

Jolanta Gwioździk

Od Redakcji

Bibliotheca Nostra : śląski kwartalnik naukowy 2/2, 9-12

2010

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

OD REDAKCJI

Współczesne zasoby informacyjne, powstałe dzięki rozwojowi technologii cyfrowych, wymagają nowego ujmowania podstawowych zadań bibliotek: gromadzenia, opracowania, udostępniania i informowania o zbiorach. Znane systemy organizowania wiedzy, jak autorytatywne zbiory nazw, klasyfikacje (np. Klasyfikacja Biblioteki Kongresu) i inne języki informacyjno-wyszukiwawcze (język deskryptorowy z tezauresem czy język haseł przedmiotowych) znajdują nowe zastosowanie w sieci, m.in. w automatycznym rozpoznawaniu i wyodrębnianiu metadanych. Stosowana tu metoda polega na przetwarzaniu syntaktycznym i wykorzystaniu algorytmów maszynowego uczenia się. Użytkownicy, przyzwyczajeni do narzędzi internetowych, np. do usług Google, oczekują podniesienia wartości i atrakcyjności informacyjnej katalogów bibliotecznych. Katalogi elektroniczne w Internecie są zatem uzupełniane o elementy dodatkowe, dodawane do rekordów bibliograficznych, takie jak odsyłacze, tagi, recenzje lub zeskanowane obrazy okładek (to ostatnie wzorem Amazona). Możliwa staje się także zmiana błędnych zapytań czy pokazywanie relacji skojarzeniowych między terminami, co służy lepszemu poznawaniu przez użytkownika i wykorzystywaniu zbiorów bibliotecznych. Do ich prezentacji w sieci wykorzystuje się programy komercyjne (jak Prima czy AquaBrowser), a także darmowe, działające na zasadzie open source (jak VuFind, SOPAC).

Takie rozwiązania są charakterystyczne dla wyszukiwarek semantycznych, korzystających z tzw. metadanych semantycznych, a więc opisujących treść dokumentu. Należą do nich po pierwsze wyszukiwarki analizujące "znaczenie" treści indeksowanych dokumen-

tów, np. Haka przeszukuje nie tyle strony internetowe, co ich zawartość znaczeniową, stosując odpowiednie metody analizy tekstu dokumentu, dzięki procedurom mechanicznego przetwarzania języka naturalnego. Po drugie są to wyszukiwarki związane z rozwojem sieci semantycznej (Web 3.0), przeszukujące zawartość sieci semantycznej, czyli zawartość indeksowanych plików RDF, stanowiących metaopis zawartości dokumentów internetowych.

Tematykę automatycznego tworzenia metadanych podjął Marek Nahotko. To zagadnienie staje się szczególnie aktualne w sytuacji wzrastających zasobów sieci, powstawania globalnej biblioteki cyfrowej, czyli „scentralizowanych usług wyszukiwania informacji w rozproszonych zasobach informacyjnych”. W bibliotekach, archiwach czy ośrodkach informacji dominuje tworzenie metadanych przez ludzi, jednak szybki wzrost treści dostępnych w sieci wymaga wykorzystywania w tym celu procedur zautomatyzowanych. W pierwszym przypadku indeksowanie semantyczne jest realizowane dzięki prototypowym pojęciom, w drugim – przez identyfikację zestawu cech, bez rozumienia sensu informacji zawartej w obiekcie cyfrowym. Często konieczna jest „kompatybilność obu sposobów oceny treści”, a nawet ich integracja, gdy osoby inicjują lub kontrolują procesy zautomatyzowane.

W wyszukiwarkach komercyjnych automatyczne tworzenie metadanych jest wykonywane przed wysłaniem zapytania przez użytkownika, służy wówczas do przygotowania zasobu metadanych, z którymi jest porównywane to pytanie, a także już podczas prowadzenia wyszukiwania, dzięki czemu przeszukuje się zasoby poza bazą danych wyszukiwarki. Autor analizuje metody automatycznego tworzenia metadanych dla zasobów cyfrowych, zwracając jednak uwagę, że wciąż brakuje badań ich efektywności dla przygotowania rekordów metadanych. Kierunkom badań w ujęciu chronologicznym jest poświęcona druga część jego artykułu. Zwraca także uwagę na generatory metadanych, specjalnie i wyłącznie przeznaczone do tworzenia rekordów metadanych.

