

Metadata, RDF

April 24, 2012

1 Rámce pro metadata popisující XML a jiné datové zdroje

1.1 Rámec RDF

RDF Model a Rdf Schema jsou doporučeními W3CSpecifikace a další informace pracovní skupiny - <http://www.w3.org/RDF>

1.2 RDF Model

RDF je obecný mechanismus pro specifikaci metadat je použitelný s libovolnými (i ne-digitálními) zdrojizákladem modelu jsou trojice:

- zdroj (resource) - např. <http://www.fi.muni.cz/tomp/xml>
- vlastnost (property) - např. popis
- hodnota (value) - např. Domovská stránka předmětu P138 na FI MU

Trojice je možné znázornit

- graficky,
- jako trojice (r, p, v) nebo
- XML syntaxí

Blíže viz

- Dobrý úvodní článek na [xml.com](http://www.xml.com/pub/a/2001/01/24/rdf.html): What is RDF? (<http://www.xml.com/pub/a/2001/01/24/rdf.html>)
- RDF Tutoriál - Zvon RDF Tutorial (<http://www.zvon.org/xxl/RDFTutorial/General/book.html>)
- RDF Tutoriál - W3Schools RDF Tutoriál (<http://www.w3schools.com/rdf/default.asp>)
- RDF Tutorial <http://www710.univ-lyon1.fr/~champin/rdf-tutorial/node1.html>
- Další RDF Tutorial (.ppt) (<http://www.aifb.uni-karlsruhe.de/WBS/sst/Teaching/Intelligente%20System%20im%20WWW%20SS%202000/RDF-Tutorial.pdf>)

1.3 RDF Schema

- Specifikuje omezení na množiny vlastností, jejich definičních oborů a oborů hodnot
- Modeluje se opět v RDF

1.4 RDF reprezentace užívaných metadatových schémat - Z39.50, Dublin Core atd.

- RDF je obecný rámec pro modelování metadat, pro konkrétní použití je obvykle nutné definovat *schéma* přípustných *vlastností*, jejich *domén* a množin (přípustných) *hodnot*.
- Tím se vytvoří RDF reprezentace daného metadatového schématu.
- Reprezentace může mít podobu *RDF Schematu*.

2 Dublin Core - příklad konkrétního metadatového schématu

2.1 Co je Dublin Core?

- je generické metadatové schéma s univerzální použitelností
- vznikl původně jako iniciativa knihovníků pro popis bibliografických informací
- dnes univerzálně používán - např. pro metadatový popis informací ve veřejné správě (*e-Government*)
- tvoří jej 15 základních elementů s rámcově definovanou sémantikou
- elementy je možné rozšiřovat - rozkladem na (obvykle disjunktní) podmnožiny (vždy to musí být podmnožiny některého z původních elementů)

2.2 Jednoduchý (Simple) Dublin Core

”Jednoduchý” nebo ”základní” Dublin Core (angl. Simple Dublin Core nebo Unqualified Dublin Core, dále jen ”jednoduchý DC”) představuje základní soubor patnácti prvků, který vyvinula a podporuje

- *Iniciativa pro metadata Dublin Core* (Dublin Core Metadata Initiative, DCMI, <http://dublincore.org>).
- Momentálně je aktuální verzí Dublin Core 1.1.
- je přijat konsorciem IETF (<http://ietf.org>) jako tzv. *dokument RFC (Request For Comment) 2431* rovněž od 2003 jako *ISO Standard 15836-2003*

2.3 Dublin Core - elementy

Název Jméno dané zdroji Tvůrce Entita primárně odpovědná za vytvoření obsahu zdroje Předmět a klíčová slova Téma obsahu zdroje Popis Vysvětlení obsahu zdroje Vydavatel Entita odpovědná za zpřístupnění zdroje Příspěvatel Entita, která přispěla k vytvoření obsahu zdroje Datum Datum spojené s určitou událostí během existence zdroje Typ zdroje Povaha nebo druh obsahu zdroje Formát Fyzická nebo digitální reprezentace zdroje Identifikátor zdroje Jednoznačný odkaz na zdroj v rámci daného kontextu Zdroj Odkaz na zdroj, z něhož je popisovaný zdroj odvozen Jazyk Jazyk intelektuálního obsahu zdroje Vztah Odkaz na příbuzný zdroj Pokrytí Rozsah nebo záběr obsahu zdroje Správa autorských práv Informace o právech vztahujících se k popisovanému zdroji

2.4 DC - příklad metadatového popisu

Název Zelená kniha o elektronickém obchodu Tvůrce Úřad pro veřejné informační systémy, Úřad vlády Předmět Elektronický obchod, elektronický podpis, bezpečnost, správa Popis Vládní návrh podpory elektronického obchodu v České republice Datum vytvoření 2001-09-20 Datum zveřejnění 2001-10-17 Identifikátor ISBN:?????

2.5 Kvalifikovaný Dublin Core

- (Qualified Dublin Core) obsahuje stejný soubor prvků jako jednoduchý DC a doporučuje další upřesnění a omezení každého prvku.
- Typicky se tak děje na základě formálního nebo de-facto mezinárodního standardu, např. může požadovat, aby prvek "jazyk" byl vyplněn v souladu se seznamem ISO pro jazyky (ISO 639).

