

Marek Nahotko

Metadane dla czasopism elektronicznych

1. Wstęp

Metadane służą do opisu dokumentów elektronicznych. Można je nazwać "danymi o danych". W zastosowaniu do dokumentów World Wide Web wykorzystywane są do wyszukiwania informacji i służą opracowaniu dokumentów elektronicznych. Istnieje wiele standardów metadanych, gdyż każde środowisko (np. bibliotekarze, archiwiści, muzealnicy, wydawcy) stara się dostosować elementy metadanych do własnych potrzeb. Dominującą rolę spełnia tu Dublin Core (<http://purl.org/dc>), jako standard metadanych dla prostego opisu zasobów. Ułatwia on także współpracę pomiędzy instytucjami używającymi różnych standardów metadanych.

Dane bibliograficzne oraz tekst abstraktu artykułu publikowanego w czasopiśmie elektronicznym udostępnianym w Internecie są niewątpliwie metadanymi tego artykułu. Przy stosowaniu tych metadanych możliwe jest wykorzystanie standardu Dublin Core.

2. Dublin Core

Dublin Core Metadata Element Set (DCMES) - tak brzmi pełna nazwa tego standardu - został wprowadzony przez organizację o nazwie Dublin Core Metadata Initiative (DCMI). Jest on efektem prac tzw. Dublin Core Working Groups, których członkami są specjaliści z różnych krajów. To międzynarodowe towarzystwo w codziennych kontaktach posługuje się głównie pocztą elektroniczną. Corocznie natomiast spotykają się podczas warsztatów Dublin Core. W 2000 roku warsztaty te odbyły się po raz ósmy. Spotkanie miało miejsce w Ottawie (zob. <http://www.ifla.org/udt/dc8/>).

Standard Dublin Core dla tworzenia metadanych jest pomyślany głównie jako narzędzie do budowy semantyki metadanych, bez poruszania zagadnień syntaktyki, dla tworzenia której istnieją inne standardy, jak np. HTML czy XML.^[8] Jest on przeznaczony do tworzenia prostych opisów zasobów elektronicznych oraz zapewnienia minimalnego stopnia przenoszalności danych pomiędzy różnymi systemami metadanych, co potencjalnie umożliwia ich wymianę pomiędzy różnymi środowiskami. Dublin Core nie pretenduje do tego, aby być narzędziem załatwiającym wszystkie potrzeby dotyczące metadanych we wszelkich ich zastosowaniach. W różnego rodzaju specyficznych zastosowaniach Dublin Core może być podstawą do tworzenia dziedzinowych schematów metadanych umożliwiających bardziej rozbudowany opis źródeł.

Podstawą Dublin Core jest zestaw piętnastu elementów metadanych ^[6] (zob. rys. 1.). Szczegółowy opis i definicja tych elementów w języku polskim znajduje się na stronach EBIB (<http://ebib.oss.wroc.pl/standard/dc.html>). Wszystkie elementy Dublin Core są opcjonalne i powtarzalne.

Dalsze doskonalenie i uszczegółowienie opisu zawartości elementów umożliwia kwalifikowany Dublin Core. Wykorzystanie kwalifikatorów pozwala na zwiększenie semantycznej szczegółowości metadanych. Zestaw kwalifikatorów Dublin Core (<http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm>) został niedawno wstępnie zaaprobowany przez Dublin Core Usage Committee po wielu miesiącach burzliwych dyskusji i setkach poprawek. Kwalifikowany Dublin Core posiada np. taki kwalifikator uszczegółowiający jak "Tytuł równoległy". Ma również możliwość określania schematów kodowania dla wartości elementów. Czyni się starania, aby schematy kodowania, którymi są np. schematy klasyfikacyjne i sformalizowane listy (np. kodów języków), były standardami międzynarodowymi.

