

**Miika Nurminen**

# **Semanttiset verkot ja niiden web-sovellukset**

Tietotekniikan (ohjelmistotekniikka)  
LuK -tutkielma  
22.5.2003

**Jyväskylän yliopisto**  
Tietotekniikan laitos

# Tietoja tutkielmasta

**Tekijä:** Miika Nurminen (minurmin@cc.jyu.fi)

**Työn nimi:** Semanttiset verkot ja niiden web-sovellukset.

**Työ:** LuK-tutkielma.

**Title in english:** Semantic networks and their web applications.

**Tiivistelmä:** Semanttiset verkot ovat formalismeja tietämyksen esittämiseen. Semanttisia verkkoja on sovellettu asiantuntijajärjestelmissä, luonnollisen kielen käsittelyssä ja muissa tekoälyä hyödyntävissä sovelluksissa. Semanttinen web on joukko teknologioita, joiden tarkoitus on saada WWW:n sisältö ja merkitykset helpommin tietokoneen ymmärrettävään muotoon. Tutkielma selvittää semanttisen webin suhdetta olemassaoleviin tietämystekniikan sovelluksiin erityisesti semanttisten verkkojen näkökulmasta.

**Avainsanat:** Aihekartat, käsitegraafit, logiikka, ontologia, RDF, tekoäly, tietämyksen esittäminen, tietämystekniikka, semanttinen web, semanttiset verkot, XML.

**Keywords:** Artificial intelligence, conceptual graphs, logic, knowledge engineering, knowledge representation, ontology, RDF, semantic networks, semantic web, topic maps, XML.

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Semanttiset verkot tietämyksen esittämisessä</b>	<b>2</b>
2.1	Tietämyksen esittämisestä . . . . .	2
2.1.1	Tavoitteet ja läheiset tutkimusalueet . . . . .	2
2.1.2	Tietämyskanta osana asiantuntijajärjestelmää . . . . .	4
2.1.3	Ontologiat . . . . .	5
2.2	Verkkoformalismeja . . . . .	6
2.2.1	Historiaa: puoliformaalit verkot . . . . .	7
2.2.2	Kehykset ja terminologialogiikat . . . . .	8
2.2.3	Käsitegraafit ja KIF . . . . .	10
2.3	Semanttisten verkkojen yhteys logiikkaan . . . . .	12
2.3.1	Logiikan laajennuksia . . . . .	12
2.3.2	Eksistentiaalisista graafeista käsitegraafeihin . . . . .	14
2.3.3	Verkkorakenne vai algebrallinen notaatio? . . . . .	17
<b>3</b>	<b>Semanttinen web</b>	<b>19</b>
3.1	Miksi semanttinen web? . . . . .	19
3.1.1	WWW:n kehitysaskeleet . . . . .	19
3.1.2	Semanttisen webin tavoitteet . . . . .	20
3.2	Semanttisen webin perustekniikat . . . . .	21
3.2.1	Metadata ja RDF . . . . .	22
3.2.2	RDFS ja ontologiat . . . . .	24
3.2.3	Päättely ja luottamus . . . . .	25
3.3	Aihekartat . . . . .	26
3.3.1	Aiheet, ilmentymät ja assosiaatiot . . . . .	26
3.3.2	RDF:n ja aihekarttojen esityskyvystä ja muunnoksista . . . . .	27
<b>4</b>	<b>Semanttiset verkot webin osana</b>	<b>29</b>
4.1	Tietämyksen esittäminen Web-sovelluksissa . . . . .	29
4.1.1	RDF(S) ja aihekartat semanttisina verkkoina . . . . .	29
4.1.2	Ontologioiden jakaminen Webissä . . . . .	31
4.1.3	Semanttisen webin mahdollisuuksista ja riskeistä . . . . .	32
4.2	Sovelluksia . . . . .	34
4.2.1	Cyc . . . . .	34
4.2.2	WordNet . . . . .	35
4.2.3	Omnigator . . . . .	36
<b>5</b>	<b>Yhteenveto</b>	<b>37</b>
	<b>Lähteet</b>	<b>39</b>
	<b>Lisämateriaalia</b>	<b>42</b>

# 1 Johdanto

Semanttiset verkot (*semantic networks*) ovat joukko toisiaan lähellä olevia formalismeja tietämyksen esittämiseen. Semanttisia verkkoja on sovellettu asiantuntijajärjestelmissä, luonnollisen kielen käsittelyssä ja muissa tekoälyä hyödyntävissä sovelluksissa. Matemaattinen perusta semanttisille verkoille on verkkoteoriassa ja predikaattilogiikassa (tosin semanttisissa verkoissa tehtävän päättelyn täsmällisyys vaihtelee formalismista riippuen). Semanttisia verkkoja on tekoälyn lisäksi käytetty kielitieteessä ja psykologiassa yli 40 vuoden ajan, nyt ne ilmenevät myös uusissa web-sovelluksissa semanttisen webin muodossa.

Semanttinen web tarkoittaa joukkoa teknologioita, joiden tarkoitus on saada WWW:n sisällön merkitykset tietokoneen ymmärrettävään muotoon. Semanttisen webin teknologiat ja standardit on määritelty pääosin XML-kielen (*eXtensible Markup Language*) sovelluksina. Semanttisen webin tulevia sovellusalueita ovat mm. tiedonhaku, tietämyksen hallinta, verkkokauppa ja sähköinen liiketoiminta. Semanttisella webillä on selvä yhteys tekoälyn perinteisiin tutkimusalueisiin erilaisesta teknisestä ympäristöstä ja kaupallisesti suuntautuneista tavoitteista huolimatta.

Tutkielman tarkoituksena on tuoda perspektiiviä ehkä ylioptimistiseen visioon semanttisesta webistä, jossa hakukoneet löytävät käyttäjän haluat tiedot vaivatta, sovellukset kommunikoivat keskenään automaattisesti ja autonomiset agentit hoitavat asioinnin käyttäjän puolesta. Samoja teemoja on tutkittu ja vuosikymmenten ajan klassisessa tekoälytutkimuksessa ja tietämystekniikassa. On kyseenalaista, pystyykö semanttinen web lunastamaan odotukset täydessä mitassa.

Tutkielman luvut jakautuvat seuraavasti: luvussa 2 käsitellään tietämyksen esittämistä yleisesti ja esitellään tärkeimpiä verkkoformalismeja. Luvussa 3 käydään läpi semanttiseen webiin liittyviä tekniikoita ja kieliä. Luvussa 4 tutkitaan semanttisten verkkojen ja semanttisen webin välistä suhdetta ja esitellään sovelluksia.

Oletan, että lukija tuntee matematiikan perusopintoihin sisältyvät predikaattilogiikan perusteet ja verkkoteorian alkeet. Semanttista webiä käsittelevän osuuden ymmärtäminen edellyttää WWW-tekniikoiden ja XML-kielen perustuntemusta. Jälkimmäisistä aiheista on runsaasti materiaalia saatavilla esim W3C:n WWW-sivuilla<sup>1</sup>.

---

<sup>1</sup><http://www.w3.org/>

## 2 Semanttiset verkot tietämyksen esittämisessä

Luvussa kuvataan tietämyksen esittämistä yleisesti ja sen asemaa tekoälyssä ja muilla sovellusalueilla. Myös tietämyksen rakenteen perustana olevia ontologioita käsitellään.

Keskeinen rakenne tietämyksen esittämiseen ovat semanttiset verkot, joista tärkeimpiä formalismeja käydään läpi luvussa 2.2 esimerkinomaisesti. Semanttisen verkon käsitteen epämääräisyydestä johtuen verkkoformalismit poikkeavat toisistaan niin notaatioltaan, semantiikaltaan kuin ilmaisukyvyltään. Kaikille on kuitenkin yhteistä verkkomainen rakenne, nimetyt solmut ja kaaret sekä käyttötarkoitus.

Luvun lopussa tutkitaan loogisten lausekkeiden ja semanttisten verkkojen välistä suhdetta. Varhaisten verkkoformalismien esittämisen aikaan kuvaustapojen kannattajat kiistelivät, mikä kuvaustapa soveltuisi parhaiten tietämyksen esittämiseen. Myöhemmin havaittiin, että oikein formalisoituna semanttiset verkot ja 1. asteen logiikka mahdollisine laajennuksineen ovat isomorfisia. Luettavuuden, päättelyn intuitiivisuuden ja osittain myös toteutuksen tehokkuuden kannalta verkkorakenteessa on kuitenkin etuja loogisiin lausekkeisiin nähden.

