



Библиотека тематической карты = лучшая библиотека: введение к проекту “Национальная библиотека Польши”

Бартоломей Влодарчик
(Bartłomiej Włodarczyk)
Национальная библиотека Польши
Библиографический институт
Варшава, Польша
E-mail: b.wlodarczyk[at]bn.org.pl

Перевод на русский язык:

Елена Загорская

(Российская национальная библиотека, Санкт-Петербург)

Session:

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

Аннотация

Предметные рубрики национальной библиотеки Польши (JНР BN) используются в Польской национальной библиографии, каталоге национальной библиотеки, а также многих других польских библиотек: публичных, педагогических, церковных и академических. Тематические поиски в нашем библиотечном каталоге все еще составляют значительную часть всех поисков, но понимание и использование предметных рубрик представляет много проблем, и не только для конечных пользователей, но также и для многих библиотекарей. Решением может быть надлежащим образом оформленное web-приложение, основанное на тематической карте, использующее соответствующую визуализацию, которая поддерживает индексирование и поиск в Национальной библиотеке Польши.

Введение

С 1957 года Национальная библиотека Польши развивает свой собственный язык индексирования – Предметные рубрики Национальной библиотеки Польши (JНР BN), которые используются в Польской национальной библиографии, каталоге национальной библиотеки, а также многих других польских библиотек: публичных, педагогических, церковных и академических (Klenczon & Stolarczyk, 2007; Klenczon, 2011). Это предкоординированная система, построенная в соответствии с действующими стандартами. Авторитетный файл предметных рубрик поддерживается в формате MARC21, включает около 69 000 рубрик (предпочтительных терминов), 87 000 вариантных терминов (не принятых терминов), и 117 000 связей между терминами. Тематический (предметный) поиск в каталоге нашей библиотеки по-прежнему составляет значительное число всех поисковых запросов - статистика логов транзакций показывает, что они составляют почти 20% - и доля отказов в получении информации не высока. Однако, изучение вопроса показало, что понимание и использование предметных рубрик имеет множество проблем, причем не только для конечных пользователей, но и для многих библиотекарей (Franz, Powell, Jude & Drabenstott, 1994; Drabenstott, Simcox & Fenton, 1999; Antell & Huang, 2008). Одной из важнейших проблем является

улучшение качества и скорости индексирования. Мы не хотим конкурировать с поисковыми машинами, но скорость предоставления информации имеет важное значение для конечных пользователей. Мы также хотим, сократить время, необходимое для обучения новых индексаторов. Это связано с усложнением предкоординированной грамматики. Другая важная проблема заключается в оказании помощи индексаторам при выборе лучшего термина для включения в сложную систему отношений при индексировании конкретной книги или иных ресурсов¹ (Library of Congress Subject Headings Pre- vs. Post-Coordination, 2007). Наши наблюдения и сведения, полученные от библиотекарей из библиотек, которые используют JHP BN, подтверждают эти выводы. Мы знаем, что требуется более эффективный способ поиска информации, но мы не хотим терять высокое качество контролируемого словаря.

Первым шагом к улучшению предметного доступа к информации была попытка упростить синтаксис и сократить предметные рубрики: формальные подзаголовки были выведены из цепочек предметных рубрик, позже они были трансформированы в индекс (указатель, список) терминов жанр/форма. Это был первый шаг к фасетизации. Фасетизация, как полагают, имеет много преимуществ (Condit Fagan, 2010), но, ни сегодня, ни в будущем, это не решает одну из важнейших проблем, стоящих перед библиотеками - каким образом представлять библиотечные ресурсы в сетевой среде. Библиотека по-прежнему предлагает текстовое представление каталогов, хотя соответствующим образом визуализированная информация упрощает понимание и навигацию. Мы будем продолжать работу по упрощению JHP BN. Наша цель заключается в разработке JHP BN словаря и упрощении его грамматики и правил использования. Мы также хотим провести исследования, которые позволят в будущем улучшить технологию индексирования в Национальной библиотеке Польши и других польских библиотеках. Мы также хотим попытаться связать (соотнести) польские предметные рубрики с нотацией Универсальной десятичной классификации.

