a list of subject headings as linear text (Figure 1), cannot support effective use of subject access points and relationships between them.

In order to convey information clearly and to enhance its use, the better visualisation of information seems to be required. The subject authority file may serve as a base for creating a new tool for the effective and user-friendly discovery, navigation and information exploration.

1	Bibliotekarstwo -- 18 Zbiornych Eski Brzesmotowa	18
2	Bibliotekarstwo	377
3	Bibliotekarstwo -- 5-15 w.	2
4	Bibliotekarstwo -- 15-18 w.	2002 1
5	Bibliotekarstwo -- 15-18 w.	2
6	Bibliotekarstwo -- 16 w.	1991 1
7	Bibliotekarstwo -- 18-20 w.	1998 1
8	Bibliotekarstwo -- 20 w.	2
9	Bibliotekarstwo -- 1918-1939 r.	9
10	Bibliotekarstwo -- a bibliofilia	1929 1
11	Bibliotekarstwo -- a cenzura	2007 1
12	Bibliotekarstwo -- a chrześcijaństwo	2002 1
13	Bibliotekarstwo -- a edytorstwo	1993 1
14	Bibliotekarstwo -- a księgarstwo	1992 1
15	Bibliotekarstwo -- a kształcenie ustawiczne -- Europa -- od 1989 r.	1996 1
16	Bibliotekarstwo -- a kształcenie ustawiczne -- Wielka Brytania	1992 1
17	Bibliotekarstwo -- a kultura informacyjna	2009 1
18	Bibliotekarstwo -- a nauczanie na odległość	4
19	Bibliotekarstwo -- a oświata	1972 1
20	Bibliotekarstwo -- a oświata dorosłych	2
21	Bibliotekarstwo -- a oświata dorosłych -- Wielka Brytania	1992 1
22	Bibliotekarstwo -- a oświata -- Niemcy -- od 1950 r.	1998 1
23	Bibliotekarstwo -- a pluralizm społeczny	2007 1
24	Bibliotekarstwo -- a pluralizm społeczny -- kraje Unii Europejskiej	2006 1
25	Bibliotekarstwo -- a praca kulturalno-oświatowa	3
26	Bibliotekarstwo -- a praca kulturalno-oświatowa -- Szwajcaria -- 1945-1985 r.	1967 1

Figure 1: The way of presenting subject headings in National Library of Poland (<http://alpha.bn.org.pl/>)

There are many ways of visualising information including graphs, maps, trees, virtual worlds and others. Even simple transformation of subject headings into a topic map may be profitable for librarians. Visualisation of a topic map enables us to look at our complex structure of terms as a whole. This may lead to an improvement of the indexing language. Therefore, topic maps can be considered as an evaluation tool. In this way all inconsistency can be seen in a detailed way (Ware, 2004, p. 3). First of all, incorrect relationships amongst the terms or lack thereof. There are a few examples of using information visualisation on web sites of libraries and publishers. One of these is the WorldCat Identities Project, which creates, “an interactive Related Identity Network Map for each Identity in the WorldCat Identities database” (WorldCat Identities Network, n.d.). Graphical representation of metadata is also used in the HighWire platform. This platform is using a hyperbolic tree to browse documents by topics (HighWire, n.d.) (Figure 2).