### 2.1 Tietämyksen esittämisestä

Tietämyksen esittäminen (*knowledge representation*) on alunperin liittynyt klassista tekoälyä hyödyntävien sovellusten tietämuskannan muodostamiseen. Tutkittavasta tai simuloitavasta ilmiöstä poimitaan tärkeimmät käsitteet ja ne mallinnetaan suhteineen osaksi tietämuskantaa. Tällöin tietämuskantaan voidaan tehdä kyselyjä ja päättelyä tunnettujen tietojen pohjalta.

Nykyisin tietämyksen esittämisen käsite on yleistynyt tekoälyjärjestelmistä osaksi kaikkea tiedon organisointia. Esimerkiksi tietokantasuunnittelu kaupallis-hallinnollisissa sovelluksissa sekä oliosuunnittelu oliopohjaisissa järjestelmissä voidaan laajasti katsottuna nähdä tietämyksen esittämisenä.

Kaupallista merkitystä tietämyksen esittäminen on saanut ontologioiden yleistymisen myötä. Ontologiaa voidaan pitää uudelleenkäytettävänä tietämyksen esittämisrakenteena tai luettelona tiettyyn sovellusalaan liittyvistä käsitteistä. Semanttisessa webissä ontologioilla tulee olemaan keskeinen asema.

#### 2.1.1 Tavoitteet ja läheiset tutkimusalueet

Tietämystekniikka on logiikan ja ontologian (ks. luku 2.1.3) soveltamista. Tietämystekniikan tavoite on muodostaa laskettavia malleja tietystä sovellusalueesta tiettyyn tarkoitukseen [Sowa, 2001, sivu 132]. Malli sijaitsee

yleensä tietämuskannassa. Tietämyksen esittämisen tavoitteita ja ominaisuuksia voidaan luonnehtia Davisin, Schroben ja Szolovitsin [Sowa, 2001, sivut 134-143] mukaan seuraavasti:

- Tietämuskannan symbolit toimivat **viitteellisenä mallina** ympäröivän todellisuuden tietystä alueesta. Viitteitä manipuloimalla tietokoneohjelma voi simuloida reaali maailmaa ja tehdä päätelmiä siitä. Päättelystä vastaa yleensä erillinen ohjelmistokomponentti, jota kutsutaan päätteykoneeksi. Simulointi voidaan suorittaa joko proseduraalisen tai logiikkaohjelmoinnin keinoin.
- Olioiden rakenteen lisäksi tietämuskannan täytyy kuvata niiden suhteet ja käyttäytyminen sekä helpottaa **päätelyä**. Tämä kuvaus voidaan määritellä aksiomaattisena **teoriana**, loogisina sääntöinä tai tietokoneohjelman osina.
- Tietämuskannan hyödyntäminen edellyttää sen koodausta **laskennallisesti käsiteltävään** muotoon. Ohjelmointikielten teoriassa on ollut merkittävä vaikutus tietämyksen esittämiskielten kehittämisessä.
- Hyvän tietämyksen esittämiskielen pitäisi auttaa automaattisen päätteilyn lisäksi myös tietämuskannan suunnittelijan ja sovellusalueiden asiantuntijoiden **viestintää**. Näin asiantuntijat voivat arvioida, onko tietämyksen esittämisen tuloksena mallinnettu teoria pätevä.
- Filosofian alueena ontologia on oppi olemassaolosta. Tietokannalle tai tietämuskannalle ontologia määrittelee sovellusalueen olemassaolevat tai sallitut asiat. Nämä asiat ovat tietämuskannan suunnittelijan sovellusalueeseen liittyviä **ontologisia sitoumuksia**. Ontologia voidaan kuvata logiikalla tai jollakin verkkoformalismilla, kuten kehyksillä. Semanttisessa webissä ontologia voidaan kuvata esim. OWL-kielellä (*Web Ontology Language*) (ks. luku 3.2.2).

Sowa ehdottaa käsittegraafeja yleiseksi kieleksi tietämyksen esittämiseen [Sowa, 2001, sivut 140-141, 424-438, 452-454]. Muita tietämyksen esittämiseen soveltuvia kieliä ovat mm. KIF (*Knowledge interface format*), kehysjärjestelmät ja terminologialogiikat sekä semanttisen webin kielet RDF (*Resource Description Framework*) ja OWL.

Tietämyksen esittämällä ja tietokantasuunnittelulla on yhtäläisyyksiä. Molemmilla aloilla pyritään löytämään sovellusalueen yleiset ominaisuudet abstrahoimalla, hyläten samalla epäolennaiset yksityiskohdat. Molemmilla aloilla käytetään (osittain samoja) käsitteitä, rajoitteita, operaatioita ja kieliä. Tietämyksen esittäminen on kuitenkin käsitteenä tietokantasuunnittelua laajempi. Tietämuskannoissa on tuki päätteleysäännöille, epätäydelliselle tiedolle, oletusarvoille sekä ajalliselle ja paikalliselle tie-

dolle. Tietokantojen päättely on yleensä rajoittunut suoriin kyselyihin ilman päättelyä. Jotkut tietokannan hallintajärjestelmät ovat ottaneet käyttöön myös perinteisesti tietämuskannoille luokiteltuja ominaisuuksia. Esimerkiksi herättimet (*triggers*) voidaan nähdä tietokannan päättelysääntöinä. [Elmasri & al., sivut 100-104]

Myös oliopohjainen analyysi ja suunnittelu voidaan nähdä tietämyksen esittämisen lähialueena. Oliomallinnusta käsitellään tarkemmin luvussa 2.2.2 kehysten yhteydessä.

### 2.1.2 Tietämuskanta osana asiantuntijajärjestelmää

Asiantuntijajärjestelmä (AJ) on tyypillinen tekoälyjärjestelmä, jossa hyödynnetään tietämyksen esittämistä. Asiantuntijajärjestelmä on tietämyspohjainen tietojärjestelmä, joka tietämyksensä ja päättelykykynsä avulla pystyy ratkomaan ongelmia, joiden katsotaan vaativan merkittävää ihmillistä asiantuntemusta [Hyvönen, 1988, sivut 32-35]. Sovellusalueita ovat esimerkiksi matematiikka, lääketiede ja CAD-suunnittelu [Rich, sivut 284-291].

Asiantuntijajärjestelmä koostuu tavallisesti seuraavista komponenteista [Hyvönen, 1988, sivut 133-153]:

- **Tietämuskanta** (*knowledge base*) sisältää sovellusalueeseen liittyvät faktat, uskomukset ja säännöt. Tietämuskanta ei ole koskaan täydellinen, sillä käytettävissä oleva tieto voi olla epätäydellistä, epävarmaa tai jopa virheellistä. Tietämuskantaa kehitetään edelleen järjestelmän käytön aikana.
- **Päättelykone** (*inference engine*) on ohjausmekanismi, jonka avulla tietämuskantaa käytetään. Päättelykoneisiin integroidaan yleensä useita eri ongelmanratkaisumekanismia. Näitä ovat esimerkiksi eteenpäätely, taaksepäätely, kehyspohjainen päättely tai [Rich, sivut 284-291] tilastolliset menetelmät. Päättelykone pyritään erottamaan järjestelmän tietämyksestä, jolloin samaa päättelykoneita voi periaatteessa soveltaa useampiin loogiselta rakenteeltaan samankaltaisiin ongelmiin. Lisäksi tietämuskannan muunneltavuus helpottuu.
- **Työmuisti** sisältää tiedon ongelmanratkaisun tilasta, lähtö- ja tavoitetiedoista, tehdyistä päätelmistä ja näiden välisistä riippuvuuksista.
- **Käyttöliittymä** tarjoaa järjestelmään näkymän AJ:n käyttäjille ja kehittäjille. Käyttöliittymä riippuu sovellusalueesta ja käyttäjäryhmästä. Se voi olla esimerkiksi luonnolliseen kieleen perustuva, jolloin tekniikkaa vähän tunteva loppukäyttäjä voi esittää helposti järjestelmälle kyselyjä. Luonnollisen kielen oikea tulkinta muodostanee tosin oman ongelmansa.

- **Selitysmekanismi** perustelee AJ:n toimintaa selväkielisesti tai vastaa sen tietämystä koskeviin kysymyksiin. Kun AJ pystyy perustelemaan, miksi se on päässyt tiettyyn lopputulokseen, voidaan paremmin arvioida sen pätevyyttä. Hyvä selitysmekanismi voi myös kouluttaa käyttäjänsä.