Zawartość (Content)	Własność intelektualna (Intellectual Property)	Dookreślenie ? Instantiation
Tytuł (Title)	Twórca (Creator)	Data (Date)
Opis rzeczowy (Subject)	Wydawca (Publisher)	Typ (Type)
Opis (Description)	Współtwórca (Contributor)	Format (Format)

Źródło (Source)	Własność (Rights)	Identyfikator (Identifier)
Język (Language)		
Relacja (Relation)		
Miejsce i Czas (Coverage)		

Rys.1. Schemat elementów Dublin Core

Podczas praktycznego stosowania Dublin Core można spotkać się z kilkoma problemami. Po pierwsze, podstawowy zestaw elementów może być zbyt mały dla pełnego opisu dokumentu elektronicznego. Po drugie, długi okres podejmowania decyzji co do definicji elementów metadanych Dublin Core może zmuszać ze względów praktycznych do stosowania pewnych rozwiązań zanim zostaną one ogólnie przyjęte, co może w efekcie prowadzić do rozbieżności w stosowaniu standardu.

Jak już stwierdzono, podstawowy zestaw elementów Dublin Core ma charakter bardzo ogólny. Zakłada się, że projektanci rozwiązań lokalnych, aby zaspokoić własne potrzeby, będą włączali do opisu dodatkowe elementy i kwalifikatory. Jednocześnie użytkownicy Dublin Core liczą na to, że rozpowszechniane będą lokalnie wprowadzone kwalifikatory, które mogą znaleźć szersze zastosowanie. Twórcy lokalnych aplikacji będą zachęceni do rejestracji swoich kwalifikatorów w Dublin Core Metadata Registry. Wpłynie to na ich szersze wykorzystanie, a co za tym idzie, zapewni przenoszalność.

Ze względu na przyjętą metodę tworzenia definicji Dublin Core przez otwarte grupy dyskusyjne, do osiągnięcia konsensusu potrzeba znacznej ilości czasu. Dotyczy to w szczególności Kwalifikowanego Dublin Core, którego pierwszą wersję dopiero niedawno zaakceptowano. Powoduje to powstawanie problemów dla twórców konkretnych aplikacji, którzy muszą tworzyć je na podstawie stale zmieniających się założeń. Jak twierdzą niektórzy, dotychczas wykorzystanie Dublin Core przypominało trafianie do ruchomego celu. W takich warunkach przyjęte już decyzje musiały ulegać zmianom, których wprowadzenie mogło być trudne w konkretnym, pracującym systemie.

3. Czasopisma elektroniczne

Istnieje kilka publikacji [1], [2], [5] na temat wykorzystania metadanych do opisu poszczególnych artykułów z czasopism elektronicznych. Można powiedzieć, że Dublin Core szczególnie nadaje się do tego rodzaju zastosowania. Niezbędne jest jednak określenie sposobu postępowania z elementami danych bibliograficznych cytowanego artykułu, takimi jak: tytuł czasopisma, rocznik i numer zeszytu, części nazwy Twórcy, przynależność Twórcy, lokalizacja i wielkość pełnego tekstu artykułu oraz rodzaj pliku (HTML, PDF itp.), konkretny typ dokumentu oraz język tych elementów, które mogą mieć różnojęzyczne wersje.

Pomimo że wszystkie elementy Dublin Core są opcjonalne, to aby wykorzystać je do opisu artykułów z czasopism elektronicznych niektóre elementy muszą być obowiązkowe. W projekcie MIMAS [5] zastosowano następujące elementy obowiązkowe: Tytuł, Wydawca, Data, Typ, Format, Identyfikator, Źródło, Język, Relacja i Własność. Wymogi te wynikają z XML DTD. [1]

Elementy metadanych dla artykułu czasopisma mogą być w następujący sposób odnoszone do standardu Dublin Core (rys.2.):

Metadane artykułu	Elementy Dublin Core
Tytuł	Tytuł
Autor	Twórca
Słowa kluczowe	Opis rzeczowy
Abstrakt	Opis
Nazwa wydawcy	Wydawca
Rocznik	Data
Język artykułu	Język
Pełny format artykułu	Format
Copyright	Własność

Rys. 2. Przeniesienie metadanych artykułu do elementów Dublin Core.

Elementy Dublin Core 'Współtwórca' i 'Miejsce i Czas' nie są w tym przykładzie wykorzystane. Przewidziano powtarzalność elementów 'Twórca' i 'Opis rzeczowy'. Za powtarzalne uznano również elementy 'Tytuł' i 'Opis', aby umożliwić wprowadzanie metadanych w wielu językach.

