

III. KRONIKA

SPRAWOZDANIE Z KONFERENCJI BEYOND LIBRARIES – SUBJECT METADATA IN THE DIGITAL ENVIRONMENT AND SEMANTIC WEB

Tallinn, 17-18 sierpnia 2012 r.

W dniach 17-18 sierpnia br. odbyła się w Tallinnie konferencja *Beyond libraries – subject metadata in the digital environment and semantic web*, zorganizowana przez IFLA Cataloguing & Indexing Section (Sekcję Opracowania Rzeczowego) jako impreza towarzysząca tegorocznemu kongresowi IFLA w Helsinkach. Konferencję zorganizowano we współpracy z Biblioteką Narodową Estonii i estońskim stowarzyszeniem bibliotekarzy. Wzięło w niej udział około 120 osób z 22 krajów, w tym duża grupa bibliotekarzy z krajów skandynawskich, Estonii i Łotwy oraz równie liczne grono specjalistów amerykańskich i kanadyjskich. Zaprezentowano 13 referatów dotyczących teorii i praktyki opracowania rzeczowego, m.in. nowej generacji kartotek wzorcowych, łączenia danych bibliotecznych z metadanymi innych zasobów, tworzenia punktów dostępu do zasobów cyfrowych, społecznego tagowania.

Sesja 1 poświęcona była potrzebom użytkowników oraz projektowaniu punktów dostępu przedmiotowego w środowisku sieciowym.

W referacie *Using authorities to improve subject searches* Edward T. O'Neill z OCLC przedstawił zalety wykorzystywania kartotek wzorcowych jako dodatkowych punktów dostępu przedmiotowego. Podkreślił fakt, że chociaż mogą one znacznie poprawić wyszukiwanie poprzez zwiększenie liczby punktów dostępu, rzadko są integralną częścią procesu wyszukiwania informacji, zwłaszcza przez tak zwanych „użytkowników końcowych”. Zaprezentował prototyp narzędzia wyszukiwawczego o nazwie searchFAST, które zostało zaprojektowane do zbadania możliwości używania kartoteki wzorcowej jako indeksu do rekordów bibliograficznych. Narzędzie searchFAST wykorzystuje FAST (Faceted Application of Subject Terminology) jako indeks do bazy danych bibliograficznych WorldCat, utrzymywanej przez OCLC. Metodologia searchFAST nie zastępuje, a uzupełnia istniejący dostęp do World Cat. Baza bibliograficzna jest przeszukiwana pośrednio: najpierw przeszukiwana jest kartoteka wzorcowa w celu identyfikacji odpowiednich haseł przedmiotowych, następnie hasła wzorcowe są używane do uzyskiwania odpowiednich rekordów bibliograficznych. Prototyp pokazuje skuteczność i użyteczność wykorzystania kartoteki wzorcowej jako indeksu. Przeszukiwanie kartoteki wzorcowej zwiększa kontrolę wzorcową nad dziełem poprzez zwiększenie liczby punktów dostępu przy jednoczesnym wsparciu prostego interfejsu, przyjaznego dla odbiorców, użytkowników końcowych.

W prezentacji *Subject metadata development for digital resources in Latvia* przedstawiciele Biblioteki Narodowej Łotwy opowiedzieli o doświadczeniach związanych z wprowadzeniem i wykorzystaniem Library of Congress Subject Headings (LCSH) do katalogowania rzeczowego zbiorów biblioteki. Od roku 2000 powstało ponad 34000 haseł przedmiotowych używanych do wszelkiego rodzaju zasobów, w tym obiektów cyfrowych. Do katalogowania rzeczowego zasobów cyfrowych Biblioteka Narodowa Łotwy stosuje system podobny do FAST. Hasła przedmiotowe mają uproszczoną składnię, hasła rozwinięte podzielone są na osobne punkty dostępu. Jednym z kamieni milowych w rozwoju katalogowania rzeczowego zasobów cyfrowych na Łotwie był projekt *In Search of Lost Latvia* z 2010 roku, dotyczący instytucji pamięci i osób, w którym użytkownicy mogli dodawać własne słowa kluczowe i komentarze. Wyzwaniem było użycie LCSH do indeksowania tych cyfrowych obiektów ze względu na poszukiwania zagadnień o wąskim zakresie tematycznym.