Tietämyskanta ja päättelykonetta käytetään asiantuntijajärjestelmien lisäksi useimmissa symbolisiin rakenteisiin perustuvissa tekoälyjärjestelmissä. Esimerkiksi luonnollisen kielen käsittelyssä [Sowa, 2001, sivut 178-179] tarvittava taustatieto tai käsitteellisessä koneoppimisessä [Rich, sivut 368-374] hankittu tieto voidaan sisällyttää tietämyskantaan.

### 2.1.3 Ontologiat

Ontologia on filosofian osa-alue, joka tutkii olemassaoloa. Sowa kuvaa logiikan ja ontologian eroa seuraavasti [Sowa, 2001, sivut 51-55]:

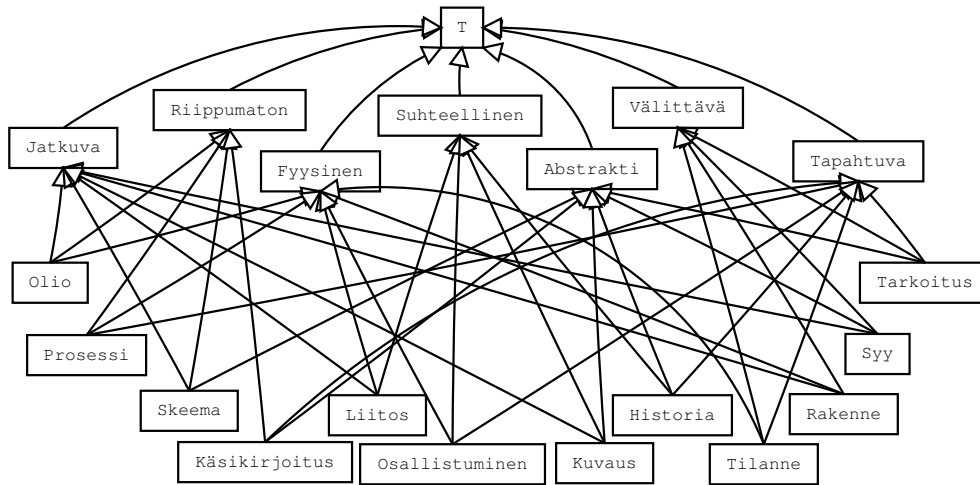
Logiikan olemassaolokvanttorilla voidaan esittää väite, että jokin on olemassa, mutta pelkällä logiikalla ei voida kuvata olemassaolevia asioita. Ontologia tuottaa predikaatit predikaattilogiikkaan ja merkityksen käsitegraafien solmuihin ja kaariin. Ontologiset kategoriat määrittävät sovellusalueen käsitteet ja asiat, joita voidaan käsitellä sovelluksessa ja joista voidaan tehdä päätelmiä.

Fensel [Fensel, sivu 8] esittää ontologian määritelmänä T. G. Gruberin tunnetun luonnehdinnan, jonka mukaan ontologia on *formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä (shared conceptualization)*. Määritelmä korostaa ontologioiden jaettavuutta erityisesti koneellista käsittelyä ajatellen.

Filosofiassa ontologioita on tutkittu Platonin ajoista lähtien. Filosofian ontologiset kategoriat ovat tyypillisesti olleet ”ylätason kategorioita”, joilla on pyritty mallintamaan kaikkeen olemassaoloon pätevät yhteiset piirteet. Näin ontologiasta saadaan mahdollisimman kattava, mutta käytännön sovelluksia varten usein liian abstrakti teoria. Esimerkkinä ylätasontontologiasta on Sowan johtama kategoriahila. Hilan huipulla on kaikelle pätevä universaalityyppi T, josta muut kategoriat on peritty. Kuvassa 1 on esitetty Sowan kategoriat. Tarkempi kuvaus kategorioista on kirjassa ”Knowledge representation”. [Sowa, 2001, sivut 67-77]

Viime vuosikymmeninä ontologia on havaittu myös filosofian ulkopuolella. Kivelä ja Hyvönen käsittelevät artikkelissaan [Kivelä & al.] eri näkökulmia ontologioihin. Ontologioita on sovellettu tekoälyssä, tietämystekniikassa, kielitieteessä ja tietojärjestelmien kehityksessä. Terminologia, menetelmät, formaalisuus ja ontologian kohteena olevan sovellusalueen laajuus vaihtelevat.





Kuva 1: Sowan ylätasen kategoriat.

Filosofiassa ontologisia kategorioita mallinnetaan yleensä yläkäsitteistä alakäsitteisiin, tietojärjestelmien kehittämisessä tai olio-ohjelmoinnissa käsitteitä (tai luokkia) mallinnetaan usein alhaalta ylös lähtien sovellusalueen konkreettisista käsitteistä. Kielitieteen sanastot eli *thesaurukset* (ks. luku 4.2.2) voivat olla hyvin epäformaaleja, mutta monien tekoälyjärjestelmien (ks. luku 4.2.1) ontologiat määritellään formaalisti, logiikan avulla. Pohjimmiltaan ontologioita käytetään näillä kaikilla aloilla todellisen maailman tai sen osa-alueen käsitteiden mallintamiseen.

Parintuhannen vuoden aikana filosofit eivät ole päässeet yksimielisyyteen kaikenkattavista ontologioista — tuskin koskaan pääseväkään. Sama pätee myös useimpiin sovellusaluekohtaisiin ontologioihin. Standardointijärjestöt yrittävät saada aikaan yleisesti hyväksytyjä määrittämiä, mutta rinnakkaisia esityksiä samoille asioille tulee varmasti olemaan jatkossakin. Ontologioiden jaettu käyttö ja hallinta on ollut ja tulee olemaan merkittävä haaste tietämyksen esittämisessä [Sowa, 2001, sivut 408-414] ja erityisesti semanttisen webin toteutuksessa [Hyvönen, 2002].

## 2.2 Verkkoformalismeja

Semanttisissa verkoissa ei itsessään ole mitään semanttista. Itse asiassa yhtä "ainoa ja oikeaa" semanttista verkkoa ei ole edes olemassa. Nimitys koskee perhettä löyhästi toisiaan muistuttavia formalismeja, joissa kuvataan tietoa verkkorakenteessa. Alunperin nimitys lienee lainattu kielitieteestä, onhan yksi tekoälyn sovellusalue luonnollisen kielen tulkinta.

Alkuperäisissä semanttisissa verkoissa ei kuitenkaan ollut semantiikkaa matematiikan ja logiikan täsmällisyyden edellyttämällä tavalla. Merkitykset näyttäytyivät käyttäjälle nimettyinä solmuina ja kaarina. Ne selventävät käsitteitä ihmiselle, mutta eivät sovellu koneellisen päättelyn pe-

rustaksi.

Tietämyksen esittämisen ”loogisen” koulukunnan kannattajat kritisoi-  
vat alkuperäisiä epäformaaleja verkkoja, ja myöhemmissä verkkoforma-  
lismeissa suhteiden merkitys määriteltiin tarkemmin. Nykyiset semantti-  
set verkot ovat formaali kieli siinä missä 1. kertaluvun predikaattilogiikka-  
kin. Verkkojen kuvauskyky vaihtelee. Esimerkiksi kehykset (ks. luku 2.2.2)  
pystyvät esittämään logiikasta vain  $\exists$  ja  $\wedge$  -operaattoreilla varustetun os-  
ajoukon. Toisaalta käsitegraafeilla (ks. luku 2.3.2) pystytään ilmaisemaan  
väittämiä, joihin predikaattilogiikka ei sellaisenaan pysty.

### 2.2.1 Historiaa: puoliformaalit verkot

Kaikille semanttisille verkkoformalismeille yhteistä on tietämyksen esittä-  
minen solmuina ja kaarina, jotka ovat yleensä nimettyjä. Niihin voi myös  
liittyä ominaisuuksina tietoa ja tietämystä. Yleensä solmut kuvaavat kä-  
siteltävän sovellusalueen käsitteitä tai olioita, kaaret näiden käsitteiden  
välisiä suhteita. [Hyvönen, 1988, sivut 118-122] Usein käytettyjä suhde-  
tyyppejä ovat nykyään olio-ohjelmoinnista tutut `subclass-of` (yläkäsité-  
alakäsite), `instance-of` (käsitteen ilmentymä) ja `part-of` (kooste).