4. Metadane artykułu z czasopisma.

Głównym problemem związanym z wykorzystaniem metadanych Dublin Core dla artykułu pochodzącego z czasopisma elektronicznego jest wydobycie informacji bibliograficznej o artykule w obrębie zeszytu czasopisma. Problem ten był rozważany przez Dublin Core Working Group, który zaproponował następujące rozwiązania. DC-Citation Working Group [3] proponuje, aby informacje na temat artykułów z czasopism umieszczać w obrębie elementu DC.Relacja. Rozważane było także użycie takich elementów, jak: 'Tytuł', 'Opis', 'Identyfikator' i 'Źródło' - jednak zostały one odrzucone. Uznano że DC.Tytuł powinien zawierać wyłącznie tytuł artykułu, a nie jakiegokolwiek inne informacje. DC.Identyfikator powinien zawierać jeden lub więcej identyfikatorów dla samego artykułu (np. SICI [7], PII, DOI[4] lub URL artykułu), ale nie powinno tu być identyfikatorów dla zeszytu, woluminu lub całego czasopisma.

Zaproponowano także wykorzystanie kwalifikatora "IsPartOf" ("JestCzęścią"). Pełna informacja o lokalizacji powinna być umieszczona w formie danych tekstowych i zawierać dodatkowo jeden lub więcej identyfikatorów wskazujących źródło, którego artykuł jest częścią. Dane tekstowe powinny zawierać lokalizację na stronach (lub odpowiednią informację o lokalizacji w źródłach, których numeracja stron nie dotyczy). Takie rozwiązanie zostało rekomendowane, gdyż:

- zakres stron, na których zamieszczony jest artykuł, pojawia się w sposób naturalny na końcu informacji bibliograficznych czasopisma;
- można oczekiwać, że stosujący standard i tak tam go umieszczą;
- robią to, gdyż brak innego odpowiedniego miejsca.

Aby utworzyć dane bibliograficzne artykułu w obrębie czasopisma wystarczające jest określenie Tytułu czasopisma, Numeru woluminu i Numeru strony początkowej (dla artykułu w czasopiśmie drukowanym). Na ogół opis bibliograficzny zawiera także rok wydania. Dla metadanych artykułu w czasopiśmie elektronicznym niezbędne jest także posiadanie numeru zeszytu zawierającego artykuł. Przydatne byłoby też posiadanie ISSN czasopisma oraz Numeru strony końcowej artykułu w czasopiśmie drukowanym. Jednym ze schematów definiującym te informacje jest Serial Item and Contribution Identifier (SICI) [7].

Nazwa autora artykułu bez wątplenia powinna zostać umieszczona w elemencie DC.Twórca. Jednak dotąd brak dla tego elementu zatwierdzonych kwalifikatorów i schematów kodowania. Wartością elementu jest nazwa Twórcy wpisywana jako ciąg znaków. W zastosowaniu do czasopism elektronicznych pożądane jest, aby możliwe było podzielenie nazwy autora na części składowe. Dla użytkowników końcowych może być przydatne, np. wyodrębnienie samego nazwiska (bez imion). W związku z tym może zaistnieć potrzeba zastosowania struktur zdefiniowanych lokalnie.

Kolejnym problemem jest uzyskanie danych o przynależności (afiliacji) autora. Dla artykułu opublikowanego w czasopiśmie uczelnianym przynależność wskazuje na instytucję zatrudniającą autora w czasie publikacji artykułu, co nie musi być jego aktualnym adresem. Należy jednak uwzględnić przypadki, gdy kilku autorów posiada tę samą przynależność (wówczas zostanie ona powtórzona dla każdego autora) lub autor posiada kilka przynależności. Aby nie powtarzać tych samych adresów w informacji wyświetlanej dla użytkownika końcowego, lepiej jest tę informację zdefiniować raz (np. w SGML) z odsyłaczami do odpowiednich nazwisk autorów.

