

OWL - Web Ontology Language

Hansi Keijonen

Seminaariraportti
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Helsinki, 1. maaliskuuta 2013

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Hansi Keijonen			
Työn nimi — Arbetets titel — Title			
OWL - Web Ontology Language			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages	
Seminaariraportti	1. maaliskuuta 2013	10	
Tiivistelmä — Referat — Abstract			
Tiivistelmä.			
Avainsanat — Nyckelord — Keywords			
avainsana 1, avainsana 2, avainsana 3			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

1	Semanttinen web	3
2	Teknologiat ja kielet jotka mahdollistavat OWL:n	3
2.1	URI ja XML	3
2.2	Tiedon esittäminen RDF-triploilla	4
2.3	Alkeelliset ontologiat RDF Schemalla	5
2.3.1	Luokat ja ilmentymät	6
2.3.2	Luokan aliluokka	6
2.3.3	Ominaisuudet ja periytetyt ominaisuudet	6
2.3.4	Rajoitukset ominaisuuksien sovellusalueissa ja arvoissa	7
2.3.5	RDF Schema esimerkki	7
3	Kehittyneitä ontologioita OWL:llä	7
3.1	OWL:n kolme alikieltä	8
3.2	OWL-ontologian rakenne	9
3.2.1	Nimiavaruudet	9
4	OWL2	9
	Lähteet	9

1 Semanttinen web

Suurin osa tämän päivän webin sisällöstä on tarkoitettu ihmisten luettavaksi sekä tulkittavaksi. Kone pystyy tulkitsemaan esim. html-tiedoston ja esittämään dokumentin siinä määritellyllä tavalla. Ongelma on, että kone ei *ymmärrä* dokumentin sisällön merkitystä, semantiikkaa [2]. Se, että kone ei ymmärrä dokumenttien semanttisia merkityksiä rajoittaa esimerkiksi haut internetissä olevista dokumenteista yksinkertaiseksi hakusanojen etsimiseksi. Sen sijaan jos hakukoneet ymmärtäisivät asioiden merkityksen ja niiden välillä vallitsevat yhteydet, olisi hakukoneiden hakutulokset tarkempia ja sisältäisivät mahdollisesti laajennettuja hakuja alkuperäisen asian ympäriltä [2]. On siis tarve olla menetelmä käsitteiden luomiseen, käsitteiden ominaisuuksien kuvaamiseen sekä käsitteiden välisten suhteiden kuvaamiseen [2]. Tim Berners-Lee, James Hendler ja Ora Lassila toteavat artikkelissaan "Semantic web", että "semanttinen web ei ole erillinen web vaan laajennos tämänhetkiseen webiin, jossa informaatiolle on annettu hyvin muotoiltu merkitys mahdollistaen koneiden ja ihmisten paremman yhteistyön." web of documents -> web of data dataa voi parsia manuaalisesti tai koneellisesti

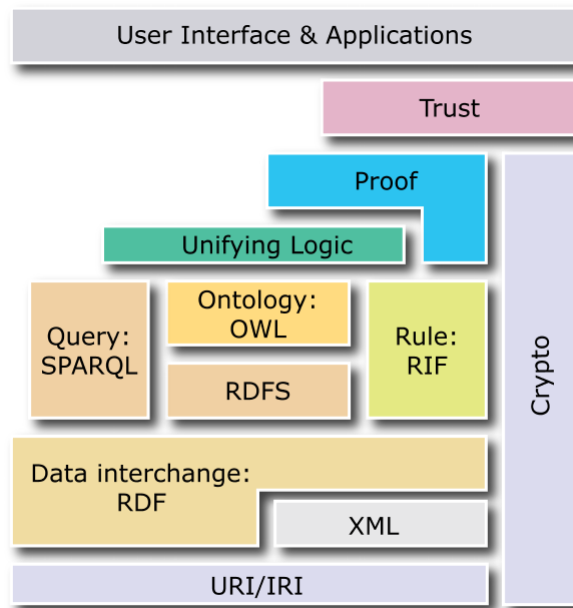
2 Teknologiat ja kielet jotka mahdollistavat OWL:n

W3c tarjoaa semanttisen webin toteuttamiseen standardit teknologioista ja kielistä. Kuvassa 1 on semanttisen webin teknologiapino sekä ajatuskonseptteja semanttisen webin toteuttamiseen. Osa teknologioista on todellisuutta ja käytössä, osa vasta ideatasolla. Jokainen kerros käyttää alemman kerroksen palveluita. Seuraavissa kappaleissa käyn läpi kaavion teknologioita ja kieliä alhaalta ylöspäin kohti OWL:ää. Jokaisesta seuraavissa kappaleissa esitetystä teknologiasta käsitellään tarkemmin ne konstruktiot, jotka ovat olennaisia ja käytössä myös OWL:ssä.