W referacie Sirje Nilbe z Biblioteki Narodowej Estonii oraz Tiitu Tarkpea z Uniwersytetu w Tartu zaprezentowano Estoński Tezaurus Tematyczny (EMS), który funkcjonuje jako główne narzędzie indeksowania w cyfrowych bazach danych bibliotek estońskich. Powstał on w wyniku połączenia dwóch uniwersalnych tezaurusów funkcjonujących w środowisku bibliotekarskim Estonii. W referacie przeanalizowano jego wykorzystanie w tradycyjnych zasobach, takich jak katalogi czy bazy danych bibliograficznych, a także w nowszych typach zasobów: archiwach cyfrowych i repozytoriach instytucjonalnych. Konkluzją było stwierdzenie, że dostępność uniwersalnego tezaury ułatwia ponowne użycie rekordów. Na koniec podkreślono, że trendem rozwojowym EMS powinna być zgodność ze współczesnymi standardami Semantic Web, takimi jak SKOS czy Open Linked Data.

Sesję 2 zatytułowaną *Library metadata beyond libraries and text-based material* rozpoczął referat Wandy Klenczon i Pawła Rygla z Biblioteki Narodowej, zatytułowany *Librarian cornered by images or How to index visual resources*. Poruszał on problemy opisu dokumentów ikonograficznych, a zwłaszcza tworzenia punktów dostępu do treści obrazów. Problem ten, do niedawna w bibliotekach traktowany był jako marginalny. Biblioteki zwykle posiadają zbiory graficzne i fotograficzne, często znaczne i wartościowe, ale udostępniały je wyłącznie wąskiemu gronu specjalistów i nie zapewniały informacji katalogowej. Digitalizacja i publiczny dostęp do tych obiektów w bibliotekach i archiwach cyfrowych uczyniły problem opisu tych zasobów niezwykle ważnym. Przedstawiono, bazując na podstawowych w tej dziedzinie schematach Panofsky'ego i Shatford, problemy indeksowania treści dokumentów ikonograficznych w katalogach bibliotecznych. Zaprezentowano model opisu, wynikający zarówno z analizy międzynarodowych standardów i praktyk, jak i z lokalnych doświadczeń. Wystąpienie spotkało się z dużym zainteresowaniem uczestników konferencji i sprowokowało wiele pytań, co potwierdza nasze obserwacje, że biblioteki nie wypracowały jeszcze metod opisu tych zasobów i utwierdza nas w przekonaniu, że temat jest ważny i wart pogłębienia.

Przedmiotem kolejnego zespołowego wystąpienia był zrealizowany przez bibliotekę University of Wisconsin-Milwaukee amerykańsko-chiński projekt udostępnienia w bibliotece cyfrowej wizerunków i opisów cennych obiektów

chińskiego dziedzictwa kulturowego – kolekcji zwojów kaligraficznych i malowanych wachlarzy (Krystyna K. Matusiak, Ling Meng, Ewa Barczyk, Chia-Jung Shih, *Multilingual Indexing of Cultural Heritage Materials: Building the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings*). Do prezentacji obiektów, zarazem tekstowych i obrazowych, zaprojektowano dwujęzyczny model biblioteki cyfrowej: zastosowano w pełni dwujęzyczny, rozszerzony zestaw elementów Dublin Core, uwzględniając możliwość ukazania kontekstu kulturowego (np. biogram artysty, oryginalny i tłumaczony tekst literacki i teksty uzupełniające) – obie wersje językowe są równoprawne i widoczne w tym samym rekordzie. Wartości metadanych pobierano z dostępnych zasobów kontrolowanego słownictwa, m.in. LC TGM, LCSH, kartoteki wzorcowej nazw osobowych LC, tezaursów Getty Institute: TGN i AAT. Ze względu na niewielki rozmiar zasobu (ok. 60 obiektów) możliwe było nie tylko mapowanie i automatyczny przekład metadanych, ale i ich weryfikacja, a także przekład niektórych tekstów przez specjalistów-sinologów. W pełni dwujęzyczna specjalistyczna biblioteka cyfrowa jest niezwykle inspirującym źródłem informacji dla badaczy z obu kręgów kulturowych.