Важность визуализации была обоснована в когнитивной теории систем (Ware, 2004). Уже в начале 19-го века, шотландский инженер и политэкономист William Playfair сказал, что "в отношении числа и доли, лучший способ включить воображение - поговорить с глазами" (Palsky, 1999). Его мнение касается только чисел, но это утверждение также справедливо и в отношении понятий и слов. Это может быть ключом для библиотек при визуализации информации. Это общая фраза - «одна картинка стоит тысячи слов». Традиционно организованный библиотечный ОПАС, представляющий перечень предметных рубрик как линейный текст (рис. 1), не может поддерживать эффективное использование предметных точек доступа и связей между ними.

Для того, чтобы явно раскрыть информацию и повысить ее использование, кажется, требуется улучшить визуализацию информации. Авторитетный файл предметных рубрик может служить базой для создания нового инструмента эффективного и удобного раскрытия информации и навигации в каталоге для конечного пользователя.

¹ Эти проблемы существуют в нашей ежедневной работе

Библиотеки нуждаются в разработке таких способов визуализации, которые подходят не только для настольных компьютеров, но и для мобильных устройств с сенсорными экранами. Сенсорные экраны уже доступны в настольных компьютерах, новые версии наиболее распространенных операционных систем Windows 8 и Internet Explorer 10 будут иметь сенсорную клавиатуру и поддержку навигации жестами (Windows Developer Preview: Windows 8 guide, 2011, p. 8-9, 13-14). Мы ожидаем, что число пользователей библиотек, имеющих доступ к библиотечным каталогам с мобильных устройств, будет расти. В Польше уже наблюдается начало этих трендов. До сих пор лишь небольшое число пользователей используют этот метод. В Электронной библиотеке Великопольской области число таких пользователей составляет лишь 0,19% от всех пользователей, независимо от используемых устройств или операционной системы (Werla, 2011). Визуализация может поддерживать более естественный способ навигации, и это одна из причин, почему мы должны работать над этим вопросом.

Еще одной проблемой в Национальной библиотеке Польши является недостаточное использование связей построенных коллективом BN JHP. Установление иерархических и ассоциативных связей требует очень много времени. Эти связи имеют ценность, и мы должны увеличить их использование. Мы также хотим увеличить количество связей для улучшения поиска информации.

Решением этой задачи может быть надлежащим образом оформленное веб-приложение, основанное на тематической карте. Тематическая карта, отвечающая требованиям ISO13250, имеет много преимуществ, которые были подтверждены большим количеством исследований (Melgar Estrada, 2011). Эти преимущества включают следующее:

- Тематическая карта представляет собой структуру, независимую от местонахождения ссылок, то есть любой документ. Поэтому тематическая карта может быть использована для навигации по многим ресурсам.

- По сравнению с тезаурусами нет жестко определяемых типов связей, например, ассоциативных, иерархических и эквивалентности. Таким образом, мы можем создать любой тип связи, какой мы хотим. Прирост числа ассоциативных связей позволяет "описывать более сложные отношения", но, в то же время, такую продвинутую сеть связей трудно создать. Связи должны создаваться очень точным и согласованным способом (Yi, 2008, p. 1899).

- В тематической карте можно представить несколько точек зрения по любому вопросу. Это стало возможным благодаря "score", которая является "контекстом, в котором имя или местоположение присваивается данной теме, и контекстом, в котором темы ассоциативно связаны" (International Organization for Standardization, 2003, p. 3). Эта особенность тематической карты позволяет поддерживать объективность в библиотеках² (Sigel, 2003, p. 427).

- Тематические карты легко визуализируются. Эта функция обеспечивается приложениями тематических карт.

- Благодаря Public Subject Identifiers, которые идентифицируют (определяют) предметные темы, несколько тематических карт могут объединяться в одну карту.

- Механизм, называемый веб-сервисом тематических карт, позволяет обмениваться фрагментами тематических карт. Он может быть использован для создания сетевой кооперации веб-приложений (Naito, 2009).

- L.M. Melgar Estrada утверждает, что "взаимодействие между тематическими картами и RDF является высоким", следовательно, проекты, основанные на тематических

² В польском Кодексе этики для библиотекарей и информационных специалистов есть положение о том, что "Основные моральные ценности, которые определяют профессиональную миссию библиотекарей и информационных специалистов, связаны с защитой интеллектуальной свободы, свободы выражения мнений, свободы доступа к знаниям, информации и культуре и соблюдением принципа идеологического, политического и религиозного нейтралитета". (Польская Ассоциация Библиотекарей, 2006).