2.1 URI ja XML

Semanttisessa webissä luokkia, ilmentymiä, ominaisuuksia ja ominaisuuksien arvoja kutsutaan resursseiksi. Jotta sekaannusta jo määritettyjen resurssien sekä uusien määritysten kanssa ei syntyisi, identifoidaan kaikki resurssit (pl. ominaisuuksien literaaliarvot) yksilöllisesti URI(Unified Resource Locator):lla. URI:n avulla voidaan viitata mihin tahansa määritettyyn resurssiin. Useinmiten URI:n toimii perinteinen URL(Unified Resource Locator)-osoite [2]. IRI (Internationalized Resource Identifier) on ainoastaan merkistölajajennos URI:in.

Semanttisen webin datan kuvaukset toteutetaan useimmiten XML-tiedostoina. XML-kieltä voidaan käyttää monimutkaisen rakenteisen tiedon esittämiseen ja tarjoaa näin standardoidun mallin tiedon vaihtamiseen prosessoijien välillä.

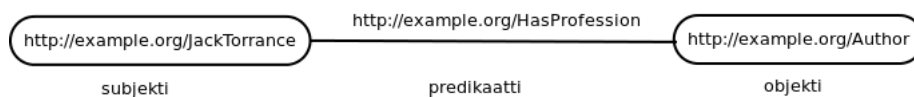


Kuva 1: Semanttisen webin toteutukseen tarvittavat teknologia, kielet ja konseptit.

Tärkeä XML:n ominaisuus on nimiavaruudet, jotka mahdollistavat resursien identifiointin URI:en avulla. Tästä kuitenkin enemmän seuraavissa kappaleissa.

2.2 Tiedon esittäminen RDF-triploilla

Resource Description Framework RDF on kieli webissä olevien resurssien kuvaamiseen. RDF perustuu asioiden identifiointiin URI:lla ja näiden asioiden kuvaamiseen ominaisuuksilla ja ominaisuuksien arvoilla. Tämä mahdollistaa yksinkertaisten lausumien esittämisen verkkoina, joissa resurssit ja ominaisuudet ovat soluja ja ominaisuudet verkon kaaria [4]. Kuvassa 2 on havainnollistettu asiaa.



Kuva 2: RDF-tripla joka kuvaa yksinkertaisen lausuman. Jokainen triplan solmu ja kaari on identifioitu URI:lla.

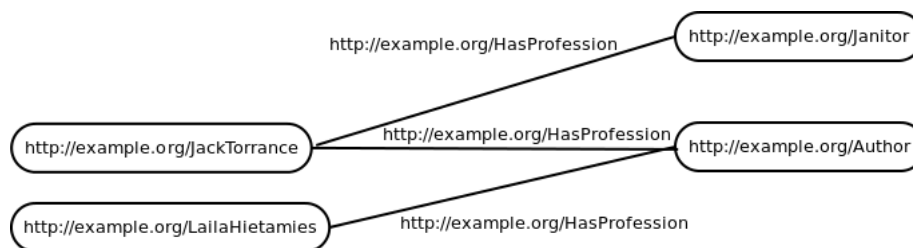
Kuvassa 2 on havainnollistettu yksinkertainen *RDF-tripla*. Triplan subjekti, predikaatti ja objekti kertovat, että "Jack Torrance on ammatiltaan kirjailija". Jokainen solu ja kaari on esimerkissä identifioitu URI:lla. Objekti

voi olla myös literaali, jolloin sitä ei identifioida erikseen, mutta formaatti määritellään ****NIIN VITTU MILLÄ?*****

Yleisin tapa esittää tripla on XML-notaatio. Myös muut tavat ovat mahdollisia, kuten esimerkiksi *JSON* ja *turtle*. Alla on esitetty kuvassa 2 triplan XML-notaatio:

```
<rdf:Description rdf:about="http://example.org/JackTorrance">
  <rel:HasProfession rdf:resource="http://example.org/Author"/>
</rdf:Description>
```

Esimerkistä näkee selvästi, kuinka subjekti, predikaatti ja objekti ovat identifioitu URI:lla. Jack Torrance on toiselta ammatiltaan talonmies ja myös Laila Hietamies on ammatiltaan kirjailija. Kun myös nämä triplat kuvataan samaan kaavioon, alkaa pieni mutta informatiivinen verkkkuva ja selitys triploista ja URI:sta? Kyllä, tähän se sopisi kokoavana elementtinä. o syntyä, kuva 3. Tästä verkosta voisimme tehdä hakuja, kuten "ketkä ovat kirjailijoita?".