Rozwój sieci semantycznych będzie wpływał na wzrost funkcjonalności wyszukiwarek, a także na integrację różnych modeli opisu danych, w tym formatów zapisu metadanych bibliograficznych. Przykładem takich działań jest aplikacja Zotero, ułatwiająca archiwizowanie i organizowanie źródeł informacji, w tym tworzenie bibliografii. Wraz dodatkiem Zotz (Firefox) pozwala na łatwiejsze przeszukiwanie i filtrowanie dużych zasobów, a pobrane dane można samodzielnie modyfikować. Interfejs jest także dostępny w języku pol-

skim. W tej sytuacji nowego znaczenia nabiera normalizacja i stosowanie standardów w zakresie formatu zapisu danych bibliograficznych w systemach komputerowych.

Agata Kyzioł w następnym artykule omówiła format MARC21 (MACHine READable Cataloging), pierwotnie opracowany przez Bibliotekę Kongresu (jako Pilot MARC w 1965 r.), a następnie modyfikowany (obecna wersja opublikowana została po raz pierwszy w 1999 r.). Wskazała oczywiste zalety formatu, jak prawidłowe odczytanie zapisu bez względu na kraj i język, możliwość uwzględniania specyfiki wielu rodzajów dokumentów, które ułatwiają wymianę rekordów i sprzyjają tworzeniu baz danych. Za twórcami formatu zdefiniowała pięć typów danych: opis bibliograficzny, zasób i lokalizacja, hasła wzorcowe, symbol klasyfikacji oraz dane faktograficzne. Autorka na przykładzie wydawnictw zwartych szczegółowo przeanalizowała opis bibliograficzny na różnych poziomach, z uwzględnieniem stosowania określonych pól i podpól, ich wzajemnych korelacji, co z pewnością posłuży praktycznemu poznaniu formatu.

Możliwość jego zastosowania w odniesieniu do opracowania zbiorów kartograficznych jest jednym z zagadnień, poruszanych w tekście autorstwa Urszuli Iwańskiej i Ireny Nowak. Charakterystyczną cechą dokumentów kartograficznych (takich jak atlasy, różnego rodzaju mapy, zdjęcia lotnicze i satelitarne, przekroje, globusy) jest zróżnicowanie jednostek opisu, jak dokument jednoczęściowy, praca współwydana, jedna część dokumentu wieloczęściowego oraz dokument wieloczęściowy jako całość. Zasadniczo wpływa to na sposób ich opracowania. Związane z tym dylematy przedstawiono na przykładzie zbiorów kartograficznych Biblioteki Wydziału Nauk o Ziemi Uniwersytetu Śląskiego. Omówiono obowiązującą normalizację i opracowanie zbiorów kartograficznych w dwóch systemach: Bitmap i Prolib, wykorzystujący format MARC21, rozważając ich cechy i zakres stosowania.

Różne aspekty współczesnego opracowania zbiorów nie tylko obrazują zmieniające się wraz z rozwojem technologii cyfrowych potrzeby użytkowników, ale także ukazują nowe możliwości ich zaspokajania, uzupełniające tradycyjne narzędzia. Automatyczne tworzenie metadanych okazuje się szybsze, tańsze i na tyle wydajne, że może być wykorzystane do indeksowania tak dynamicznych zasobów, jak dostępne we współczesnej Sieci: „...funkcjonuje poprawnie, można nawet powiedzieć, że daje równie dobre rezultaty, jak indeksowanie realizowane przez człowieka, jest tylko inne...”. Ważna jest także możliwość wykorzystania wiedzy internatów na potrzeby opra-

cowania zbiorów. Interesującym – i zakończonym sukcesem – przykładem takiego udziału społeczności internetowej jest projekt o nazwie „The Commons”, animowany przez Bibliotekę Kongresu, służący udostępnianiu fotografii historycznej poprzez serwis Flickr, co znacząco wpłynęło także na promocję zbiorów fotografii, bowiem zbiory biblioteczne bywają nie w pełni wykorzystywane z powodu braku wiedzy o nich potencjalnych odbiorców.

Jolanta Gwioździk

