

Marko Milenković
(Matematički fakultet, Beograd)

INICIJATIVA “DUBLIN CORE”

Prikazujemo jedan lak i fleksibilan način opisivanja resursa. Jednostavno kreiranje i održavanje, razumljiva semantika i mogućnost lokalizacije su ciljevi Dublin Core Metadata Initiative. Implementacija je moguća preko XML-a i HTML-a pri čemu prva od njih pruža daleko veće mogućnosti. Opis nekog resursa se može čuvati unutar ili van njega, što je pogodno za davanje opisa netekstualnih podataka. Mogućnost lokalizacije je dokazana, pošto su vlade mnogih zemalja uvele Dublin Core kao standardni način opisivanja podataka.

Ključne reči: metapodaci, XML, HTML

Uvod

Metapodaci su skup atributa potrebnih da se opiše neki resurs. Na primer, metapodaci se koristi u bibliotekama za opis knjiga, pa tako bibliotekarski katalog sadrži kratak opis svake knjige (ime autora, naslov, datum izdavanja, mesto na polici...). Postoje dva načina čuvanja podataka i metapodataka:

1. metapodaci i podaci su razdvojeni, kao u slučaju bibliotekarskog kataloga.
2. metapodaci su postavljeni uz sam podatak.

Drugi način se koristi pri štampanju knjiga, gde se podaci o knjizi stavljaju u samu knjigu (*Cataloging In Publication – CIP*).

Dublin Core Metadata Initiative DCMI

Dublin Core Metadata Initiative (DCMI, www.dublincore.org) je organizacija koja se bavi izradom standarda i rečnika za opis informacija u cilju bolje i lakše manipulacije. Fokus rada DCMI je izrada semantike standarda za metapodatke. Dublin Core (DC) je mali skup metapodataka koji služe za opis informacije. Standard je ustanovljen kroz konsenzus međunarodne grupe koju su sačinjavali profesionalci iz raznih oblasti: bibliotekari, informatičari, kustosi... Ukratko, "Dublin Core je mali jezik za pravljenje klase iskaza o resursu" [3]. U ovom jeziku postoje dve vrste termova: elementi i kvalifikatori, koji mogu biti uređeni u prost šablon iskaza.

DC ima za cilj:

jednostavnost u kreiranju i održavanju: teži se što većem pojednostavljivanju, kako bi svako mogao da napravi skup opisnih iskaza.

razumljivu semantiku: kako bi se olakšala pretraga preko globalne mreže svima koji imaju potrebu za informacijom.

lokalizaciju: DC je originalno implementiran u engleskoj, ali postoje verzije koje su napisane za mnoge druge jezike (ruski, kineski, finski, norveški, japanski...)

Sintaksa

Sam DC se može implementirati preko nekoliko različitih sintaksi: HTML, RDF/XML (Resource Description Framework koristeći eXtensible Markup Language) ili u generičkoj formi (element="vrednost"). HTML se može iskoristiti za demonstraciju DC koncepta, ali složenije aplikacije je najbolje praviti u RDF/XML-u.

DC je veoma prilagodljiv način opisa podataka; pored standardnih tipova, bibliotekarskih, muzejskih tipova, moguće je opisati i mesta na osnovu geografskog položaja, poštanskog broja ili slično [5].

Osnovni principi deskriptivnih elemenata

Svaki element se može više puta pojaviti, što doprinosi višeznačnosti i opcionalan je. Metapodaci se mogu pojavljivati u bilo kojem redosledu. Višestruko pojavljivanje može biti od koristi, ali se mora voditi računa da neki tipovi reprezentacije objekata to ne podržavaju. HTML je jedan od njih.

Elementi za opis nekog podatka se mogu izabrati iz "kontrolisanog rečnika", koji je ograničen skup često korišćenih i precizno definisanih izraza. Bez bazične kontrolne terminologije može se drastično smanjiti uspeh pretraživanja. Pravljenje dobrog "kontrolisanog rečnika" je moguće uz korišćenje kvalifikatora. U tabeli su prikazani osnovni elementi koji se koriste pri opisu resursa.