Kuva 3: RDF-triplat muodostavat verkon.

RDF:n peruskonstruktioita ovat: description, resource, about ***selitä nämä*****

RDF tarjoaa siis vain melko alkeellisen tavan esittää lausumia, jotka muodostavat haut mahdollistavan verkon. RDF toteuttaa myös joukon muita ominaisuuksia, kuten *säiliöitä* (container) tiedon säilömiseen sekä *kokoomatietorakenteita* (collections) asioiden listaamiseen [4]. Näitä käsitellään myöhemmin niiltä osin kuin ne ovat relevantteja OWL-ontologioiden muodostamisessa.

2.3 Alkeelliset ontologiat RDF Schemalla

Semantiikkaa voidaan webissä ilmaista ontologioilla. Tietojenkäsittelytieteessä ontologialla tarkoitetaan dokumenttia, jossa kerrrotaan asioiden välisistä yhteyksistä [2]. ***vajaa läppä!! kuuluuko edes tähän???!*****

RDF:llä ilmaistut ominaisuudet voidaan ajatella resurssien attribuuteiksi samassa mielessä kuin perinteiset attribuutti-arvo -parit [3]. Vaikka RDF:llä

voidaan kuvata resursseja, se ei tarjoa keinoja kuvata ominaisuuksia tai ominaisuuksien välisiä suhteita. Tämä on mahdollista RDF:n *sanastonkuvauslaajennoksella* RDF Schemalla. RDF Schemalla on mahdollista määritellä luokkia ja ominaisuuksia joita voidaan käyttää luokkien, ominaisuuksien ja resurssien kuvaamiseen [3]. RDF Schema on RDF:n semanttinen laajennos.

Seuraavassa esitellään RDF Scheman niitä peruskonstruktioita, joita myös OWL käyttää määrittelyssään lähes sellaisenaan. *RDFS Schema määrittelee myös suurehkon joukon muita konstruktioita, mutta niiden esittely OWL:n selostamisen kannalta on tässä tarpeetonta.*

2.3.1 Luokat ja ilmentymät

Resurssit voidaan jaotella luokkiin. Luokkien jäseniä nimitetään luokan ilmentymiksi [3]. Luokat ovat itsessään resursseja. Useimmiten luokat identifioidaan RDF:n URI:lla ja niitä voidaan kuvata RDF:n ominaisuuksilla (property). Luokka ilmaistaan `rdfs:Class` tagilla, `rdf:type` ominaisuudella voidaan ilmaista, että resurssi kuuluu luokkaan [3]. Esimerkiksi kustannusosakeyhtiö "WSOY" on itse luokka ja luokan "Kustantamo" ilmentymä:

```
Esim:WSOY rdf:type rdfs:Class
Esim:WSOY rdf:type esim:Kustantamo
```

2.3.2 Luokan aliluokka

`rdfs:subClassOf` ominaisuudella voidaan ilmaista, että jonkin luokan ilmentymät ovat myös jonkin toisen luokan ilmentymiä [3]. Esimerkiksi luokka "Kustantamo" on luokan "Liikkeyritys" aliluokka:

```
esim:Kustantamo rdfs:subClassOf esim:Liikkeyritys
```

2.3.3 Ominaisuudet ja periytetyt ominaisuudet

Resursien välisiä suhteita kuvataan ominaisuuksilla. Esimerkiksi "Matilla on veli Teppo". Voidaan ajatella, että resurssien "Matti" ja "Teppo" välinen suhde on ominaisuus "Veljeys". Ominaisuus "Veljeys" voi olla ominaisuuden "Perhesuhde" aliluokka tai paremminkin aliominaisuus. Tällaisia konstruktioita kuvataan RDF Schemassa termillä `subPropertyOf`:

```
esim:Veljeys rdfs:subPropertyOf esim:Perhesuhde
```

`subPropertyOf` ilmaisee määritelmällisesti, että resurssit joiden välistä suhdetta kuvataan jollain suhteella, voidaan kuvata myös tämän suhteen ylisuhteella, josta ko. suhde on periytynyt [3].