Wystąpienie Harriet Aagaard z Biblioteki Narodowej Szwecji oraz Elisabet Viktorsson z agencji bibliograficznej Bibliotekstjänst (BTJ) poświęcone było opracowaniu rzeczowemu literatury pięknej (*Subject Headings for Fiction in Sweden – a Cooperative Development*). Omówiono w nim działalność Swedish Library Association's Fiction Indexing Committee, powołanego w 2005 r. w celu wsparcia idei stworzenia standardu opisu rzeczowego literatury pięknej. W skład komitetu weszli przedstawiciele bibliotek różnego typu oraz specjaliści z agencji bibliograficznej Bibliotekstjänst (BTJ). Dzięki temu standard może być użyteczny dla różnego typu instytucji. Jest to szczególnie przydatne i szeroko wykorzystywane w bibliotekach publicznych i szkolnych. Inną zaletą jest praktyczny wpływ instytucji na listy haseł przedmiotowych, dzięki czemu zaspokajane są potrzeby użytkowników. Przyczyniło się to również do budowy standardu, który jest użyteczny w branży księgarskiej. W rezultacie powstał standard, który działa na różnych platformach, oferując tym samym nowe możliwości eksploracji i eksploatacji danych bibliograficznych.

Sesję 3, której mottem były nowe i tradycyjne metody zapewniania dostępu przedmiotowego, rozpoczęło wystąpienie Marcy Bidney i Kevina Clair, przedstawicielei Pennsylvania State University, którzy zaprezentowali ciekawy referat zatytułowany *Harnessing the Geospatial Semantic Web: Toward Place-Based Information Organization and Access*. Podstawowym przeznaczeniem geoprzestrzennej sieci semantycznej (geospatial Semantic Web) było do tej pory tworzenie mashup'ów map, projektów współpracy przy tworzeniu map oraz innych działań badawczych. Siłę geoprzestrzennej sieci semantycznej można również wykorzystać dla rozwoju punktów dostępu opartych na miejscach do dalszego wykorzystania zbiorów informacji cyfrowych i drukowanych. Utworzenie interfejsu dla geograficznego wyszukiwania zbiorów informacyjnych pozwala użytkownikom na wyszukiwanie według lokalizacji. Podstawowe zasady danych powiązanych (linked data), opisujące jednostki wykorzystujące unikalne identyfikatory i zapewniające połączenia z obiektami powiązanymi, wiążą się z chęcią łączenia przez biblioteki własnych zasobów cyfrowych

z pokrewnymi materiałami posiadanymi przez inne instytucje kultury publikujące treści w Internecie. Referat zawierał też przegląd zasad dotyczących danych powiązanych (linked data), omówienie korzyści i wyzwań związanych z dostarczaniem informacji geograficznej w rekordach metadanych oraz podanie przykładów na to, jak wartościowe dla użytkowników są wyszukiwania oparte na lokalizacji i jakie możliwości badawcze daje to na przyszłość.