Porpyruksen puu lienee filosofian varhaisin semanttista verkkoa muis-  
tuttava rakenne. Porphyryus kuvasi 300-luvulla j.a.a. Aristoteleen ontolo-  
giset kategoriat puuksi, jonka mukaan mm. *ihminen on rationaalinen eläin*  
[Sowa, 2001, sivut 1-11, 274-284]. 1800-luvun lopulla Peirce kehitti graafi-  
sen notaation loogisten lausekkeiden kuvaamiseen ja päättelyyn, ns. ek-  
sistentiaaliset graafit. Eksistentiaalisia graafeja käsitellään tarkemmin lu-  
vussa 2.3.2.

Tekoälytutkimuksessa semanttiset verkot keksittiin 1960-luvulla, mm.  
Quillianin tutkimusten tuloksena. Verkkoja käytettiin ihmismuistin psy-  
kologisina malleina. Perusajatuksena oli, että toisiinsa liittyvien käsittei-  
den tuli sijaita toistensa ”semanttisessa läheisyydessä”. Näin käsitteisiin  
liittyvä tieto voitaisiin löytää laskennallisesti helposti [Hyvönen, 1988, si-  
vut 118-122]. Epäformaalein muoto tästä ovat oppimisen apuvälineinä  
käytettävät miellekartat eli mind mapit.

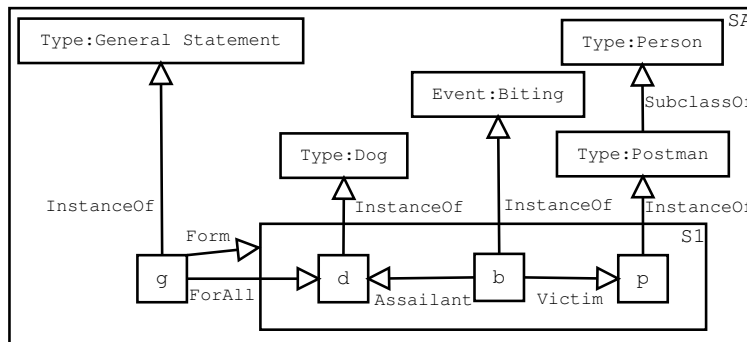
Tällaiset puoliformaalit verkot saivat runsaasti kritiikkiä täsmällisen  
semantiikan puutteen vuoksi. Eräs kuuluisa semanttisia verkkoja kriti-  
soiva artikkeli on W.A. Woodsin *What's in a link: Foundations for semantic  
networks* vuodelta 1975 [Woods, 1991]. Puutteiksi hän luetteli mm. seuraa-  
vat:

- Käsitteiden yleistyksessä käytettyä suhdetta ei ole määritelty sel-  
keästi, jolloin olioiden luokittelu käsitteisiin ei toimi oikein.
- Käsitteiden määrittelyssä ei ole erotettu käsitteen intensiota (määri-  
tys) ja ekstensiota (ilmentymien joukko). Esim. käsitteiden ”aamu-  
tähti” ja ”iltatähti” ekstensio on planeta Venus.

- Määrittelyksiä kuvaavia (rakenteellisia) linkkejä ei ole erotettu väittämiä kuvaavista linkeistä. Määrittelyjen ja väitteiden erottaminen vaikuttaa verkossa tehtävään päättelyyn.
- Verkoissa on yleisesti käytetty linkkejä, joiden semantiikkaa ei ole määritelty formaalisti.

Edellisten lisäksi puoliformaaleissa verkoissa ei yleensä pystytä esittämään loogista negaatiota,  $\forall$ -operaatiota eikä kvantifiointia ( $\forall$ ). Epäkohdista ja puutteista huolimatta semanttisten verkkojen käyttöä puoltaa niiden ymmärtämisen helppous ja tapa kuvata käsitteiden yleistyshierarkia. Semanttisiin verkkoihin perustuvat tietämysjärjestelmät voivat olla myös nopeampia kuin logiikkaan perustuvat. [Russell & al., sivut 316-325].

Semanttisten verkkojen ilmaisykykyä voidaan laajentaa. Ositetut semanttiset verkot ovat Hendrixin 1977 esittämä formalismi, joka vastaa ilmaisukyvyltään 1. asteen predikaattilogiikkaa. Hendrixin ratkaisu on verkon osittaminen avaruuksiin (kontekstit), jotka määräävät kvantifioinnin tai negaation vaikutusalueen. Avaruudet muodostavat hierarkian [Rich, sivut 215-222] ja voivat lisäksi limittyä, jolloin samoihin käsitteisiin voidaan viitata useammasta kontekstista [Sowa, 2001, sivut 274-284].



Kuva 2: "Every dog has bitten a postman" — ositettu semanttinen verkko.

Kuvassa 2 on esitetty ositettu semanttinen verkko lauseelle *Every dog has bitten a postman*. Esimerkki pohjautuu Richin kirjan vastaavaan esitykseen.

## 2.2.2 Kehykset ja terminologilogiikat

Epäformaalien verkkojen epäkohtien tultua esille tekoälytutkimuksessa etsittiin uusia tapoja tietämyksen esittämiseen. Marvin Minsky esitteli kehykset vuonna 1975 artikkelissa *A framework for representing knowledge*. Kehykset ovat rakenteinen, formaali ja nykynäkökulmasta olioita muistuttava tapa tietämyksen esittämiseen.

Kehys on tiettyä tilannetta, oliota tai käsitettä kuvaava tietorakenne, jonka tiedot on sijoitettu kenttiin. Kehyksen tyyppi määrää, millaisia kenttiä kehyksessä voi olla. Kentillä voi olla oletusarvoja tai monipuolisempia rajoitteita järjestelmästä riippuen. Joissakin järjestelmissä sallitaan myös määritelyihin tapahtumiin liittyvän ulkoisen ohjelmakoodin liittäminen järjestelmään. Tapa muistuttaa tietokantojen herättimiä tai tapahtumapohjaisessa ohjelmoinnissa käytettäviä tapahtumankäsittelijöitä. Kehystyyppit muodostavat moniperinnän sallivan hierarkian.

Kehykset voidaan tulkita myös semanttisina verkkoina, joissa linkkien tyypit on määritelty ja niiden osoittamien solmujen ominaisuuksia on rajoitettu. Kehysjärjestelmät toteuttavat yleensä 1. kertaluvun predikaattilogiikan ilman negatiota ja  $\vee$ -operaattoria [Sowa, 2001, sivut 143-156]. Kehyspohjaisista järjestelmistä tuli suosittuja monissa tekoälyprojekteissa Minskyn artikkelin julkaisun jälkeen. Eräs syy kehysjärjestelmien suosioon lienee niiden helppo koodaus LISP-kielen tietorakenteiksi. Nykyisistä järjestelmistä esim. Cyc (katso luku 4.2.1) suunniteltiin alunperin kehyspohjaiseksi, joskin järjestelmä toimii nykyään predikaattilogiikan laajennukseen perustuvalla CycL-kielillä.

Eräs varhainen formaali toteutus semanttisille verkoille on R. Brachmanin 1979 suunnittelema KL-ONE -kieli (*Knowledge Language One*). Aiemmista verkkoformalismeista poiketen KL-ONE:ssa on tarkasti määritelty semantiikka. KL-ONE -kielen pohjalta on tehty useita järjestelmiä, joita Woods tosin kritisoi. Kaikissa toteutuksissa ei ole saavutettu kielen täyttä ilmaisukykyä — juuri väärinymmärretyt semantiikan takia [Woods, 1991].

KL-ONE -kielen pohjalta on kehitetty useita samansukuisia laajennuksia, joista käsitellään esimerkkinä CLASSIC-järjestelmää [Brachman & al.]. Järjestelmän tietämyskantaan voidaan määritellä käsitteitä (kehykset, luokat), rooleja (kentät, attribuutit) sekä yksilöitä (oliot), joiden ominaisuuksien perusteella päätellään niiden edustamat käsitteet. Edelleen voidaan määritellä yksinkertaisia sääntöjä yksilöiden edustamien käsitteiden johtamiseen jo havaittujen käsitteiden pohjalta.