2.3.4 Rajoitukset ominaisuuksien sovellusalueissa ja arvoissa

RDF Schemassa on mahdollista antaa ominaisuuksille rajoituksia sen suhteen, että minkä luokkien välistä suhdetta ominaisuus kuvaa. Ensinnäkin voidaan rajoittaa suhteen sovellusaluetta (domain). Suhteen "Työskentelee" sovellusalue voidaan rajata koskemaan ainoastaan luokkaa "Työntekijä". Samaten suhteen saamat arvot (range) voidaan rajata olemaan ainoastaan luokan "Yritys" tai sen aliluokkien ilmentymiä.

```
esim:Työskentelee rdfs:domain esim:Työntekijä
esim:Työskentelee rdfs:range esim:Yritys
```

2.3.5 RDF Schema esimerkki

Myös RDF Scheman varsinainen notaatio on toteutettu XML:llä. Alla on lyhyt esimerkki RDF Schemalla toteutetusta ontologiasta.:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

  <rdfs:Class rdf:ID="liikeyritys" />

  <rdfs:Class rdf:ID="kustantamo">
    <rdfs:subClassOf rdf:resource="#liikeyritys"/>
  </rdfs:Class>

</rdf:RDF>
```

Esimerkissä on yksinkertainen ontologia, joka kertoo, että "kustantamo" ja "liikeyritys" ovat luokkia ja että "kustantamo" on myös luokan "liikeyritys" aliluokka. RDF Schema -ontologiat ovat RDF-tiedostoja samoin kuin OWL-ontologiatkin. Tiedoston alussa olevat nimiavaruusmäärittelyt varmistavat, että rdf- ja rdfs-alkuisilla tageilla ympäröidyt elementit todellakin ovat rdf- ja rdfs- elementtejä. Nimiavaruudet käsitellään takemmin OWL:n esittelyn yhteydessä.

3 Kehittyneitä ontologioita OWL:llä

Tällä hetkellä kehittynein tapa ilmaista ontologioita on OWL-ontologiat. Lyhenne OWL tulee sanoista Web Ontology Language. Vaikka määritelmässä on sana language, kieli, on OWL-ontologiat ymmärrettävä ennemminkin sanastoina, joita on kuvattu RDF-kielellä. Eräs tapa hahmottaa RDF-triplojen ja OWL-ontologioiden välinen ero on verrata niitä perinteiseen relaatiotietokantaan. RDF-triplat on tapa tallettaa tietoa olioiden ominaisuuksista samalla

tavalla kuin relaatiotietokannan taulun riveillä tallennetaan rakenteista tietoa tietokantaolioista. Jokaista riviä relaatiotietokannassa yksilöi yksilöivä avain kun taas RDF-triploissa avaimen virkaa hoitaa URI. Relaatiotietokannoissa tietokantaolion attribuuttien suhteita ilmaistaan taulurakenteilla ja tietokantaolioiden suhteita toisiin tietokantaoloihin ilmaistaan viitteillä taulujen välillä. Vastaavasti voidaan ajatella, että OWL-ontologiat kuvaavat ja jäsentävät samalla tavalla RDF-triploilla ilmaistua tietoa: ontologialla voidaan määritellä monipuolisesti luokkien, ilmentymien ja suhteiden ominaisuuksia. RDF on kuin kieli jolla ilmaistaan lausumia asioista kun taas OWL on sanasto, jonka avulla lausumien merkitykset ymmärretään.

OWL tarjoaa ontologioiden määrittelyyn [1]

- *hyvin määritellyn syntaksin* , jotta ontologiat olisivat koneluettavissa
- *hyvin määritellyn semantiikan*, jolla voi ilmaista merkityksiä tarkasti ja konsistentisti
- *tuen koneelliselle päättelylle*, jotta esim. ontologioiden eheys voidaan tarkastaa automaattisesti
- *riittävästi ilmaisuvoimaa* ilmaisemaan kaikki tarvittavat merkitykset
- *miellyttävän ilmaisutavan*, jotta työskentely olisi sujuvaa

Idealisesti OWL on RDF:n ja RDF Scheman laajennos [1]. OWL käyttää RDF:n luokkia ja suhteita lisäten niihin omia laajennoksiaan. RDF Schemassa on joitain hyvin vahvoja konstruktioita, kuten rfd:Class (kaikkien luokkien yliluokka) sekä rfd:Property (kaikkien suhteiden yliluokka). Näiden primitiivien ilmaisuvoima yhdistettynä OWL:n tarjoamaan laajennoksiin on ristiriidassa sen tavoitteen kanssa, että ontologiat olisivat koneellisesti pääteltävissä. Tämä tasapainotila mielessäpitäen on määritelty kolme OWL:n alikieltä sen perusteella, painotetaanko ilmaisuvoimaa vai koneellista päättelyä [1].

3.1 OWL:n kolme alikieltä

W3C:n Web Ontology Working Group on määritellyt OWL:lle kolme alikieltä, joiden on tarkoitus toteuttaa eri aspektit (ilmaisuvoima, koneellinen päättely), joita ontologioiden kuvaamiskieleltä vaaditaan [5]. Alikielien ovat ilmaisuvoiman mukaisesti kasvavassa järjestyksessä:

- *OWL Lite* tarjoaa ainoastaan yksinkertaisen luokitteluhierarkian ja yksinkertaiset rajoitteet[5]. Kardinaalisuusrajoitteet ovat ainoastaan muotoa 0 ta 1. Työkalutuen tarjoaminen on OWL Litelle helpompaa kuin ilmaisuvoimaisemmille versioille [5].

- *OWL DL* on tarkoitettu niille käyttäjille, jotka haluavat mahdollisimman hyvän ilmaisukyvyn siten, että ontologia on koneellisesti pääteltävissä (kaikki päätelmät tehtävissä järjellisessä ajassa) [5]. OWL DL tarjoaa kaikki kielen konstruktiot, mutta niitä voi käyttää tietyin rajoituksin, esimerkiksi luokka voi olla monen luokan aliluokka mutta ei voi olla samalla luokan ilmentymä. DL tulee sanoista Description Logics, deskriptiivinen logiikka, joka on tieteenala, joka tutkii logiikkaa ja on OWL:n perusta [5].
- *OWL Full* on tarkoitettu niille käyttäjille, jotka haluavat maksimaalisen ilmaisukyvyn välittämättä siitä, onko ontologiat enää koneellisesti pääteltävissä [5]. Toisin kuin OWL DL:ssä, luokka voi olla kokoelma yksilöitä (instansseja) samalla kuin luokka itsessään on jonkin luokan yksilö. OWL Full mahdollistaa jo olemassa olevien ontologioiden laajentamisen. On epätodennäköistä, että mikään ohjelmisto pystyy täydellisesti päättelemään OWL Full ontologioita [5].

3.2 OWL-ontologian rakenne

3.2.1 Nimiavaruudet

XML-dokumentissa tulee määritellä nimiavaruudet (namespace). Nimiavaruuksien avulla voidaan ratkaista mm. samannimisten elementtien aiheuttamia tulkintaongelmia sekä kertoa lukijalle (koneelle tai ihmiselle) konteksti, jonka mukaan elementtien kahvoja tulee tulkita. OWL-ontologiassa nimiavaruudet määritellään rdf:RDF -kahvojen sisään. Alla olevassa esimerkissä on eräs mahdollinen nimiavaruusmäärittely [2].