Content	Intellectual Property	Instantiation
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		

Dublin Core Metadata Element Set

Prikaz DCMES (Dublin Core Metadata Element Set) je definisan po RFC2413 [11]. Ovu definiciju zovemo verzija 1.1. Svaki element DCMES je opisan uz korišćenje attribute iz standarda ISO/IEC 11179 (ISO11179) [12]. Atributi koji se koriste po ovom standardu su:

- **Name** Oznaka dodeljena elementu podatka;

- **Identifier** Jedinstveni identifikator koji je dodeljen elementu podatka;
- **Version** Verzija elementa;
- **Registration Authority** Organizacija zadužena za registraciju elementa;
- **Language** Jezik na kome je preciziran element;
- **Definition** Definicija elementa koja jasno opisuje njegovu svrhu;
- **Obligation** Definiše potrebu za pojavljivanjem nekog elementa;
- **Datatype** Precizira koje vrste je podatak koji može biti preciziran kao vrednost elementa;
- **Maximum Occurrence** Maksimalni broj pojavljivanja elementa;
- **Comment** Komentar.

Šest od deset atributa je zajedničko za sve DC elemente, pa se oni pri opisu elemenata skupa ne navode. Ti atributi su:

Version: 1.1

Registration Authority: Dublin Core Metadata Initiative

Language: en

Obligation: optional

Datatype: Character string

Maximum Occurrence: Unlimited

Prikaz standardnih elemenata, sa pratećim atributima:

Element: Title

Name: Title

Identifier: Title

Definition: Ime dato objektu

Comment: Naslov je ime po kome je resurs poznat.

Element: Creator

Name: Creator

Identifier: Creator

Definition: Podatak koji prikazuje ko je odgovoran za pravljenje sadržaja resursa

Comment: Creator treba da sadrži ime osobe, organizacije ili servisa koji je kreirao resurs.

Element: Subject

Name: Subject and Keywords

Identifier: Subject

Definition: Prikaz sadržaja resursa

Comment: Obično Subject sadrži ključne reči ili fraze koje opisuju sadržaj resursa.

Element: Description

Name: Description

Identifier: Description

Definition: Opis sadržaja resursa

Comment: Description služi za opis objekta. Može da sadrži sadržaj, reprezentaciju nekog grafičkog objekta...

Element: Publisher

Name: Publisher

Identifier: Publisher

Definition: Odgovara na pitanje ko je zaslužan za objavljivanje resursa

Comment: Publisher obično sadrži ime osobe, organizacije ili servisa

Element: Contributor

Name: Contributor

Identifier: Contributor

Definition: definiše ko je dao doprinos sadržaju resursa

Comment: Contributor sadrži ime osobe, organizacije ili servisa.

Element: Date

Name: Date

Identifier: Date

Definition: Datum je asociran sa nekim događajem iz života resursa.

Comment: Obično se datum vezuje za sa kreiranje resursa. Način zapisa datuma je po ISO8601 yyyy-mm-dd [13]

Element: Type

Name: Resource Type

Identifier: Type

Definition: "Priroda" sadržaja resursa

Comment: pogledati dodatak A

Element: Format

Name: Format

Identifier: Format

Definition: Psihička ili digitalna manifestacija objekta

Comment: Format može sadržati vrstu medija ili veličinu resursa, a može sadržati i softver ili hardver potreban za manipulaciju rasursom. Preporučuje se da se za vrednosti Formata uzimaju vrednosti po MIME [14] specifikaciji. Osnovni opis MIME specifikacije je dat u dodatku B.

Element: Identifier

Name: Resource Identifier

Identifier: Identifier

Definition: Nedvosmisleno pozivanje na resurs

Comment: u praksi vrednost identifikatoru treba dati prema dogovorenim pravilima.

Formalni identifikacioni sistem sadrži Uniform Resource Identifier (URI) Digital Object Identifier (DOI) i International Standard Book Number (ISBN)

Element: Source

Name: Source

Identifier: Source

Definition: Referiše na objekat iz koga je sadašnji objekat nastao

Comment:

Element: Language*Name:* Language*Identifier:* Language*Definition:* Jezik intelektualnog sadržaja resursa*Comment:* Preporučuje se da se za vrednosti ovog elementa koriste vrednosti definisane u RFC1766 [15].**Element: Relation***Name:* Relation*Identifier:* Relation*Definition:* Relacija ka srodnim resursima*Comment:* Relacija treba da bude predstavljena kao reč ili broj, tako da se preko nje može pristupiti srodnim resursima. U tesnoj je vezi s elementom Identifier.**Element: Coverage***Name:* Coverage*Identifier:* Coverage*Definition:* Prošireni sadržaj resursa*Comment:* Coverage sadrži imena, lokaciju ili vremenski period koji su važni za resurs.**Element: Rights***Name:* Rights Management*Identifier:* Rights*Definition:* Informacije o intelektualnoj svojini resursa.*Comment:***RDF/XML**

RDF dozvoljava da se primeni više šema metapodataka, koje su čitljive i razumljive, kako za ljude tako i za mašine. Koristi se XML kako bi se izrazila struktura. Prednost je u tome što ne postoji određena semantika već se ona definiše u zavisnosti od potrebe.