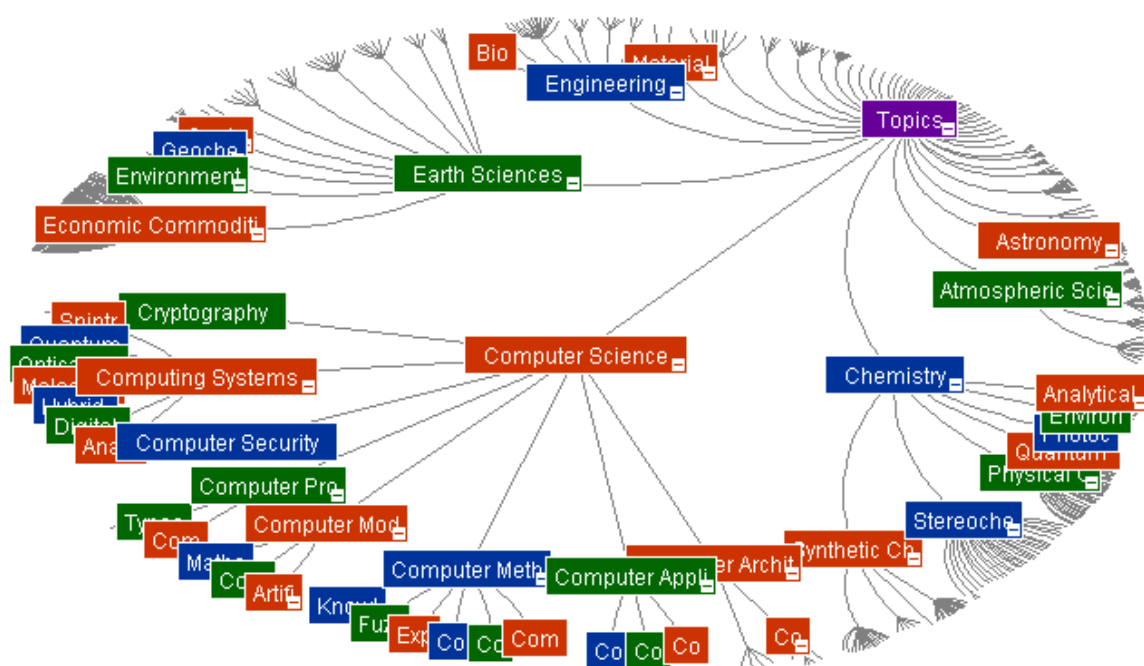


Figure 2: Browse by topics in HighWire (<http://highwire.stanford.edu/>)

In the Aqua-Browser software it is possible to see only one term and its relationships. It seems convenient. However, it causes difficulties when we perceive only a small spectrum of

relationships. Likewise, it provides limited context to search terms and, consequently, one of the main advantages of visualisation is lost.

In libraries there is a need to design ways of visualisation that are appropriate, not only for desktop computers, but also for mobile devices with touchscreens. Touchscreens are already available in desktop computers and the new version of the most widespread operating system, Windows 8 and Internet Explorer 10, will have a touch keyboard and will support gesture-based navigation (Windows Developer Preview: Windows 8 guide, 2011, p. 8-9, 13-14). We expect that the number of library users who access library catalogues with mobile devices will increase. In Poland this is just the beginning of this trend. So far only a small number of users use this method. In the Digital Library of the Wielkopolska Region, the number of such users is only 0.19% of all users regardless of the devices or the operating systems being used (Werla, 2011). Visualisation can support more natural navigation and that is one of the reasons why we should work on this issue.

Another problem in the National Library of Poland is the insufficient use of relationships built by JHP BN team. Building hierarchical and associative relationships is very time-consuming. These relationships are valuable and we have to increase their use. We also want to enlarge the number of relationships to improve information retrieval.

The solution could be a properly designed web application based on a topic map. The topic map, ISO13250, has many advantages that have been confirmed through lots of research (Melgar Estrada, 2011). Such advantages include the following:

- A topic map constitutes a structure independent of referenced occurrences, meaning any document. So one topic map can be used to navigate many resources.
- In comparison to the thesaurus there are no strict relationships e.g., associative, hierarchical and equivalence. Therefore, we can create any type of relationship we want. The increment number of associative relationships enables us to “describe more complex relationships”, but, at the same time, such an advanced net of

relationships is difficult to create. The relationships must be designed in a very well-considered way (Yi, 2008, p. 1899).

- In a topic map a few points of view on any subject could be presented. This is possible thanks to “scope”, which is “the context in which a name or an occurrence is assigned to a given topic, and the context in which topics are related through associations” (International Organization for Standardization, 2003, p. 3). This is the feature of topic maps that enables the maintenance of objectivity in libraries² (Sigel, 2003, p. 427).
- Topic maps can be visualised easily. This feature provides topic maps applications.
- Thanks to Public Subject Identifiers, which identify subjects of topics, several topic maps can be merged into one map.
- A mechanism called Topic Maps web services enables us to exchange topic map fragments. It may be used to create the net of cooperative web applications (Naito, 2009).
- L. M. Melgar Estrada states that, “interoperability between Topic Maps and RDF is high”; consequently, projects based on topic maps can be included harmoniously in the semantic web (Melgar Estrada, 2011, p. 56).