Verrattaessa kehysjärjestelmiä olio-ohjelmointiin yhtäläisyytenä on perinnän käyttäminen käsitteiden mallinnuksessa ja tietorakenteiden esitys. Erona olio-ohjelmointiin kehysjärjestelmissä ei yleensä määritellä metodeja tai muuta toiminnallisuutta. Toisaalta olio-ohjelmoinnista puuttuvat säännöt ja rajoitteet attribuuttien arvoille. Merkittävämpi ero on, että oliokiellisessä olion luokka on aina tiedettävä sen luomisen yhteydessä. KL-ONE -tyyppiset kielet lähtevät yksilöiden luokittelusta (*classification*): yksilö voi edustaa useampaa käsitettä, jotka voivat vaihdella yksilön tilan mukaan. Kaikkien kenttien arvoja ei tarvitse tietää yksilön luontihetkellä, joten kehysjärjestelmät tukevat myös epätäydellisen tiedon esittämistä.

Terminologilogiikat (*description logics, terminological logics*) ovat formaali malli kehyksille ja semanttisille verkoille. Esimerkiksi CLASSIC-jär-

jestelmää tai Woodsin esittämää luokittelukieltä [Woods, 1991] voidaan pitää terminologilogiikkoina. Terminologilogiikat eroavat tavanomaisesta predikaattilogiikasta keskittymisessä käsitteiden kuvaamiseen väittämien sijasta. Käsitteillä operoiminen mahdollistaa selkeän notaation ja tehokkaan päättelyn predikaattilogiikkaan verrattuna — tosin kielen ilmaisuvoiman kustannuksella [Russell & al., sivut 298, 316-325]. Yksinkertaisena esimerkkinä olkoon CLASSIC-kielinen määrittely

$$Bachelor = And(Unmarried, Adult, Male),$$

joka vastaa predikaattilogiikan lausetta

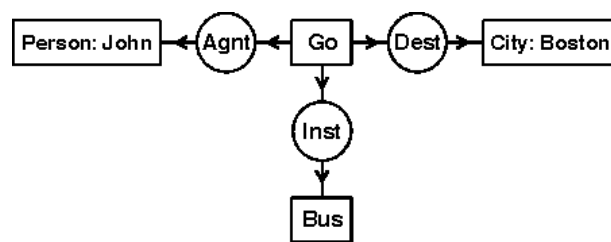
$$\forall x : (bachelor(x) \Leftrightarrow unmarried(x) \wedge adult(x) \wedge male(x)).$$

### 2.2.3 Käsitegraafit ja KIF

Luku perustuu pääosin Sowan kirjaan "Knowledge Representation". Kirjassa kuvataan laajasti käsitegraafien suhdetta muihin tietämyksen esittämisessä käytettyihin menetelmiin ja kieliin [Sowa, 2001]. Tiiviimpi kuvaus käsitegraafeista on artikkelissa *Toward the Expressive power of Natural Language* [Sowa, 1991].

John Sowa esitteli käsitegraafit (*Conceptual Graphs*) kirjassa "Conceptual Structures — Information processing in Mind and Machine" 1984. Puoliformaalien verkkojen tapaan käsitegraafeilla on nimetyistä solmuista ja kaarista koostuva notaatio. Graafin komponentteja voidaan koota ositetujen verkkojen tapaan hierarkkisiksi (mutta ei limittäisiksi) konteksteiksi.

Kuvassa 3 on esimerkki käsitegraafista. Kuva on otettu Sowan käsitegraafeja esittelevältä WWW-sivulta <sup>2</sup>.



Kuva 3: "John is going to Boston by bus" — käsitegraafiesitys.

Graafisesta notaatiosta huolimatta järjestelmän lähtökohta on 1. kertaluvun predikaattilogiikka ja erityisesti loogikko Peircen jo vuonna 1897 keksimä notaatio sen esittämiseen: eksistentiaaliset graafit. Tästä johtuen käsitegraafeilla on täsmällisesti määritelty semantiikka ja niillä voidaan

<sup>2</sup>©1999 John F. Sowa. Kuvaa saa käyttää vapaasti, kunhan tekijä ja alkuperäisen dokumentin URL mainitaan. <http://www.jfsowa.com/cg/cgexampw.htm>

kuvata kaikki predikaattilogiikan mukaiset lauseet. Luvussa 2.3.2 kuvataan tarkemmin käsitegraafien ja logiikan suhdetta.

Käsitegraafit suunniteltiin alun perin yksinkertaistamaan muunnosta luonnollisten kielten ja logiikan välillä. Ne soveltuvat kuitenkin periaatteessa kaikkeen tietämyksen esittämiseen johtuen niiden läheisestä yhteydestä logiikkaan ja toisaalta semanttisiin verkkoihin. Käsitegraafeilla voidaan määritellä kehysjärjestelmistä tutut tyyppihierarkiat ja kieltä voidaan laajentaa sen itsensä käsitteitä käyttäen. Jokaiselle käsitegraafille on olemassa esitys myös predikaattilogiikan kaavana. Käsitegraafit soveltuvat myös muunnosten määrittelyyn formalismista toiseen.

Käsitegraafidatan syötön helpottamiseksi vaihtoehtona graafiselle notaatiolle on myös tekstuaalinen (lineaarinen) notaatio. Kuvan 3 graafi voidaan esittää lineaarisella notaatiolla seuraavasti:

$$[Go] - \{ \\ \rightarrow (Agnt) \rightarrow [Person : John] \\ \rightarrow (Dest) \rightarrow [City : Boston] \\ \rightarrow (Inst) \rightarrow [Bus] \}.$$

Tyypitetyn predikaattilogiikan merkinnöillä vastaava lause näyttää seuraavalta:

$$(\exists x : Go)(\exists y : Person)(\exists z : City)(\exists w : Bus) \\ (name(y, John) \wedge name(z, Boston) \wedge agnt(x, y) \wedge dest(x, z) \wedge inst(x, w))$$

Lukijalla ei liene epäselvyyttä, kumpi muoto on useimmille ihmislukijoille helpompi hahmottaa.

KIF (*Knowledge Interchange Format*) on käsitegraafien tapaan 1. kertaluvun predikaattilogiikan toteuttava ja tietämyksen esittämiseen soveltuva formalismi [Sowa, 2001, sivut 134-143]. KIF on M. Geneserethin ja R. Fikesin suunnittelema versio tyypitetystä predikaattilogiikasta. Sen päätarkoitus on toimia välikielenä eri tieto- ja tietämuskantojen välillä. Maksimoidakseen toimivuuden eri järjestelmien ja toteutusten välillä KIF käyttää rajoitettua merkkijoukkoa. Esimerkiksi logiikasta tutut symbolit  $\wedge$  ja  $\forall$  on korvattu tekstuaalisilla kuvauksilla. Taulukossa 1 on vertailtu kvantifioinnin esittämistä eri kielillä.

Sekä käsitegraafit<sup>3</sup> että KIF<sup>4</sup> ovat ANSI/ISO:n standardoinnissa tietämyksen esittämiseen [Sowa, 2001, sivu 424]. Määrittämiä on kehitetty rinnakkain niiden keskinäisen muunnettavuuden varmistamiseksi. KIF soveltuu käsitegraafeja paremmin koneelliseen käsittelyyn ja sitä pidetään

<sup>3</sup>ISO/JTC1/SC 32/WG 2

<sup>4</sup>NCITS.T2/98-004

Englanti	Tyypitetty logiikka	Käsitegraafit	KIF
some student	$(\exists x : Student)$	$[Student]$	<code>(exists(?x student))</code>
every student	$(\forall x : Student)$	$[Student : \forall]$	<code>(forall(?x student))</code>

Taulukko 1: Esimerkki kvantifioinnista eri kielillä.

yhtenä mahdollisena standardina semanttisen webin logiikkakerroksen kuvaamiseen [Hyvönen, 2002]. Käsitegraafit soveltuvat esitysformaatiksi niiden paremman luettavuuden takia.

## 2.3 Semanttisten verkkojen yhteys logiikkaan

1960-luvulla semanttisten verkkojen tultua tekoälytutkimuksen piiriin logiikan kannattajat vastustivat verkkojen käyttöä tietämyksen esittämisessä. Tietämyksen esittämisessä olikin pitkään vallalla kaksi koulukuntaa, verkkorakenteiden ja logiikalla ilmaistavien sääntöjen kannattajat. Tässä valossa on ehkä yllättävää, että nykyisten logiikassa käytettyjen symbolien keksijä Peirce oli jo 1800-luvulla ottamassa käyttöön verkkoformalismia logiikan havainnoistamiseen, ns. eksistentiaaliset graafit. Peircen graafit voidaan tulkita myös semanttisina verkkoina, joissa on täsmällinen semantiikka.