Niezbędne jest ujęcie w jakiś sposób formatu pełnego tekstu artykułu, którego dotyczą metadane. Teksty te mogą być udostępniane w różnych formatach, z których najpopularniejsze są HTML i PDF. Informacja o formacie pełnego tekstu artykułu ujmowana jest w elemencie DC.Format. Dodatkowo można zamieścić informację o rozmiarze pliku, co ułatwi użytkownikowi podjęcie decyzji o jego ściągnięciu.

Aktualnie przyjęty schemat kodowania wartości elementu DC.Typ, DCT1 Type Vocabulary (<http://purl.org/DC/documents/wd-typelist.htm>), nie dostarcza żadnych środków pozwalających wskazać, że tworzone metadane dotyczą artykułu z czasopisma. Schemat kodowania 'DCT1' cechuje się wysokim stopniem abstrakcji. Obecnie zatwierdzonymi terminami są:

- Kolekcja (collection);
- Zestaw danych (dataset) - ustrukturyzowana informacja w postaci list, tabeli, baz danych itp.;
- Wydarzenie (event);
- Obraz (image);
- Źródło interaktywne (interactive resource) - źródło wymagające interakcji ze strony użytkownika w celu jego zrozumienia, wykonania lub zbadania (np. formularze na stronach Web, aplety itp.);
- Model (model) - abstrakcja dotycząca przedmiotów rzeczywistych (np. model kosztów);
- Grupa (party);
- Obiekt fizyczny (physical object);
- Miejsce (place);
- Usługa (service) - np. usługi bankowe, wypożyczenia międzybiblioteczne, Z39.50.;
- Oprogramowanie (software);
- Dźwięk (sound);
- Tekst (text).

Wymienione typy mają być wzbogacone o wykazy podtypów. Przy pomocy powyższego schematu, artykuły z czasopism mogą być określone po prostu jako teksty. Jednak w odniesieniu do czasopism naukowych przydatne są także inne typy. Lista przydatnych podtypów może zawierać np.: Ogłoszenie, Przegląd

publikacji, Errata, Od wydawcy, Forum dyskusyjne, Komentarz od redakcji, Artykuł naukowy, Artykuł przeglądowy, Raporty, Relacje, Spis treści itp. Czynnione są starania, aby podobna lista typów odpowiednia dla czasopism elektronicznych została zarejestrowana jako słownik kontrolowany Dublin Core.

W czasopiśmie mogą zdarzać się artykuły, abstrakty, czy tytuły równoległe artykułów w więcej niż jednym języku. Powinno to znaleźć swoje odniesienie w metadanych i zostać zaprezentowane użytkownikowi wraz z inną informacją o artykule. W tym celu należy w elementach DC.Tytuł i DC.Opis umieścić powtarzalny atrybut dotyczący języka. Kody języków są przedstawione w RFC 1766 (<http://www.ietf.org/rfc/rfc1766.txt>).

Poniżej przedstawiono przykład metadanych wykonanych przy użyciu schematu Dublin Core, dla niniejszego artykułu, napisane w HTML:

```
<HEAD>
<TITLE>Metadane dla czasopism elektronicznych</TITLE>
<META name="DC.title" content="Metadane dla czasopism elektronicznych">
<META name="DC.creator" content="Nahotko, Marek">
<META name="DC.subject" content="metadane, Dublin Core, Czasopisma elektroniczne, katalogowanie">
<META name="DC.description" content="W artykule omówiono sposoby opracowania artykułów czasopism elektronicznych przy użyciu schematu metadanych Dublin Core">
<META name="DC.publisher" content="Ossolineum">
<META name="DC.date" content="(Scheme =ISO 8601) 2001-01-06">
<META name="DC.type" content="Text.Serial.Journal.Article">
<META name="DC.format" content="text/html">
<META name="DC.identifier" content="www.oss.wroc.pl/biuletyn/ebib19/nahotko.html">
<META name="DC.source" content="EBIB 2001, 1">
<META name="DC.language" content="pl">
<META name="DC.rights" content="©Wyd. Ossolineum 2001">
</HEAD>
```

Rys. 3. Przykładowe metadane artykułu w schemacie Dublin Core (HTML).