Until now topic maps were used in Poland within the European Exchange of Documents – Poland (EWD-P) system. This application serves to support, “management of elaboration of the official Polish standpoint concerning adjustment of the newly associated states law to the EU procedures and regulations” (Momotko, Nowicki & Strychowski, 2004). The only project using topic maps in MLA in Poland is being created in the National Museum in Warsaw and deals with art in Warsaw from 1901 to 1920 (Kopszak, 2011). This project will end with an implementation of this map on the museum’s website.

² In the Polish Code of Ethics for Librarians and Information Professionals there is a statement, that, “Basic moral values which define the professional mission of librarians and information professionals involve the protection of intellectual freedom, freedom of expression, freedom of access to knowledge, information and culture and the compliance with the principle of ideological, political and religious neutrality” (The Polish Librarians Association, 2006).

The issue of presenting subject headings as a topic map is the subject of Motomu Naito's work. It shows a way of such transformation, visualisation and examples of the practical use of subject headings (Naito, 2010). The way of visualising large hierarchies was also a subject of research. For example, Nihar Sheth and Qin Cai used a radial tree layout to visualise the MeSH data set, and developed the MeSH browser as enabled to display, not only hierarchical, but also non-hierarchical relationships on demand (Nihar & Cai, 2003). Ioannis Papadakis, Michalis Stefanidakis and Aikaterini Tzali (Papadakis, Stefanidakis & Tzali, 2008), developed a prototype web application based on AJAX technology and OWL. It visualised subject headings and provided user-friendly library resources navigation. The topic map was applied by Hak-Keun Kim, Teuk-Seob Song, Yoon-Chul Choy and Soon-Bum Lim as a navigation aid in a 3D virtual environment (Kim, Song, Choy & Lim, 2005). The visualisation of subject headings has also been introduced in the Library of Congress Subject Headings in order to browse vocabulary (Authorities & Vocabularies (Library of Congress)).

Among the inspirations for our project are the works of Motomu Naito (Naito, 2009; Naito, 2010), Myongho Yi (Yi, 2008) and Jadwiga Woźniak-Kasperek (Woźniak-Kasperek, 2011). The object of our project is to create a web application using appropriate visualisation that supports indexing and retrieval in the National Library of Poland. We are also considering the adaptation of the Wandora application. Five people from the JHP BN team as well as a number of IT professionals from the National Library of Poland will be involved in the project. It is expected to take approximately two years. It consists of three stages:

1. The creation of a topic map from the National Library of Poland Subject Headings.

In the National Library of Poland, the MARC standard is used to encode subject headings.

The first step is to transform the subject headings list. It should be determined how to divide headings and design relationships in the future topic map. This part rests upon designing ontology. We understand this term as a set of typing entities, that is, topics, names and associations (Melgar Estrada, 2011, p. 46). The clear solution is to keep the division according to the MARC fields – names: personal, corporate, meeting, topical, geographic, uniform title and genre/form term. Similarly,

relationships can be kept from those that already exist in a subject authority file: broader/narrower term, related term and use-use for. This solution was adopted by Motomu Naito (Naito, 2010). We will try to increase the number of relationships and change the division of subject headings to improve information retrieval in the catalogue of the National Library of Poland. It would be ideal to design a set of more specified relationships. At first we would choose one domain, for example, history, or even one historical period. At first sight it seems to be very difficult. One of the many difficult issues is the lack of relationships in JHP BN between a person and an historical event in which she or he participated, or between a person and an organization. The creation of these relationships manually is the only solution. Perhaps in the future this process will be automated. Once the ontology has been designed, we could then transform the subject headings into the XML format and add relationships. We will use the XML topic map syntax to represent the topic map based on our subject headings. This syntax has been designed to be “syntax for the interchange of Topic Maps” (International Organization for Standardization, 2007, p. 2).