Oksana Zavalina (University of North Texas) poświęciła swoje wystąpienie opisom kolekcji w bibliotekach cyfrowych (*Exploring Mutual Complementarity of Free-Text and Controlled-Vocabulary Collection-Level Subject Metadata in Large Scale Digital Libraries: a Comparative Analysis*). Przedmiotem analizy były metadane opisowe kolekcji zamieszczanych w trzech wielkich bibliotekach cyfrowych, agregujących zasoby dziedzictwa kulturowego: American Memory (LC), Opening History (University of Illinois-Champaign) i The European Library. Biblioteki cyfrowe rzadko udostępniają metadane całych kolekcji (powszechnie natomiast występują one w archiwach cyfrowych), zasoby biblioteczne zwykle są opisywane wyłącznie na poziomie poszczególnych obiektów. Celem analizy było stwierdzenie, czy metadane kolekcji podawane swobodnym tekstem (w elemencie Description/Opis oraz Object/Obiekt) i jako kontrolowane punkty dostępu rzeczowego (elementy Subject/Temat, Coverage/Zasięg, Date/Data, Type/Typ) uzupełniają się wzajemnie, czy też powielają informacje. Autorka stwierdziła wysoki poziom komplementarności i spójność obu rodzajów metadanych, a nieznaczną jedynie redundancję. Brak niestety kodeksu dobrych praktyk w tym zakresie – często kolekcje są opisywane jedynie wolnym tekstem. Dla ukazania pełnego kontekstu treści zasobu wskazane jest posługiwanie się oboma rodzajami metadanych, w tym kierunku zmierzają zalecenia przygotowywane przez IFLA Working Group razem z World Digital Library Project (Guidelines for Digital Libraries).

W następnym wystąpieniu Ulrike Junger (Deutsche Nationalbibliothek) omówiła założenia i podsumowanie pierwszego etapu bardzo interesującego projektu automatycznego indeksowania zasobów elektronicznych gromadzonych przez Bibliotekę Narodową Niemiec na podstawie prawa o egzemplarzu obowiązkowym (*Can indexing be automated? - the example of the Deutsche Nationalbibliothek*). Zasoby elektroniczne, ze względu na liczebność, nie mogą być opracowywane w sposób tradycyjny. Z drugiej strony, Biblioteka Narodowa Niemiec podtrzymuje decyzję o stosowaniu kontrolowanego słownictwa (SWD -Schlagwortnormdatei) w opracowaniu wszystkich zasobów. Przygotowanie narzędzia i procedur automatycznego indeksowania z wykorzystaniem SWD było celem projektu Petrus, realizowanego w latach 2009-2011 we współpracy z firmą Averbis, dostawcą systemu Averbis Extraction Platform, dotychczas stosowanego w indeksowaniu wydawnictw medycznych. Do testowania narzędzia wybrano około 20 000 z ponad 187 000 gromadzonych rocznie dokumentów online (prace doktorskie w języku niemieckim). System analizuje tytuł i tekst, wybierając terminy znaczące i ważne, które następnie porównuje z zasobem pobranym z SWD do własnego słownika. Ewaluację i pomiar jakości (tzn. trafności i użyteczności) haseł rzeczowych generowanych automatycznie wykonali na reprezentatywnej próbie specjaliści z działu opracowania rzeczowego. Wynik nie jest jeszcze satysfakcjonujący, system wybrał zbyt dużo haseł nieprawidłowo-

wych i zbyt mało wartościowych. Jeszcze zbyt wcześnie na wdrożenie systemu, zdecydowano jednak kontynuować projekt przez kolejne dwa lata.