Dokumenti gde su Dublin Core Metadata Element Set (DCMES) kodirani preko RDF/XML-a podležu sledećim ograničenjima:

- samo DC elementi koji su opisani u DMCES mogu biti korišćeni; nijedan drugi element ne sme biti korišćen.
- rezultujući RDF/XML dokument se ne može staviti unutar polaznog dokumenta.

Znaci < > ' " se kodiraju na način na koji se to radi po HTML standardu (& ≈...). Svi ostali znaci koji nisu saglasni sa US-ASCII standardom ne mogu se kodirati na ovaj način pošto ih XML neće prepoznati. Alternativno se oni mogu kodirati &#ddd, u decimalnoj ili ઼, u heksadecimalnoj notaciji. Ili pak mogu biti kodirani po UNICODE rasporedu UTF-8 koji je i podržan od strane XML-a.

XML deklaracija

Svaki dobro formatirani XML dokument mora da sadrži naredbu u kojoj je precizirana verzija XML-a koji je korišćen. Trenutno jedini validan način da se ovo

uradi je po W3C preporukama. Tako ovaj dokument mora da sadrži u prvoj liniji koda sledeće: `<?xml version="1.0"?>`

Referisanje XML DTD-a

Definisanje vrste dokumenta (Document Type Definition, DTD) se radi po važećem XML standardu.

```
<!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD
2002/07/31//EN"
"http://dublincore.org/documents/2002/07/31/dcmes-
xml/dcmes-xml-dtd.dtd">
```

Deklaracija RDF-a

Ovaj korak je neophodan kako bi aplikacije prepoznale da se koristi XML-RDF dokument. Prvo se piše `rdf:RDF` pa zatim XML tag koji opisuje `rdf` i XML tag koji opisuje DCMES elemente.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Opisivanje resursa

Ovaj način može opisati više resursa u jednom dokumentu. Opis svakog resursa počinje XML tagom `rdf:Description`

```
<rdf:Description>
...
</rdf:Description>
```

Unutar `rdf:Description`

Svaki DC elemenat unutar `Description` taga mora početi sa `dc`; element koji označava naslov se piše sa `dc:title`, sva slova su mala i nalaze se unutar `rdf::Description` taga. Kontejner u kojem su sadržani iskazi izgleda ovako:

```
<rdf:Description rdf:about="http://www.ncd.matf.bg.ac.yu/">
<dc:title>Nacionalni centar za digitalizaciju</dc:title>
</rdf:Description>
```

Ovo se primenjuje za sve DCMES elemente. Potpun izgled RDF-XML zaglavlja:

```
<?xml version="1.0"?> <!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN
CORE//DCMES DTD 2002/07/31//EN"
"http://dublincore.org/documents/2002/07/31/dcmes-
xml/dcmes-xml-dtd.dtd">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description>
<dc:title>Duble Core Metadata Initiative</dc:title>
<dc:title xml:lang="fr">L'Initiative de mÃ©tadonnÃ©es
```

```

du Dublin Core</dc:title>
<dc:creator>Marko Milenković</dc:creator>
<dc:format>pdf</dc:format>
<dc:identifier>ISBN 00000000</dc:identifier>
<dc:description>Kratak uvod u DCMI... </dc:description>
<xml:language>Sr</xml:language>
</rdf:Description>
</rdf:RDF>

```

Metapodatak sadržan u resursu

Ovaj način implementacije se najčešće koristi pri implementaciji DC preko HTML-a, ali nije nemoguće da se koristi i na drugi način. Metapodaci se upisuju unutar HTML fajla i služe za njegov opis. Ovako opisani fajlovi se mogu pretraživati preko pretraživačkog softvera (SWISH-E, ISEARCH, HARVEST...). Ova vrsta implementacije ima široku primenu, ali jednom standardizovan XML u saradnji s RDF-om pruža veću slobodu u opisu. Postoji mnoštvo alata koji su kreirani radi lakše manipulacije (dodavanje i menjanje metapodatka u telu podatka).