2. Finding an appropriate way to visualise a topic map for library use and create a web application.

One of the ways of presenting topic maps in the Wandora application is a graph (Wandora – WandoraWiki). Benedicte Le Grand has noticed the following about this visualisation technique: “representation may become cluttered rapidly as the number of topics and associations increases” (Le Grand, 2003, p. 273). As I previously mentioned, the Polish subject headings now consist of approximately 69000 subject headings, so the graph created in Wandora based on this language will have about 69000 nodes. Such a huge topic map is difficult to browse. In the Wandora application, when we click on one node, another node opens and so on. After some time the user could see a completely illegible structure. The solution may be a limitation of displayed nodes at the same time or using focus + context visualisation techniques. It is also intended to explore other visualization methods including 3D

techniques. A good example of using 3D layout is the Tag Galaxy browser (Tag Galaxy, n.d.) (Figure 3).

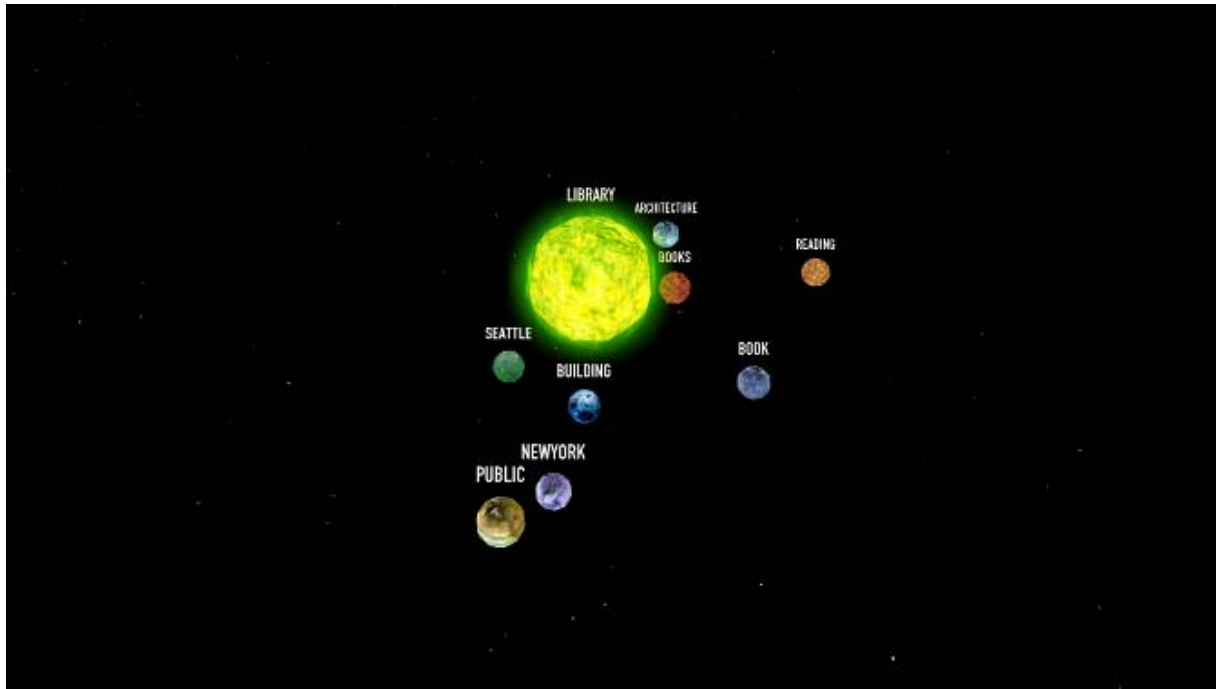


Figure 3: Tag Galaxy browser (<http://taggalaxy.de/>)

The important issue is how to represent the relationships in an appropriate way. The result of this stage will be a web application presenting the National Library of Poland Subject Headings in a visual way. This visualisation could be created in Flash or HTML5, CSS3 and Javascript.

