

Luento 12: XML ja metatieto

AS-0.110 XML-kuvauskielten perusteet

Janne Kalliola

XML ja metatieto

- Metatieto
 - rakenne
 - sanasto
- Resource Description Framework
 - graafikuvaus
 - XML
- Semanttinen Web
 - agentit

Metatieto

- Metatieto on tietoa tiedosta
 - tarkoituksena kuvata sisältöä sovitulla tavalla
 - luodaan kontekstia tiedolle
 - metatiedon avulla voidaan
 - ryhmittää samankaltaisia sisältöjä
 - arvioida sisällön sopivuutta tietyille kohderyhmälle
 - hakea sisältöjä arkistosta
- Ilman metatietoa tietomassat ovat käyttökeltvottomia
 - koneet eivät ymmärrä semantiikkaa

3

Tieto ja metatieto

- Metatieto voi kuvata periaatteessa minkä tahansa muotoista sisältöä
 - teksti, kuvat, audio, video
- Metatieto täytyy kytkeä tietoon jollakin tavalla
 - toinen viittaa toiseen
 - tieto sisältää metatiedon
- Yleensä tiedon ja metatiedon välillä on yksi-yhteen – suhde
 - periaatteessa tietosirpale voi viitata useampaan metatietosirpaleeseen ja yksi metatietosirpale voi määrittää useampaa tietosirpaletta

4

Metatiedon rakenne

- Metatieto koostuu yleensä avain-arvopareista
 - avain on jokin tietoa kuvaavan tyyppin nimi
 - muutospäivämäärä
 - kirjoittaja, hyväksyjä, toimittaja yms.
 - arvo on avaimen sidottu kuvaus
 - yhdellä avaimella voi olla yksi tai useita arvoja
 - riippuu avaimen määritelmästä

Sanasto

- Sanasto määrittää sallittujen avainten joukon
 - myös avainten arvoille voidaan määrittää rajoituksia
 - edistyneemmissä sanastoissa on perintämahdollisuuksia
 - nopeuttaa uusien avain-arvoparien laadintaa



Resource Description Framework

- Määrittelykieli
 - W3C:n suositus standardiksi
 - esityskieliriippumaton
 - yksi esitysmalli pohjautuu tällä hetkellä XML:en
- Metatieto esitetään lausekkeilla
 - kuvaavat dokumenttien ominaisuuksia ja keskinäisiä suhteita
 - dokumenttia kutsutaan resurssiksi
 - voi sisältää mitä tahansa tietoa, kunhan on viitattavissa URI:lla

8

RDF:n tarkoitus

- Määritellä yleinen mekanismi, jolla resursseja voidaan kuvata ilman oletuksia aihe- tai sovellusalueesta
 - ei ota kantaa käytettyyn sanastoon
 - sanastoa kutsutaan RDF Schemaksi
- RDF on itsessään abstrakti ja formaali määrittely, jolla voidaan kuvata avoimia metatietorakennelmia
 - perustuu yksinkertaiseen graafimalliin

RDF:n rakenne

- RDF-metatieto koostuu kolmikoista (triplet):
 - subjekti – määrittelyn kohde
 - objekti – määrittelyn arvo
 - predikaatti – määriteltävä ominaisuus
- Kolmikossa objekti sidotaan subjektiin predikaatilla
 - eli objekti määrittää subjektia
 - predikaatti määrittää määrittelyn tyyppin
 - objekti voi olla myös toinen subjekti
 - tällöin syntyy puu

RDF:n graafiesitys

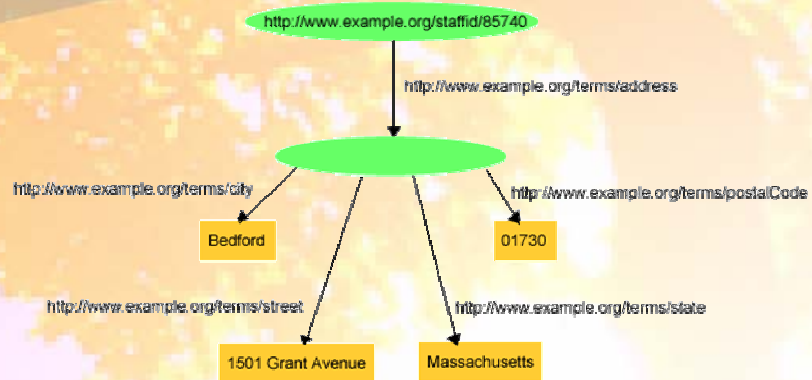
- RDF-lauseke voidaan piirtää alla olevana graafina
 - nuoli on aina subjektista objektiin



Määrittelyt

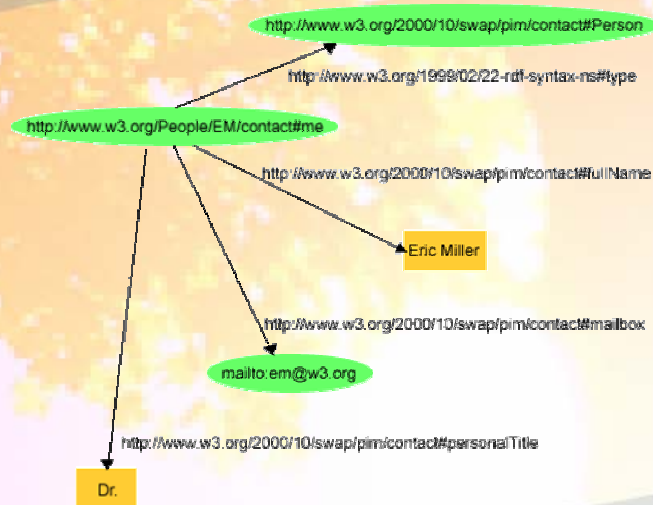
- Subjekti voi olla joko URI-viittaus tai tyhjä solmu
- Objekti voi olla URI-viittaus, literaali tai tyhjä solmu
- Predikaatti on URI-viittaus
- URI-viittauksen avulla RDF:ssä voidaan viitata määritettäviin asioihin
 - URI-viittauksia täytyy ajatella sanastoina
 - RDF määrittää joukon URI-viittauksia
 - metatietosanasto määrittää metatiedossa käytettävät URI-viittaukset