Nykytutkimuksessa verkkojen ja logiikan kannattajien kiista on menettänyt merkityksensä, kun on ymmärretty, että pohjimmiltaan semanttiset verkot ja logiikka ovat eri esityksiä formaalista järjestelmästä. Vastakkainasettelun sijasta ne täydentävät toisiaan. Luotettavan päättelyn edellytys on hyvin määritelty logiikka päättelysääntöineen. Luettavuuden, päättelyn intuitiivisuuden ja osittain myös toteutuksen tehokkuuden kannalta verkkorakenteessa on etuja loogisiin lausekkeisiin nähden.

### 2.3.1 Logiikan laajennuksia

Tietämyspohjaista järjestelmää suunniteltaessa eräs keskeinen kysymys on päättelykoneessa käytettävän kielen valinta. Periaatteessa sovellusalueen tiedot voidaan kuvata 1. kertaluvun predikaattilogiikalla, mutta kaa-voista saattaa tulla vaikeaselkoisia ja päättely voi olla laskennallisesti vaativaa. Esimerkiksi kehysjärjestelmissä ilmaisukykyä on tietoisesti pudotettu poistamalla kielestä yleinen negaatio ja  $\forall$ -operaattori, mutta vastaavasti suorituskyky ja kuvauksen havainnollisuus on parantunut.

Predikaattilogiikan lisäksi logiikassa on määritelty lukuisia muita formaaleja systeemejä, jotka ilmaisukyvyltään ylittävät 1. kertaluvun logiikan (tosin suuri osa näistä laajennuksista voidaan määritellä 1. kertaluvun logiikan käsittein). Semanttiset verkot ovat myös formaaleja systeemejä –

ainakin, jos solmujen ja kaarien semantiikka on määritelty riittävän tarkkaan.

Eräs yksinkertainen laajennus predikaattilogiikalle on tyyppin käsitteen lisääminen. Semanttisten verkkojen termein tämä vastaa *instance-of*-operaattorin käyttöönottoa. Tyypitetyt kvanttorit voidaan määritellä predikaattilogiikalla seuraavasti [Sowa, 2001, sivut 18-29]:

$$\begin{aligned}(\forall x : T)P(x) &\equiv \forall x : (T(x) \Rightarrow P(x)) \\(\exists x : T)P(x) &\equiv \exists x : (T(x) \wedge P(x))\end{aligned}$$

Ensimmäinen kaava on siis määrittely väitteelle *Kaikille x tyyppiä T pätee P* ja jälkimmäinen väitteelle *On olemassa x tyyppiä T, jolle pätee P*. Tyypin käyttöönotto selkeyttää huomattavasti loogisia lauseita.

Muita esimerkkejä logiikan laajennuksista ovat korkeamman kertaluvun logiikat (*Higher Order Logics, HOL*), intensionaaliset eli modaaliset logiikat ja epämonotoninen logiikka. Korkeamman kertaluvun logiikka eroaa tavallisesta predikaattilogiikasta sallimalla kvantifioinnin predikaattien yli (1. kertaluvun logiikassa kvantifiointi sallitaan vain muuttujien yli) [Hyvönen, 1988, sivut 59-66]. Sowa antaa esimerkkinä HOL-kaavasta aritmetiikasta tutun induktion määrittelyn:

$$(\forall P : \text{Predikaatti})((P(0) \wedge (\forall n : \mathbb{N})(P(n) \Rightarrow P(n+1))) \Rightarrow (\forall n : \mathbb{N})p(n))$$

Intensionaaliset logiikat lisäävät logiikkaan kaksi uutta operaattoria, joilla voidaan kuvata *mahdollisuuden* ja *välttämättömyyden* käsitteitä. Käsitepareja voidaan lisätä edelleen tai niille voidaan esittää erilaisia tulkintoja, kuten *tietää* ja *uskoa* tai *aina* ja *joskus*. Motivaatio intensionaaliselle logiikalle on luonnollisen kielen käsitteiden tuonti loogisiin lauseisiin. Modaaliset laajennukset ovat käytössä esim. käsitegraafeissa.

Esimerkkinä intensionaalisen logiikan teoreemasta esitetään määrittely *Lause on välttämättä totta jos ja vain jos se ei ole mahdollisesti epätotta*. Kaavassa  $\square$  siis merkitsee välttämättömyyttä ja  $\diamond$  mahdollisuutta.

$$\square P \equiv \neg \diamond \neg P$$

Epämonotoninen logiikka mallintaa epävarman ja virheellisen tiedon kuvaamista. Tavanomaisen logiikan päättely on monotonista siinä mielessä, että uudet faktat ja päättelysäännöt voivat vain lisätä teoreemojen joukkoa. Epämonotonisessa päättelyssä teoreemojen määrä voi tilapäisesti vähentyä lisätiedon paljastaessa osan päättelyssä käytetyistä oletuksista virheellisiksi.

Formaalien systeemien ominaisuuksia arvioidaan usein validisuuden ja täydellisuuden kriteereillä. Validisuus tarkoittaa, että jokainen kielen aksioomista todistettava väite (teoreema) on semanttisesti tosi (validi). Täydellisyys tarkoittaa, että jokainen validi kaava on teoreema. Täydellisyys

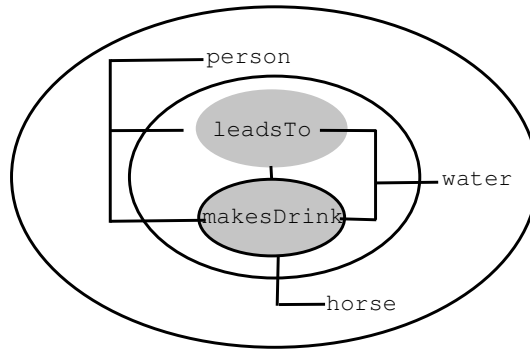


on hyvin vahva vaatimus, jonka toteutuminen on mahdollista vain rakenteeltaan yksinkertaisilla formaaleilla systeemeillä. Propositio- ja 1. kertaluvun predikaattilogiikka ovat täydellisiä, mutta esimerkiksi korkeamman kertaluvun logiikat eivät ole. Määritelmänsä takia epämonotoninen logiikka ei ole edes validia. Se toteuttaa vain heikomman konsistenttiuden ehdon: kaikkien tosien lauseiden täytyy olla konsistentteja (ristiriidattomia), mutta kaikki ristiriidattomat lauseet eivät ole tosia [Sowa, 2001, sivut 297-307].

### 2.3.2 Eksistentiaalisista graafeista käsittegraafeihin

Eksistentiaalisten graafien notaatio perustuu predikaattilogiikan operaatioiden graafiseen esittämiseen. Verkon solmut esittävät  $n$ -paikkaisia predikaatteja (vastaavat semanttisten verkkojen käsitteitä), vaakasuuntaiset linkit esittävät  $\exists$ -operaattoria ja pystysuorat linkit konjunktioita  $\wedge$ . Negaatio  $\neg$  esitetään ovaalisulkeumalla (konteksti), joita voi olla useampia sisäkkäin. Muistettakoon, että epäformaaleissa verkkoformalismeissa negaatiota ei ole yleensä pystytty esittämään.

Eksistentiaalisilla graafeilla pystytään esittämään 1. kertaluvun predikaattilogiikan lisäksi myös sen laajennuksia. Kuvassa 4 on esitetty eksistentiaalisena graafina modaalisen logiikan lause *You can lead a horse to water, but you can't make him drink*. Harmaa kontekstiellipsi ilman reunaviivaa esittää mahdollisuutta, reunaviivan kanssa mahdollisuuden negaatiota (eli välttämätöntä epätotuutta). [Sowa, 2001, sivut 274-278]



Kuva 4: "You can lead a horse to water, but you can't make him drink" — eksistentiaalinen graafi.