Kratko se osvrnimo na implementaciju DC preko HTML-a [HTML4. 0], korišćenjem META i LINK tagova (detaljnije u [10]). META tag se koristi kada želimo da opišemo postojeći metapodatak. Da bi se naznačilo da se koristi DC elementat, dopisuje se prefiks ispred DC elementa.

```
<meta name = "DC.Creator" content = "Milenković, Marko">
```

Tako ovaj tag kazuje da je kreator dokumenta Marko Milenković. U META tagu ime DC elementa se piše velikim početnim slovom. LINK tag služi da asocira prefiks skupa elemenata sa definicijom skupa elemenata.

```
link rel = "schema.DC"
href = http://purl.org/DC/elements/1.0/">
```

Konačan izgled opisa:

```

<html>
<head>
<title> Dublin Core </title>
<link rel = "schema.DC"
href = "http://purl.org/DC/elements/1.0/">
<meta name = "DC.Title"
content = "Dublin CORE">
<meta name = "DC.Creator"
content = "Milenković, Marko">
<meta name = "DC.Type"
content = "rad">
<meta name = "DC.Date"
content = "March, 2003">
<meta name = "DC.Format"
content = "html/pff">
<meta name = "DC.Language"
content = "sr">

```

```

</head>
<body><pre>
Metapodatak (metadata) je u upotrebi...
</pre></body>
</html>

```

Metapodatak sadržan van resursa

Ovakva implementacije je moguća u bilo kakvoj bazi podataka; metapodatak pruža link ka podatku. Ovo je pogodan način za netekstualne podatke. Ovaj način se implementira tako što se naprave dve tabele u bazi podataka: jedna sadrži metapodatke, a druga same podatke. Tako se pri pretrazi pretražuje tabela s metapodacima, i ako traženi pojam postoji u tabeli isčitava se link na sam podatak.

Lokalizacija

DC je uspešno lokalizovan na najmanje dvadesetak jezika. Sedam vlada koristi ovu tehniku čuvanja informacija (Australija, Kanada, Danska, Finska, Irska, Novi Zeland, Velika Britanija). Postoji poprilično dokumentacije o lokalizaciji na druge jezike. Korišćenje DC-a sa UNICODE (UTF-8) rasporedom znatno bi uvećalo mogućnost lokalizacije na srpski jezik (pogledati [16]).

Softver

Postoji dosta aplikacija u javnom vlasništvu, koje služe za kreiranje šablona, menjanje, "pakovanje" podataka... Postoje aplikacije koje kreiraju metapodatak iz HTML-a i MS-Office ekstenzija (doc...). Autor nije isprobao ni jedan od pronađenih programa i ne može da proceni njihov kvalitet i spretnost pri baratanju sa YUSCII skupom karaktera (pogledati [17]). Takođe ima dosta javno dostupne literature [18].

Dublin Core i Z39. 50

Postoji dogovoreni mehanizam koji omogućava da DC bude pretražen pod Z39.50 Preuređivanje DC, kako bi bio pretraživ u Z39.50, minimalno je.

Dodatak A: Rečnik DCMI tipova

Rečnik DCMI tipova pruža listu elemenata koji mogu da se koriste kao atribut **type** DC elementu.

Collection označava da je resurs predstavljen kao grupa, njegovim delovima se može pojedinačno pristupiti.

Dataset prikazuje da je informacija kodirana kao neka struktura (tabela, lista...)

Event opisuje netrajan, vremenski događaj.

Image je netekstualna vizuelna reprezentacija. Image može da se označava: fotografije, slike, dijagrame, mape, filmove...

InteractiveResource označava resurs koji zahteva interakciju od korisnika da bi bio razumljiv. Atribut opisuje formulare na webu, aplete, multimedijalno učenje..

Service je sistem koji obezbeđuje funkcije koje su značajne za krajnjeg korisnika (na primer servis za identifikaciju, web server, Z39.50...)

Software je kompjuterski program u prevedenom ili izvornom kodu.

Sound je resurs koji je primarno namenjen da bude tumačen kao zvuk.

Text je resurs namenjen za čitanje (knjige, pisma, novine...).