W kolejnym referacie Yvonne Jahns (Deutsche Nationalbibliothek) przedstawiła opracowane przez grupę roboczą Sekcji Opracowania Rzeczowego zalecenia dla agencji opracowujących bibliografie narodowe „Guidelines for Subject Access in National Bibliographies”¹. Dokument składa się z dwóch części: opisu podstawowych zasad i procedur, jakich należy przestrzegać, opracowując rzeczowo zasoby rejestrowane w bibliografii narodowej oraz opisu narzędzi i praktyk indeksowania w wybranych kilkudziesięciu narodowych agencjach bibliograficznych (w tym również w Bibliotece Narodowej). Zalecenia te, jak wiele analogicznych dokumentów IFLA, zawierają zarówno bardzo ogólne wskazówki (adresowane do instytucji, które rozpoczynają prace nad rejestracją bibliograficzną lub dopiero ją planują), jak i określają kluczowe dziś i w najbliższej przyszłości zadania. Większość zaleceń powinna zainteresować nie tylko głównego ich adresata, jakim są narodowe agencje bibliograficzne, ale wszelkie instytucje opracowujące i udostępniające bibliografie (np. regionalne, dziedzinowe), a szerzej: instytucje, w których zbiory są opracowywane rzeczowo, a OPAC oferuje punkty dostępu do treści zasobów. Podkreślono odpowiedzialność narodowych agencji bibliograficznych za tworzenie, rozwój i promocję zaleceń i standardów opisu rzeczowego na poziomie krajowym – zaleca się, o ile to możliwe i uzasadnione, adaptację międzynarodowo uznawanych narzędzi indeksowania. Stosowane języki informacyjno-wyszukiwawcze – zarówno o notacji paranaturalnej, jak i klasyfikacje – powinny mieć charakter uniwersalny, obejmować wszystkie dziedziny, problemy, obszary wiedzy i życia. Zasoby słownictwa i/lub symboli powinny być kontrolowane, aczkolwiek zaleca się udostępniać użytkownikom również niekontrolowane punkty dostępu do treści, na przykład możliwość wyszukiwania za pomocą słowa/słów z dokumentu i/lub z opisu (tytuł, abstrakt, uwagi). Hasła/deskryptory/symbole oraz relacje pomiędzy nimi powinny być czytelnie prezentowane w opisach bibliograficznych i w indeksach.

Wiele uwagi poświęcono polityce indeksowania, którą każda z agencji powinna opracować i upubliczniać, ustalając, na jakim poziomie, w zależności od znaczenia lub wartości zasobu, są opracowywane różne typy publikacji (opis pełny czy minimalny). Podjęto problem opisu rzeczowego zasobów elektronicznych – ze względu na ogromną ich liczbę, nie jest możliwe indeksowanie całego „national output”, jak to ma miejsce w przypadku dokumentów drukowanych. Jednym z możliwych rozwiązań jest nadawanie metadanych jedynie tym dokumentom, które mają szczególną wartość intelektualną, naukową lub artystyczną. Ten problem będzie jednym z głównych punktów dyskusji w trakcie prac nad nowelizacją i uzupełnieniem zaleceń, którą zapowiedziała w imieniu Grupy Roboczej Yvonne Jahns, prosząc wszystkie zainteresowane tematem instytucje o aktywny udział w przygotowaniu i opiniowaniu nowej wersji dokumentu.

Ostatnia sesja poświęcona była użytkownikom jako twórcom metadanych przedmiotowych. Pierwszy referat zatytułowany *Folksonomies as Subject Access – A Survey of Tagging in Library Online Catalogs and Discovery Layers*

¹ Guidelines for Subject Access in National Bibliographies, ed. by Yvonne Jahns. IFLA Series on Bibliographic Control nr 45.

został zaprezentowany przez Yan Yi Lee z Wagner College oraz Sharon Q. Yang z Rider University. Przedstawiły one wyniki badania ankietowego, w którym zadano pytania o to, jak dostawcy systemów oraz biblioteki obsługują funkcje tagowania w katalogach OPAC i innych sieciowych systemach wyszukiwawczych (np. AquaBrowser, Endeca, Primo, VuFind itp.). W badaniach uwzględniono też zachowania użytkowników w momencie, gdy dano im sposobność do oznaczania zasobów własnymi słowami kluczowymi. Przebadano możliwości 37 zintegrowanych systemów bibliotecznych pod kątem możliwości dodawania i obsługi tagów oraz 15 innych systemów. Badanie wykazało, że 7 sieciowych systemów daje możliwość obsługi funkcji tagowania, wśród przebadanych systemów bibliotecznych jedynie 2 dają taką możliwość – Genesis 3G oraz Koha. Dalsze badania wykorzystania funkcji tagowania przez 307 bibliotek z systemem Koha (publiczne, szkolne i akademickie) wykazały, że w 49% bibliotek posiadających tę funkcję wyłączono możliwość tagowania przez użytkowników. Tylko 40% bibliotek, które wykazały działanie funkcji tagowania, rzeczywiście wykorzystywało słowa kluczowe dodawane przez użytkowników jako przedmiotowe punkty dostępu. Zauważono też, że użytkownicy bibliotek akademickich są mniej aktywni w tagowaniu zasobów niż użytkownicy bibliotek publicznych.