картах могут быть гармонично включены в семантический веб (Melgar Estrada, 2011, p. 56.).

До сих пор тематические карты использовались в Польше в рамках Европейской системы обмена документами - European Exchange of Documents – Poland (EWD-P). Это сервис-приложение служит для поддержки "управления разработкой официальной польской точки зрения, касающейся процедур и правил корректировки законодательства новых присоединившихся к ЕС государств» (Momotko, Nowicki & Strychowski, 2004). Единственный проект с использованием тематических карт в MLA в Польше создается в Национальном музее в Варшаве и касается искусства в Варшаве с 1901 по 1920 гг. (Kopszak, 2011). Этот проект завершится внедрением этой карты на сайте музея.

Представление предметных рубрик в виде тематических карт является предметом исследования Мотому Найто (Motomu Naito). В этом исследовании предложен способ подобного преобразования, визуализации и примеры практического использования предметных рубрик (Naito, 2010). Способ визуализации больших иерархий также был предметом исследования. Например, Nihar Sheth и Qin Cai (Цинь Кай) использовали радиальный макет дерева для визуализации комплексов данных MeSH, они также разработали MeSH-браузер для просмотров не только иерархических, но и не-иерархических связей, соответствующих запросу (Nihar & Cai, 2003). Ioannis Papadakis, Michalis Stefanidakis и Aikaterini Tzali (Papadakis, Stefanidakis & Tzali, 2008), разработали прототип веб-приложений на основе технологии AJAX и OWL. Они визуализировали предметные рубрики и обеспечили удобную навигацию по библиотечным ресурсам. Тематическая карта была применена Hak-Keun Kim, Teuk-Seob Song, Yoon-Chul Choy и Soon-Bum Lim, в качестве средства навигации в 3D виртуальной среде (Kim, Song, Choy & Lim, 2005). Визуализация предметных рубрик также была использована в предметных рубриках Библиотеки Конгресса для просмотра словаря (Authorities & Vocabularies. (Library of Congress)).

Наш проект был инициирован под влиянием исследований Мотому Найто (Naito, 2009; Naito, 2010), Митонго Йи (Yi, 2008) и Ядвиги Возняк-Касперек (Woźniak-Kasperek, 2011). Целью нашего проекта является создание веб-приложений, использующих визуализацию, которая поддерживает индексирование и поиск в Национальной библиотеке Польши. Мы также изучаем возможность адаптации приложения Wandoga. В проекте будут участвовать пять человек из команды JHP BN, а также ряд IT-специалистов из Национальной библиотеки Польши. Ожидается, что работа займет около двух лет. Она состоит из трех этапов:

1. Создание тематической карты из предметных рубрик Национальной библиотеки Польши. В Национальной библиотеке Польши для кодирования предметных рубрик используется стандарт MARC. Первый шаг состоит в преобразовании списка предметных рубрик. Следует определить, как разделить рубрики и дизайн связей в будущей тематической карте. Эта часть основывается на структуре онтологии. Под этим термином мы понимаем набор типовых сущностей (объектов), таких как темы, имена/наименования и ассоциации (Melgar Estrada, 2011, p. 46). Четко решено сохранить разделение в соответствии с полями MARC-формата – имена/наименования: лиц, организаций, совещаний, тематические, географические названия, унифицированные заглавия и термин жанра/формы. Кроме того, связи могут быть установлены с уже существующими в авторитетном файле предметных рубрик связями: широкий/узкий термин, связанный термин и термин для использования-использования. Это решение было принято Мотому Найто (Naito, 2010). Мы пытаемся увеличить количество связей и изменить разделение предметных рубрик для улучшения поиска

информации в каталоге Национальной библиотеки Польши. Было бы идеально разработать комплект более специфических связей. Сначала мы бы выбрали одну область, например, историю, или хотя бы один исторический период. На первый взгляд это кажется очень трудным. Одной из многих проблем является отсутствие в JHP VN связей между человеком и историческим событием, в котором он или она принимали участие, или между человеком и организацией. Создание этих связей вручную является единственным решением. Возможно, в будущем этот процесс будет автоматизирован. После того как онтология будет разработана, мы могли бы преобразовать предметные рубрики в XML-формат и добавить связи. Мы будем использовать синтаксис тематических карт XML для того, чтобы представить тематическую карту, основанную на наших предметных рубриках. Этот синтаксис был разработан как "синтаксис для обмена тематическими картами". (International Organization for Standardization, 2007, p.2).