Modaalisen predikaattilogiikan symbolein vastaava väite voidaan esittää seuraavasti:

$$\forall x : \forall y : \forall z : ((person(x) \wedge horse(y) \wedge water(z)) \Rightarrow (\diamond leadsTo(x, y, z) \wedge \neg \diamond makesDrink(x, y, z)))$$

Predikaattilogiikassa päättely hoidetaan usein *modus ponensin* kaltaisilla päättelysäännöillä joko suoraan todistukseen tai ristiriitaan edeten. Peirce kehitti eksistentiaalisia graafeja varten omat päättelysäännöt, jotka ovat havainnollisesti sovellettavissa graafeihin. Säännöt soveltuvat luonnollisesti minkä tahansa formaalin systeemin päättelysäännöiksi. Peircen päättelysäännöt ovat seuraavat: [Sowa, 2001, sivut 299-300]

1. **Poisto.** Minkä tahansa (ali)graafin voi poistaa positiivisessa kontekstissa.
2. **Lisäys.** Minkä tahansa (ali)graafin voi lisätä negatiivisessa kontekstissa.
3. **Iteraatio.** Jos (ali)graafi  $p$  on jossakin kontekstissa  $C$ , toinen  $p$ :n kopia voidaan sijoittaa samaan kontekstiin tai mihin tahansa  $C$ -sisällä olevaan kontekstiin.
4. **Deiteraatio.** Mikä tahansa graafi, jonka voisi johtaa iteraation avulla voidaan poistaa: (ali)graafi  $p$  voidaan poistaa, jos siitä on kopia samassa tai jossain kontekstissa, johon  $p$ :n konteksti sisältyy.
5. **Kaksoisnegaatio.** Kaksi negaatiota, joiden välissä ei ole mitään voidaan poistaa tai lisätä graafiin mihin tahansa kontekstiin. Erityisesti kaksoisnegaatio voidaan lisätä tyhjän kokoelman ympärille.

Intuitiivisesti päättelysäännöt perustuvat periaatteelle, että yleistys positiivisessa kontekstissa tai erikoistaminen negatiivisessa kontekstissa säilyttää totuuden. Väitetään esimerkiksi, että *Koira syö luuta keittiössä*. Tällöin myös seuraavat väitteet pätevät:

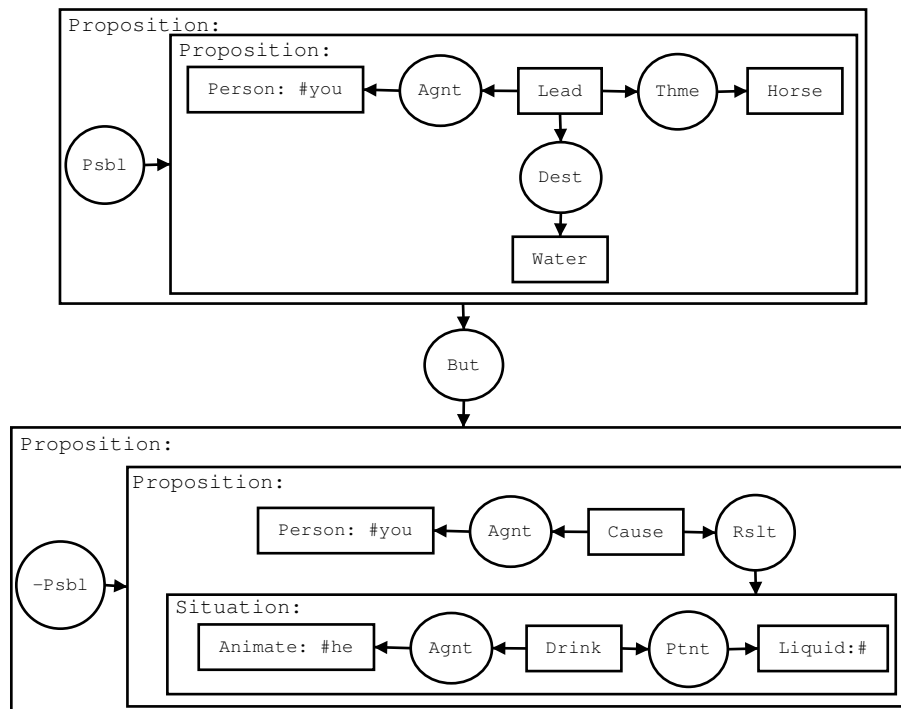
- *Koira on keittiössä.*
- *Eläin syö.*
- *Eläin syö jotain huoneessa.*

Käsitegraafeissa konteksteja voidaan käyttää eksistentiaalisten graafien tapaan. Kuvassa 5 on eräs mahdollinen käsitegraafiesitys edellä kuvatulle väitteelle. Graafissa on otettu huomioon kuvaus luonnolliselle kielelle. Oletuksena on, että kaikki graafin solmu- ja linkkityypit on määritelty formaalisti. Esimerkiksi kielellinen suhde *Agnt* (agentti) voidaan määritellä käsitegraafien lineaarisella notaatiolla seuraavasti:

$$Agnt = (\lambda x1, x2) [Act : *x1] \rightarrow (Has) \rightarrow [Agent] - - - [Animate : *x2]$$

*Agnt*-relaatio liittää teon  $x_1$  liikkuvaan olioon  $x_2$  [Sowa, 2001, sivu 272]. Määrittäminen on 2-paikkainen  $\lambda$ -lauseke [Sowa, 2001, sivut 22-25].  $\lambda$ -lausekkeet perustuvat Churchin 1941 kehittämään lambda-laskentaan.  $\lambda$ -merkinnällä voidaan määrittellä lausekkeen formaalit parametrit, joita yllä olevassa kaavassa ovat  $x_1$  ja  $x_2$ . — — — -symboli tarkoittaa viitelinkkiä, joka samastaa olion  $x_2$  käsitteeseen *Agent*.

Muistettakoon, että graafissa käytettävät kielelliset relaatiot ja käsitteet liittyvät sovellusalueen ontologiaan, eivät käytettävään logiikkaan. Käsitteet määritellään logiikan avulla.



Kuva 5: "You can lead a horse to water, but you can't make him drink" — käsitegraafi.

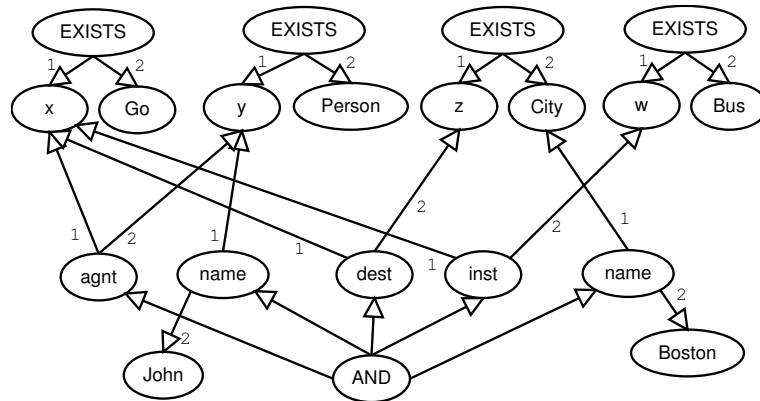
Käsitegraafissa olevat symbolit *#you*, *#he* ja *#* ovat indeksejä, edelleen Peircen määrittelemiä käsitteitä. Indeksit ovat kontekstiriippuvaisia viitteitä, joille ei ole vastaavaa esitystä logiikan puolella (indekseillä voidaan kuvata myös käsitteet *siellä*, *täällä*, *nyt*). *#* esittää yleisindeksiä, joka voidaan indoeurooppalaisissa kielissä tulkita määräiseksi artikkeliksi.

Indeksit helpottavat luonnollisen kielen analysointia ja tulkitsemista graafiksi, mutta ennen kuin graafin voi muuntaa logiikan lauseeksi, indeksit täytyy selvittää jossakin kontekstissa. Yleinen muunnos indekseistä logiikkaan voi olla hankalaa. Muunnoksessa saatetaan tarvita taustatietoa, jota lauseesta itsessään ei suoraan näe. Useissa järjestelmissä muunnos perustuu Hans Kampin 1981 esittämään diskurssien esitysteoriaan (*discourse representation theory*). Indeksit muunnetaan muuttujiksi tai vakioiksi.