3. Testing the prototype.

Tests will involve two groups of users:

- indexers
- end users

The first group will be tested on indexing with a web application based on a topic map. The participants of the test will be indexers from the National Library of Poland and one of the big public libraries, for example, from Warsaw. The tests will show how the new tool influences speed and quality of indexing, and will involve indexing, observations, and interviews about traditionally organised subject headings and the

new tool. We will compare how fast and how precisely indexers will index resources, which will have previously been indexed by members of the JHP BN team. Comparing these two approaches will demonstrate whether a visualised topic map is advantageous for the National Library of Poland. Indexing with the new tool will differ from the JHP BN method. There will be no strict subject heading strings, but some of the rules existing in the old tool will be transferred to the new one. It could be advantageous to establish schemas, which would help in assigning specified classes of terms to specified types of resources. For example, for the exhibition catalogue it could be terms for form, exhibits, topic and place of an exhibition and an institution organising it (Włodarczyk, 2012).

The second group will be tested on information retrieval in the traditional OPAC of the National Library of Poland and on browsing in the new environment based on a topic map. Their task will be to find resources related to a specified topic. Apart from research on visualised subject headings, we want to test searching based on the created associative relationships. The scientific techniques will include observations and interviews.

We believe that the research will show the differences in indexing and browsing between the traditional OPAC and a web application based on a visualised topic map. It will also show the best way of visualising subject headings for library use and whether it is advantageous to represent more specified relationships. Although until now, no visual search engine has gained broader popularity, it still seems to be a promising field of research. Further study seems to be necessary and could be very profitable for future libraries. One of the features of topic maps is also the capability of merging several maps or exchanging information, features that also could be profitable in the context of library cooperation.

This is only the beginning of the project and there are many problems to be solved. Nonetheless, research on topic maps and information visualisation are quite advanced and we can benefit from them. We hope that the project will show whether visualised topic maps are suitable for our needs.

References:

Antell, K. & Huang, J. (2008). Subject Searching Success Transaction Logs, Patron Perceptions, and Implications for Library Instruction. *Reference & User Services Quarterly*, 48 (1), 68–76.

Authorities & Vocabularies (Library of Congress)]. Retrieved from: <http://id.loc.gov/search/>

Condit Fagan, J. (2010). Usability Studies of faceted browsing: a literature review. *Information Technology and Libraries*, 29 (2), 58-66.

Drabenstott, K. M., Simcox, S. & Fenton, E. G. (1999). End-user understanding of subject headings in library catalogs. *Library Resources & Technical Services*, 43 (3), 140-160.

Franz, L, Powell, J., Jude, S., & Drabenstott, K. M. (1994). End user understanding of subdivided subject headings. *Library Resources & Technical Services*, 38 (3), 213-226.

Highwire. Retrieved from <http://highwire.stanford.edu>

International Organization for Standardization. (2003). ISO/IEC 13250 Information Technology—SGML Applications—Topic Maps. Geneva, Switzerland: Author.

International Organization for Standardization. (2007). ISO/IEC 13250-3 Information Technology —Topic Maps – XML syntax. Geneva, Switzerland: Author.

Kim, H.-K., Song, T.-S., Choy, Y.-Ch. & Lim, S.-B. (2005). Guided navigation techniques for 3D virtual environment based on topic map. In O. Gervasi, M. L. Gavrilova, V. Kumar, A. Laganà & H. P. Lee (2005), *ICCSA'05 Proceedings of the 2005 international conference on Computational Science and its Applications - Volume Part I* (pp. 847-856). Berlin, Germany: Springer-Verlag.

Klenczon, W. (2011). Język Haseł Przedmiotowych Biblioteki Narodowej (National Library of Poland Subject Headings) – from Card Catalogues to Digital Library: Some Questions about the Future of a Local Subject Headings System in the Changing Word of Information Retrieval. In P. Landry [et al.], *Subject Access: Preparing for the Future* (pp. 169-179). Berlin, Germany: Walter de Gruyter GmbH & Co.

Klenczon, W., & Stolarczyk, A. (2007). Subject Headings of the Polish National Library (JHP BN). *Polish Libraries Today*, 7, 60-64.

Kopszak, P. (2011, January 11). Sztuka w Warszawie w latach 1901-1920 – szkic mapy topik [Video file]. Retrieved from <http://wiedzaiedukacja.eu/archives/47948>

Le Grand, B. (2003). Topic map visualization. In J. Park & S. Hunting (Eds.), *XML Topic Maps: Creating and Using Topic Maps for the Web* (pp. 267-282). Boston, United States: Addison-Wesley Professional.