Dodatak B: Internet Media Types

U svakodnevnom radu, pretraživači weba moraju da znaju koju vrstu informacije primaju kako bi je pravilno obradili. Pretraživač prepoznaje dolazeću informaciju tako što mu web server najavi koju informaciju šalje. Server koristi elemente iz MIME skupa kako bi saopštio podatke o poslatoj informaciji, a ona je oblika *type/subtype*. Type precizira o kojoj vrsti informacije se radi i ona može da bude: tekst, slika, zvuk, video, aplikacija, višedelna i poruka (text, image, audio, video, application, multipart, message). Prva četiri elementa su jasna; aplikacija predstavlja binarni podatak, npr. instalaciju nekog programa. Multipart označava dokument s nekoliko različitih MIME tipova u sebi. Subtype daje detaljniji opis o tipu koji se šalje. Tako ako je type = text, subtype može biti rtf, html... Najčešće korišćeni MIME tipovi su:

text/plain: običan tekst; ako server ne uspe da prepozna vrstu fajla šalje ovu poruku.

text/html: HTML dokument.

text/richtext: Rich Text Format. Standard tekstualnih dokumenata podržan od mnogih programa za obradu teksta.

image/gif: gif format slika je uobičajan komprimovani format namenjen za razmenu slika između različitih platformi.

image/jpeg: još jedan popularan grafički format.

audio/basic: komprimovani audio fajl, obično ima ekstenziju .au.

audio/x-wav: Microsoft Windows audio format.

video/mpeg: MPEG komprimovani video.

video/quicktime : Quicktime video.

video/x-msvideo: Microsoft video, obično ima ekstenziju .avi.

application/octet-stream: Svaki binarni fajl koji server ne prepoznaje. Kada pretraživač primi ovakvu poruku od servera, on obično daje opciju čuvanja fajla.

application/postscript: Postskript fajlovi.

applicatio/pdf: Portable Document Format fajlovi.

Literatura i korisni linkovi

- 1 Diane Hillmann, *Using Dublin Core*, <http://dublincore.org/resources/translations/>
- 2 Dave Beckett, Eric Miller, Dan Brickley, *Expressing Simple Dublin Core in RDF/XML*, <http://www.dublincore.org/documents/dcmes-xml/>
- 3 Baker, Oktobar 2000, <http://www.dlib.org/dlib/october00/baker/10baker.html>
- 4 Ralph LeVan, *Dublin Core and Z39. 50*, <http://www.dublincore.org/documents/dc-z3950/>
- 5 Simon Cox, *DCMI Box Encoding Scheme: specification of the spatial limits of a place, and methods for encoding this in a text string*, <http://www.dublincore.org/documents/dcml-box/>
- 6 *Dublin Core Metadata Element Set, Version 1.1: Reference Description*, <http://dublincore.org/documents/dces/>
- 7 *Dublin Core Qualifiers*, <http://dublincore.org/documents/dcmes-qualifiers/>
- 8 *DCMI Type Vocabulary*, <http://dublincore.org/documents/dcml-type-vocabulary/>
- 9 Steven Holzner, *XML Complete*

- 10 J. Kunze, *Request for Comments: 2731*
- 11 *Dublin Core Metadata for Resource Discovery*,
<http://www.ietf.org/rfc/rfc2413.txt>
- 12 *Specification and Standardization of Data Elements*,
<ftp://sdct-sunsv1.ncsl.nist.gov/x3l8/11179/>
- 13 *Date and Time Formats*, <http://www.w3.org/TR/NOTE-datetime>
- 14 *Internet Media Types*, <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>
- 15 *Tags for the Identification of Languages*, <http://www.ietf.org/rfc/rfc1766.txt>
- 16 <http://dublincore.org/groups/languages/>
- 17 <http://www.dublincore.org/tools/>
- 18 <http://www.dublincore.org/documents/>

DUBLIN CORE METADATA INITIATIVE (DCMI)

We present an easy and flexible way of describing resources. The simplicity of the design and maintenance, understandable semantics and possibility of localization are the aims of Dublin Core Metadata Initiative. The implementation is possible through XML and HTML, the former one having greater possibilities. Description of a particular resource can be kept inside or out of it, which is convenient for describing nontextual information. Possibility for localizations is proved, since the governments of many countries introduced Dublin Core as the standard way of describing information.

Key words: metadata, XML, HTML