2. Найти подходящий для использования в библиотеке способ визуализации тематической карты и создать веб-приложения. Один из способов представления тематических карт в приложении Wandora представляет собой график (Wandora – WandoraWiki). Об этой технологии визуализации Бенедикт Ле Гранд (Benedicte Le Grand) отметил следующее: "изображения (образы) могут быстро превратиться в хаос по мере увеличения количества тем и ассоциаций" (Le Grand, 2003, p. 273). Как я уже упоминал ранее, польские предметные рубрики в настоящее время включают примерно 69 000 предметных рубрик, поэтому график, разработанный в Wandora, основан на этом языке и будет иметь около 69 000 точек пересечения (узлов). Такую огромную тематическую карту сложно просматривать. При использовании Wandora, когда мы кликаем на одном узле (пересечении), открывается другой узел и так далее. Через некоторое время пользователь может увидеть совершенно неудобочитаемую структуру. Решением может быть ограничение количества показываемых узлов, при одновременном использовании фокуса + контекстного метода визуализации. Предполагается также рассмотреть другие методы визуализации, включая 3D-технологии. Хорошим примером использования 3D-макета является браузер тегов Galaxy (Tag Galaxy, n.d) (рис. 3).



Рисунок 3. Браузер тэгов Galaxy (<http://taggalaxy.de/>)

Важной проблемой является подходящий способ представления связей. Результатом данного этапа будет веб-приложение, представляющее Польские предметные рубрики визуальным способом. Эта визуализация может быть реализована в Flash или HTML5, CSS3 и JavaScript.

3. Тестирование прототипа. К тестированию привлекались две группы пользователей:

индексаторы
конечные пользователи

Первая группа будет проверять процесс индексирования с использованием веб-приложения, основанного на тематической карте. Участниками тестирования будут индексаторы из Национальной библиотеки Польши и одной из самых больших публичных библиотек, например, из Варшавы. Тесты покажут, как новый инструмент влияет на скорость и качество индексирования, и будут включать индексирование, наблюдение и интервью о традиционно организованных предметных рубриках и новых инструментах. Мы будем сравнивать, как быстро и насколько точно индексаторы будут индексировать ресурсы, которые ранее были проиндексированы членами группы JHP BN. Сравнение этих двух подходов продемонстрирует, эффективна ли визуализированная тематическая карта для Национальной библиотеки Польши. Индексирование с помощью нового инструмента будет отличаться от методики JHP BN. Там не будет строгих цепочек предметных рубрик, но некоторые правила, существующие в старой технологии, будут действовать и в новой. Это может быть полезно - создавать схемы, которые помогут при использовании указанных классов терминов для индексирования определенных видов ресурсов. Например, для каталога выставки это могут быть термины формы, экспоната, темы и места проведения выставки, а также учреждения, организовавшего ее (Włodarczyk, 2012).

Вторая группа будет проверять поиск информации в традиционном ОРАС Национальной библиотеки Польши и просмотр в новых условиях на базе тематической карты. Их задачей будет найти ресурсы, связанные с указанной темой. Помимо исследований по визуализации предметных рубрик, мы хотим проверить поиск на основе созданных ассоциативных связей. Методы научного исследования будут включать в себя наблюдения и интервью.

Мы считаем, что исследование покажет различия в индексировании и поиске информации между традиционным ОРАС и веб-приложениями на основе визуализации тематической карты. Оно также покажет лучший способ визуализации предметных рубрик для использования библиотекой и выявит эффективность более точного представления определенных связей. Хотя до сих пор визуальный способ поиска не приобрел широкую популярность, он представляется нам перспективным направлением исследований. Дальнейшие исследования представляются необходимыми и могут быть очень полезными для будущего библиотек. Одной из особенностей тематической карты является также возможность объединения нескольких карт и обмена информацией, возможности, полезные в контексте сотрудничества библиотек.

Это только начало проекта, и есть много проблем, которые предстоит решить. Тем не менее, исследования тематических карт и визуализации информации достаточно популярны, и мы можем извлечь из них пользу. Мы надеемся, что проект покажет, подходят ли для наших нужд визуализированные карты тем.

Ссылки:

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/cpso/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/Nz11.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-urzedzenmobilnych-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 Informatyki 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.