Library of Congress Subject Headings Pre- vs. Post-Coordination and Related Issues. (2007). Retrieved from http://www.loc.gov/catdir/cpsd/pre_vs_post.pdf

Melgar Estrada, L. M. (2011). Topic Maps from a Knowledge Organization Perspective. *Knowledge Organization*, 38 (1), 43-61.

Momotko, M., Nowicki, B., & Strychowski J. (2004). The EWD-P system. Polish government - European Commission Interoperability Achieved, accepted to The 4th European Conference on E-Government, 17-18 June 2004, Trinity College Dublin, Ireland, Retrieved from <http://www.rodan.pl/editor-cm-web-portlet/content/Nzl1.pdf>

Naito, M. (2010). Subject Headings make information to be topic maps. In L. Maicher & L.M. Garshol (Eds.), *Information Wants to be a Topic Map Sixth International Conference on Topic Maps Research and Applications, TMRA 2010, Leipzig, Germany, September 29 – October 01, 2010, Revised Selected Papers* (pp. 43-51), Leipzig, Germany: Leipziger Informatik-Verbund. Retrieved from http://tmra.de/2010/documents/TMRA2010_proceedings.pdf

Naito, M. (2009) Topic Maps Web Service: Case Examples and General Structure. In L. Maicher & L.M. Garshol (Eds.), *Linked Topic Maps. Fifth International Conference on Topic Maps Research and Applications, TMRA 2009, Leipzig, Germany, November 12-13, 2009, Revised Selected Papers* (pp. 179-184). Leipzig, Germany: Leipziger Informatik-Verbund. Retrieved from http://tmra.de/2009/documents/TMRA2009_Proceedings.pdf

Nihar, S. & Cai, Q. (2003) *Visualizing mesh data-set using radial tree layout*. Retrieved from <http://iv.slis.indiana.edu/sw/papers/radialtree.pdf>

Palsky, G. (1999). The debate on the standardization of statistical maps and diagrams (1857-1901). elements for the history of graphical language. *Cybergeo: European Journal of Geography*. doi: 10.4000/cybergeo.148

Papadakis, I., Stefanidakis, M., & Tzali, A. (2008). Visualizing OPAC subject headings. *Library Hi Tech*, 26 (1), 19-23. doi: 10.1108/07378830810857762

Polish Librarians Association (2006). *Code of Ethics for Librarians and Information Professionals*. Retrieved from http://archive.ifla.org/faife/ethics/poland_code_of_ethics.htm

Sigel, A. (2003). Topic maps in Knowledge Organization. In J. Park & S. Hunting (Eds.), *XML Topic Maps: Creating and Using Topic Maps for the Web* (pp. 383-476). Boston, United States: Addison-Wesley Professional.

Tag Galaxy. Retrieved from: <http://taggalaxy.de/>

Wandora – WandoraWiki. Retrieved from http://www.wandora.org/wandora/wiki/index.php?title=Main_Page

Ware, C. (2004). *Information visualization : perception for design*. San Francisco, United States: Morgan Kaufman.

Werla, M. (2011, October 4). Użytkownicy urządzeń mobilnych w Wielkopolskiej Bibliotece Cyfrowej [Web blog]. Retrieved from <http://dl.psnc.pl/2011/10/04/uzytkownicy-urazden-mobilnych-w-wielkopolskiej-bibliotece-cyfrowej/>

Windows Developer Preview: Windows 8 guide. (2011). Retrieved from <http://windows.microsoft.com/pl-PL/windows-8/preview>

Włodarczyk, B. (2012). Mapa tematów jako system reprezentacji wiedzy. *Zagadnienia Informacji Naukowej*, 2 (98), (in press).

WorldCat Identities Network. Retrieved from <http://experimental.worldcat.org/idnetwork>

Woźniak-Kasperek, J. (2011). *Wiedza i język informacyjny w paradygmacie sieciowym*. Warszawa, Poland: Wydawnictwo SBP.

Yi, M. (2008). Information organization and retrieval using a topic maps-based ontology. *Journal Of The American Society For Information Science And Technology*, 59 (12), 1898-1911.