Maira Bundża (Western Michigan University Libraries) w prezentacji *The Choice is Yours! Researchers Assign Subject Metadata to Their Own Materials in Institutional Repositories* przedstawiła platformę Digital Commons przeznaczoną dla repozytoriów instytucjonalnych. Omówiła możliwości systemu w zakresie metadanych przedmiotowych. Repozytorium zapewnia trypoziomową taksonomię dyscyplin naukowych dla każdej pozycji złożonej w repozytorium. Wykładowcy są zachęceni do osobistego umieszczania materiałów w repozytorium, muszą też sami przydzielać je do odpowiednich dyscyplin lub kategorii przedmiotowych. Rozszerzalne i rozwijane menu z około 1000 kategorii jest łatwe w użyciu i ułatwia rozwój instytucjonalnego repozytorium oraz dostęp do materiałów za pośrednictwem Internetu.

W zamykającej konferencję dyskusji podsumowano poruszane problemy i zastanawiano się nad przyszłą obecnością kontrolowanych zasobów słownictwa w Sieci. Niewątpliwie, konkurencja ze wszechobecną informacją – dostępną łatwo, szybko i bezpłatnie – jest trudna. W ocenie większości uczestników panelu przewagą zasobów tworzonych przez specjalistów informacji powinna być rzetelność i wysoka jakość oferowanych danych. Duże znaczenie ma również stabilność usług i warunków ich dostarczania – serwisy tworzone ze środków publicznych w instytucjach będących w pewnym sensie „instytucjami zaufania publicznego”, w przeciwieństwie do wielu serwisów prowadzonych przez firmy komercyjne, mają – nawet w czasach kryzysu – gwarancje rozwoju i bezpłatnego oferowania usług. Coraz częściej słyszymy opinie, iż era bezpłatnego Internetu kończy się i że wysokiej jakości usługi informacyjne prędkiej czy później będą świadczone odpłatnie – w takim otoczeniu biblioteki mogą odnaleźć i zagospodarować ważną społecznie niszę. Niewątpliwie, przedmiotem troski wszystkich osób odpowiedzialnych za rozwój międzynarodowych i lokalnych narzędzi indeksowania powinno być jak najbardziej efektywne udostępnianie zasobów, zwłaszcza zapewnienie ich widoczności dla wyszukiwarek interneto-

wych. Wysokiej jakości dane to dziś takie, które mogą współpracować z innymi danymi – trudno przecenić znaczenie idei Linked Data, nawet jeśli na razie jej realizacja pozostaje raczej w sferze eksperymentów. Wiele wiodących ośrodków bibliograficznych już prezentuje swe dane jak otwarte zasoby, inne niebawem będą się musiały zmierzyć z tym problemem. Zamykając dyskusję i zarazem całe spotkanie, przewodnicząca Sekcji Jo-Ann Belair wyraziła przekonanie, że przed specjalistami opracowania rzeczowego są czasy trudne, pełne wyzwań, ale i niezwykle inspirujące.

Teksty wszystkich wystąpień, prezentacje oraz galeria zdjęć są dostępne pod adresem <http://www.nlib.ee/tallinnsatellite>. Referaty będą również publikowane jako materiały pokonferencyjne w serii IFLA Series on Bibliographic Control.

Informując o planowanych działaniach Sekcji Opracowania Rzeczowego, nie podano na razie informacji, czy sekcja zorganizuje konferencję towarzyszącą przyszłorocznemu kongresowi IFLA, który odbędzie się w Singapurze. Bardziej prawdopodobne jest, że kolejne spotkanie będzie miało miejsce w 2014 r. we Francji, przy okazji zaplanowanego kongresu IFLA w Lyonie.

Wanda Klenczon
Paweł Rygiel
Biblioteka Narodowa