Zakończenie

Różne próby zastosowania Dublin Core dla opisu treści czasopism elektronicznych, często pozostające w fazie eksperymentów, wykazują, że możliwe jest jego wykorzystanie w tym zakresie. Niezbędne jest jednak uzupełnienie schematu o pewne lokalnie funkcjonujące schematy kodowania danych, głównie wartości strukturalnych. Nie jest natomiast konieczne dodawanie nowych elementów metadanych ponad standardowe piętnaście.

Prowadzone są obecnie starania o przekształcenie Dublin Core w międzynarodowy standard dla metadanych służący prostemu opisowi zasobów i zapewnieniu przynajmniej minimalnej przenoszalności danych. Prace te realizowane są zarówno w CEN/ISSS Workshop on Metadata for Multimedia Information - Dublin Core (<http://www.cenorm.be/iss/Workshop/MMI-DC>), jak i w amerykańskim NISO (<http://www.niso.org/Z3985.html>).

Wykorzystanie standardowych metadanych, takich jak Dublin Core, w lokalnych zastosowaniach zapewni w przyszłości możliwość współpracy pomiędzy tymi aplikacjami. W określonej realizacji możliwe będzie prowadzenie wyszukiwania w obrębie wielu zastosowań, np. w bazach danych i zasobach pełnotekstowych odpowiedniej literatury naukowej jednocześnie, jak również wyświetlanie tylko części informacji, np. na ekranie telefonu komórkowego.

Jak dotąd głównym problemem związanym z wykorzystaniem metadanych Dublin Core i zapewnieniem odpowiedniej jakości pracy jest brak stabilizacji tego standardu. Jednak i w tym względzie należy odnotować pewien postęp - podstawowe elementy Dublin Core w wersji 1.1 zostały już ustalone i nie będą ulegać zmianom. Powstała także pierwsza wersja Kwalifikatorów Dublin Core. Wydaje się więc, że już obecnie Dublin Core jest dostatecznie dojrzałym tworem, akceptowanym jako standard wszędzie tam, gdzie istnieje potrzeba używania metadanych.

Przypisy:

[1] Definicja Typu Dokumentu (DTD ang. Document Type Definition) - Opis składników określonego dokumentu lub klasy dokumentów. Opis DTD zawiera:

- Części lub elementy, które tworzą dokument, np.: paragrafy, akapity, nagłówki, listy, rysunki, tablice itp.
- Logiczną strukturę dokumentu, np.: rozdziały z podrozdziałami itp.
- Dodatkową informację związaną z elementami (nazywaną atrybutem), np.: identyfikatory, data, znaki itp.

Literatura:

1. A. Apps: *SuperJournal Metadata Specification*. SuperJournal Project Report.
<http://www.superjournal.ac.uk/sj/sjmc141.htm>
2. A. Apps, R. MacIntyre: *Metadata for the Nature Digital Archive*. <http://epub.mimas.ac.uk/natpaper.html>
3. DC Citation Working Group <http://purl.org/dc/documents/wd-citation-19990702.htm>
4. DOI, Digital Object Identifier <http://www.doi.org>
5. Electronic Publishing at MIMAS web site. <http://epub.mimas.ac.uk>
6. M. Nahotko: *Metadane*. EBIB 2000 Nr 6 (14): <http://www.oss.wroc.pl/biuletyn/ebib14/nahotko.html>
7. Serial Item and Contribution Identifier (SICI) ANSI/NISO Z39.56-1996 (Version 2)
<http://sunsite.berkeley.edu/SICI/sici.pdf>
8. XML. <http://www.w3.org/XML>

Pierwotny adres: <http://ebib.oss.wroc.pl/2001/19/nahotko.html>



Metadane dla czasopism elektronicznych / Marek Nahotko// W: Biuletyn EBIB [Dokument elektroniczny] / red. Bożena Bednarek-Michalska - Nr 1/2001 (19) styczeń. - Czasopismo elektroniczne. - [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich. KWE, 2001. - Tryb dostępu: <http://www.ebib.pl/2001/19/nahotko.php>. - Tyt. z pierwszego ekranu. - ISSN 1507-7187