```
<rdf:RDF
  xmlns = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
  xmlns:vin = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
  xml:base = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
  xmlns:food = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/food#"
  xmlns:owl = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd = "http://www.w3.org/2001/XMLSchema#">
```

Esimerkin nimiavaruusmäärittelyissä on määritelty oletusnimiavaruus niille kahvoille, joissa ei käyteä mitään etuliitettä sekä nimiavaruudet niille kahvoille, joissa käytetään etuliitteitä vin: tai food:. Nimiavaruudet on määritelty myös owl:-, rdf:- ja rdfs:-etuliitteille kertomaan, että näillä etuliitteillä varustetut kahvat edustavat OWL:n, RDF:n ja RDFS:n termistöä. OWL-ontologia on riippuvainen myös XMLSchema-datatyypeistä, joten myös niiden nimiavaruus tulee määrittää.

4 OWL2

Lähteet

- [1] Antoniou, Grigoris ja Harmelen, Frank Van: *Web Ontology Language: OWL*. Handbook on ontologies, 2(September):91–110, 2009. <http://www.springerlink.com/index/10.1007/978-3-540-92673-3>.
- [2] Berners-Lee, Tim, Hendler, James ja Lassila, Ora: *The Semantic Web*. Scientific American, 284(5):34–43, 2001. <http://www.nature.com/doifinder/10.1038/scientificamerican0501-34>.
- [3] Brickley, Dan ja Guha, R V: *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C Recommendation, 2009(10 February 2004), 2004. <http://www.w3.org/TR/rdf-schema/>.
- [4] Manola, Frank ja Miller, Eric: *RDF Primer*. W3C Recommendation, 10(February):1–107, 2004. <http://www.w3.org/TR/rdf-primer/>.
- [5] McGuinness, Deborah L ja Van Harmelen, Frank: *OWL Web Ontology Language Overview*. W3C recommendation, 10(February):1–22, 2004. <http://www.w3.org/TR/owl-features/>.
- [6] Smith, Michael K, Welty, Chris ja McGuinness, Deborah L: *OWL Web Ontology Language Guide*. W3C Recommendation, 10(February):1–46, 2004. <http://www.w3.org/TR/owl-guide/>.