Graafiesimerkki URI:n kanssa



- Tyhjä subjetti kokoaa yhteen osoitteen

Toinen graafiesimerkki



RDF ja XML

- W3C on määrittelyt RDF:n abstraktin mallin lisäksi erillisen RDF-XML –sarjallistuksen
 - RDF:lla kuvattu malli kirjoitetaan XML-dokumenttina
 - URI-viittaukset muuttuvat nimiavaruuksiksi
- RDF käyttää tämän lisäksi XML Schemassa määriteltyjä tietotyyppisiä

15

XML-esimerkki

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ex="http://example.org/stuff/1.0/">

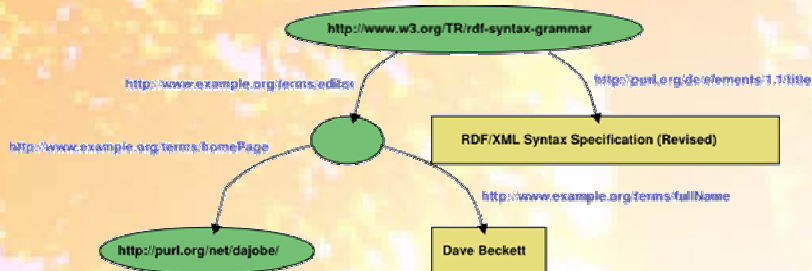
  <rdf:Description
    rdf:about="http://www.w3.org/TR/rdf-syntax-grammar"
    dc:title="RDF/XML Syntax Specification (Revised)">

    <ex:editor>
      <rdf:Description ex:fullName="Dave Beckett">
        <ex:homePage
          rdf:resource="http://purl.org/net/dajobe/" />
        </rdf:Description>
      </ex:editor>

    </rdf:Description>
  </rdf:RDF>
```

16

XML-esimerkki graafina



RDF:n käyttö

- RDF-dokumentit luetaan koneellisesti XML-muodossa
 - graafit ovat pitkälti ihmisiä varten
- RDF-dokumentti lisää koneymmärrettävää tietoa määrittämäänsä dokumenttiin tai muuhun resurssiin
 - käytetty sanasto (ts. URI-viittaukset) vaihtelevat käyttötarkoituksesta toiseen
 - RDF-dokumentin pohjalta sovellus voi esimerkiksi kytkeytyä toiseen palveluun tai korostaa dokumentin esitystä jollakin tavalla
- RDF ei ota itsessään kantaa siihen, kuinka sitä pitää käyttää
- Tällä hetkellä RDF-tuki on vielä kovin hajanaista
 - verkossa on kuitenkin useita mielenkiintoisia RDF:a käyttäviä projekteja, kuten www.foaf-project.org

RDF ja WWW

- RDF-dokumentteja voidaan helposti siirtää WWW-ympäristössä
 - RDF-määrytykset voidaan myös sisällyttää toiseen dokumenttiin
 - esimerkiksi XHTML-dokumentti voi sisältää RDF-määrytyksiä
 - nämä määrytykset eivät näy välttämättä koskaan käyttäjälle, mutta selain tai plug-in osaa tulkita niitä ja muokata toimintaansa

Semanttinen Web

Semanttinen web

- Semanttinen web on kehys (framework), jonka avulla tietoa voidaan jakaa helposti
 - sovellusten välillä
 - yritysten välillä
 - yhteisöjen välillä
- Semanttisen webin ajatuksena on mahdollistaa koneiden välinen tiedonvaihto metatiedon avulla
 - metatieto kuvataan RDF:na
 - käytetään erilaisia luokitteluja (ontologia)
- Yleensä siirrettävät dokumentit ovat XML-pohjaisia

21

Semanttisen webin käyttöympäristö

- Semanttinen web ei ole sama WWW-ympäristö, johon olemme tottuneet
 - periaatteessa esimerkiksi puhelin ja stereot voivat keskustella semanttisen webin yli: stereot hiljentävät äänenvoimakkuutta puhelimen soidessa
 - stereot ja puhelin eivät välttämättä ole kuitenkaan sovitettu puhumaan keskenään
- Oleellista on välittää riittävästi semanttista tietoa (rakennetta ja metatietoa), jotta kone pystyy tulkitsemaan vastaanottamaansa tietoa
 - sama tietosisältö voidaan esittää myös ihmiselle
 - ihminen osaa tulkita sisällön suoraan

22

Agentit

- Agentit ovat ohjelmia, jotka tekevät pieniä (ja myöhemmin ehkä suuriakin) toimintoja ihmisten puolesta
 - toiminta pohjautuu metatietoon
 - palvelusta
 - käyttäjästä (ihmisestä, agentin isännästä)
 - agentti pystyy tekemään yksinkertaisia valintoja
 - esimerkiksi valitsemaan sopivan erikoisruokavalion lennolle
- Tällaiset toimenpiteet vaativat nykyistä tiukempaa turvallisuutta ja siksi digitaaliset allekirjoitukset ovat tärkeä osa semanttista webiä

Kysymyksiä? Kommentteja?