2.6 Kódování DC v XML

DTD - <http://dublincore.org/documents/2001/11/28/dcmes-xml/dcmes-xml-dtd.dtd> (<http://dublincore.org/documents/2001/11/28/dcmes-xml/dcmes-xml-dtd.dtd>) XML Schema - <http://dublincore.org/documents/2001/11/28/dcmes-xml/dcmes-xml-xsd.xsd> (<http://dublincore.org/documents/2001/11/28/dcmes-xml/dcmes-xml-xsd.xsd>) RDF Schema - <http://www.dublincore.org/documents/2001/11/28/dcmes-xml/dcmes-xml-rdf-schema-cz.rdf> (/~tomp/xml/rdf/dc-rdf-schema-cz.rdf) RDF Schema pro slovník typů (Type Vocabulary) - /tomp/xml/rdf/dc-tv-rdf-schema-cz.rdf (/~tomp/xml/rdf/dc-tv-rdf-schema-cz.rdf)

2.7 Nástroje pro práci s RDF

Jena Java RDF API and toolkit (<http://www.hp1.hp.com/semweb/>) The ICS-FORTH RDFSuite (<http://139.91.183.30:9090/RDF/>) DC Creator na University of Bath (<http://www.ukoln.ac.uk/cgi-bin/dcdot.pl>) další viz <http://www.w3.org/RDF> (<http://www.w3.org/RDF/>)

3 Příklady praktického použití metadat - veřejná správa

3.1 Rámec pro metadata ISVS ČR

Kroky budování

- Přijmout doporučení **Dublin Core** a osvojit jej jako **Národní metadatový standard (NMS)**.
- Rozšířit tento standard tak, aby vyhovoval potřebám veřejné správy jak pro snadné vyhledávání informací, tak pro správu informačních zdrojů.
- Vyvinout **Aplikační profil NMS**, který bude obsahovat předepsaná kódovací schémata a závazný výklad jednotlivých metadatových prvků.
- Připravit **Tezaurus veřejné správy**.

3.2 Adaptace Dublin Core pro potřeby veřejné správy

pro potřeby veřejné správy v zemích Evropské Unie, Austrálie, Kanady a Nového Zélandu je rozpracováván specifický *aplikační profil* Dublin Core. Cílem MIREG je vytvořit metadatový rámec (metadata framework), příslušné referenční softwarové nástroje a soubor osvědčených postupů (best practice) pro implementaci rámce v jednotlivých zemích a sektorech. Přitom spolupracuje také s evropskou standardizační autoritou CEN, což dává předpoklad celoevropského respektování vzniklého doporučení.

- proces zahájen na sérii pracovních seminářů **Managing information resources for e-government** (MIREG) a stal se součástí programu *Interchange of Data between Administrations (IDA)* Evropské Unie.
- Dalším partnerem při vytváření evropského metadatového rámce je též projekt **ParlML**, zaměřený na zpřístupňování informací Evropského parlamentu.
- Příslušná pracovní skupina připravuje doporučení **DC-Gov Application Profile**

3.3 Aplikační profil NMS

zahrnuje:

- **Upřesnění** (zjemnění, kvalifikaci, specializaci angl. element refinement) metadatových prvků, které přesněji určuje sémantiku daného prvku a tím jej rozděluje na jemněji (přesněji) určené podprvky - např. obecné datum lze kvalifikací rozdělit na menší části, a místo "datum" uvádět přesněji např. "datum vytvoření", "datum zveřejnění", "datum platnosti", "nástupnické datum".
- Kvalifikovaný prvek lze však i nadále zpracovávat nástroji, které příslušné kvalifikaci "nerozumějí" - tyto nástroje potom chápou prvek jako by zůstal nekvalifikovaný (všeobecnější), tj. "datum zveřejnění" mohou chápat jako

prosté "datum", čímž je sice část sémantiky ztracena, ale prvek může být stále užitečný např. pro vyhledávání.

- **Kódovací schémata** (též kvalifikace hodnoty, angl. encoding scheme nebo value qualification) specifikující formát, ve kterém bude uložena hodnota pro příslušný metadatový prvek, např. "datum" vždy bude uváděno ve formátu *rrrr-mm-dd* (rok-měsíc-den), což definuje standard ISO 8601.
- Kromě formátu může být kvalifikací hodnoty též např. specifikace *měrné jednotky*, v níž bude hodnota uváděna.

4 Ontologie

4.1 Co jsou ontologie?

prostředek jak popisovat znalostimnožina pojmů a konstruktů, jak je odvozovat, spojovat atd. základní kategorie ontologií jsou

- **Classes** (general things) in the many domains of interest
- The **relationships** that can exist among things
- The **properties** (or **attributes**) those things may have

používá metadatové rámce (např. RDF), ale jebohatší s přesnější sémantikou jsou vybudovány obecné rámce pro tvorbu ontologií pro specifické domény

4.2 Aplikace ontologií (Use Cases)

- Webové portály, integrace dat na webu
- Multimediální kolekce
- Správa velkých webů
- Dokumentace návrhu
- Inteligentní agenti
- "Všudypřítomné počítání"

Pracovní skupina při W3C (<http://www.w3.org/2001/sw/WebOnt/>)

4.3 XML Topic Maps

Další návrh pracovní skupině WebOnt - <http://www.topicmaps.org/xm/1.0/>
(<http://www.topicmaps.org/xm/1.0/>)