Koska käsitegraafeilla voidaan esittää indeksejä, voidaan sanoa, että käsitegraafit ovat predikaattilogiikkaa ilmaisuvoimaisempia. Toisaalta indeksit ovat lähinnä välivaiheen esitystapa muunnettaessa luonnollista kieltä logiikan kielelle, joten niitä voidaan pitää myös syntaktisina apuvälineitä. Varsinainen päättely suoritetaan yleensä vasta indeksien selvittämisen jälkeen. [Sowa, 2001, sivut 278-284]

### 2.3.3 Verkkorakenne vai algebrallinen notaatio?

Tässä vaiheessa lukijalle lienee selvää, että täsmällisillä määrittelyillä mikä tahansa semanttinen verkko voidaan muuntaa predikaattilogiikan (tai jonkin sen laajennuksen) kaavoiksi. Eräs muunnostapa logiikasta verkoksi saadaan määrittelemällä solmujen tyypeiksi predikaattilogiikan vakiot, muuttujat, funktiot ja operaattorit. Jokainen predikaattilogiikan kaava saadaan helposti muunnettua suunnatuksi syklittömäksi graafiksi [Schubert]. Schubert kutsuu esimerkkinotaatiota nimellä *s-net*. Edellisen luvun esimerkki *John is going to New York by bus* näyttäisi tällä notaatiolla seuraavalta:



Kuva 6: "John is going to Boston by bus" — s-net -esitys.

Koska myös käsitegraafien perustana olevan eksistentiaaliset graafit ovat isomorfisia predikaattilogiikan kanssa, voidaan perusteellisesti kysyä, miksi semanttisia verkkoja kannattaa käyttää. Verkkojen etuna on toki niiden helpompi hahmotettavuus ihmisille, mutta kone ei ymmärrä käsitteiden merkityksiä sen enempää verkoista kuin kaavoistakaan. Russell & Norvig eivät näe semanttisissa verkoissa mitään erityistä logiikkaan verrattuna juuri tämän vastaavuuden vuoksi [Russell & al., sivut 316-325]. Schubertin mielestä yleistermi "semanttiset verkot" on menettänyt merkityksensä johtuen formalismien kirjavuudesta ja limittymisestä. Useimmat tietämyspohjaiset järjestelmät käyttävät joka tapauksessa semanttisten verkkojen kaltaisia mekanismeja. Vain nimitykset vaihtelevat.

Toisaalta mm. Shastri puolustaa semanttisten verkkojen käyttöä niiden laskennallisella tehokkuudella. Käsitteiden väliset suhteet ja kokonaisuus ovat kuvattavissa graafinotaatiolla predikaattilogiikkaa tehokkaammin. Verkkomuotoisesta tietämuskannasta tieto on löydettävissä helpommin ja päättelyssä voidaan käyttää tavallisten logiikan päättelysääntöjen lisäksi verkkoteorian algoritmeja. Päättely voi perustua myös verkon topologiaan tai solmujen etäisyyteen (ns. ”semanttinen etäisyys”) [Shastri]. Schubertin mielestä tosin myös logiikan päättelyt ovat abstraktilla tasolla etäisyyteen perustuvia — ainakin, jos logiikan kaavan tulkitsee *s-net*-verkkona [Schubert].

Tietämyskielen tulee sisältää runsaasti päättelyä tukevia elementtejä. Hyvä esimerkki on tyyppihierarkia. Sallimalla moniperintä ja poikkeukset ylityypeissä tehtyihin määrittelyihin päättely on kuitenkin monimutkaisempaa. Lisäksi semanttiset verkot mahdollistavat rinnakkaishaun: haettaessa jotain käsitettä haku aloitetaan monesta eri verkon solmusta samanaikaisesti, jonka jälkeen sitä laajennetaan rintamahaun tapaan. Lopputuloksena todettakoon, että logiikan asemaa päättelyn perustana ei voi väheksyä, mutta formaalisti määritellyssä verkkonotaatiossa on etuja luettavuuden, tietojen luontevan organisoinnin ja laskennallisen tehokkuuden ansiosta.

## 3 Semanttinen web

Luvussa kuvataan semanttisen webin taustaa ja tavoitteita. WWW on ilmiönä suhteellisen uusi ja kasvaa räjähdysmäisesti. Webissä on tietoa valtavasti, mutta tiedon luotettavuus ja organisointi on kyseenalaista ja hallitsematonta. W3C:n Semanttinen web -ohjelman tarkoitus on tuoda WWW-sivuille varsinaisen sisällön lisäksi uudenlaista metadataa. Metadata auttaa ihmisiä ja hakukoneita ymmärtämään sivujen sisältöä ja merkityksiä.

Semanttisen webin toteutuminen vaatii useita uusia standardeja metatiedon kuvaamiseen ja tiedon rakenteellisten suhteiden esittämiseen. Sisällöllisesti valtaisa haaste on sopia sovellusaluekohtaisista ontologioista, yhteisestä käsitteistöstä ja sanastosta. Ontologioiden pohjalta voidaan luoda semantiikkaa metatiedolle. Semanttisen webin puolestapuhujien mukaan tulevaisuuden www muistuttaa yhtä suurta tietämyskantaa.

### 3.1 Miksi semanttinen web?

WWW:n tieto on tällä hetkellä esitetty niin, että sen välittäminen ja näyttäminen on helppoa ihmisille. HTML ja muut dokumenttien kuvauskielet keskittyvät lähinnä dokumenttien ulkoasun kuvaamiseen. HTML-kielen määrittäminen on kehitetty rakenteisempaan suuntaan, mutta parhaimmillaankin sillä pystytään kuvaamaan vain dokumentin looginen rakenne, ei dokumentin tietojen merkitystä. Tämän takia Internetin hakukoneet keskittyvät lähinnä sivulla olevien sanojen indeksointiin. Tämä johtaa tietoa haettaessa valtavaan määrään hakutuloksia, jotka eivät kuitenkaan välttämättä sisällä hakijaa kiinnostavaa tietoa.

Tekoälyssä tiedonhaun ongelmat on perinteisesti yritetty ratkaista kehittämällä älykkäitä järjestelmiä luonnollisen kielen tulkintaan — tähän asti huonolla menestyksellä. Semanttisen webin idea on muuttaa WWW-sivuilla oleva tieto muotoon, josta merkitykset ovat helpommin luettavissa ohjelmallisesti. Semanttinen web ei siis ole kaikenkattava asiantuntija-järjestelmä, vaan joukko tekniikoita, jotka helpottavat WWW-sivuilla olevan tietämyksen esittämistä. [Hyvönen, 2002, Berners-Lee & al.].

#### 3.1.1 WWW:n kehitysaskleet

WWW:n kehitys voidaan jakaa kolmeen sukupolveen tietojen organisoinnin suhteen [Hyvönen, 2002]. 1. sukupolven WWW:n perusidea on esityksen erottaminen sijainnista ja sen sisältö koostuu mm. HTML- ja PDF-muotoisista dokumenteista. Dokumentteihin pystytään viittaamaan ja linkittämään niitä keskenään URI:n (*Universal Resource Identifier*) avulla. Suurin osa WWW-sivuista edustaa edelleen 1. sukupolvea.

2. sukupolven WWW perustuu tiedon rakenteen erottamiseen ulkoasusta — tämä oli HTML-määrittämyksen alkuperäinen idea, jota jatkuvasti

väärinkäytetään epävalidien dokumenttien ja esitystapaan vaikuttavien merkintöjen muodossa. Rakenne ja esitystapa voidaan erottaa toisistaan kirjoittamalla dokumentit jollakin XML-määrittelyn mukaisella kielellä, joita on runsaasti lähes kaikilla ajateltavissa olevilla sovellusalueilla<sup>5</sup>.

WWW-dokumenteissa yleisimpiä XML-pohjaisia kieliä ovat XHTML (HTML-kieli muotoiltuna XML:ksi), MathML (matematiikan esittämiseen) sekä SVG (vektorigrafiikan esittämiseen). XSLT on XML-kielellä muotoiltu muunnos- haku- ja ohjelmointikieli rakenteisten dokumenttien käsittelyyn. Ulkoasuun liittyvät määrittelyt voidaan tehdä erillisellä tyyli tiedostolla, jonka käyttäjä voi halutessaan vaihtaa. (X)HTML-dokumenteille tyyli tiedostoina käytetään CSS-dokumenteja, jotka nykyään ovat jo laajassa käytössä WWW:ssä.

3. sukupolven WWW — tulevaisuuden semanttinen web -visio — perustuu merkitysten erottamiseen rakenteesta. XML-kieli tarjoaa yhtenäisen syntaksin, tuen kaikille maailman kielille Unicoden muodossa ja puumaisen tietomallin tiedon esittämiseen. Kielellä ei kuitenkaan voi määrittellä rakenteiden merkityksiä eikä siinä ole tukea päättelylle. Tähän semanttisen webin tekniikat pyrkivät vastaamaan.

### 3.1.2 Semanttisen webin tavoitteet

WWW:n kehittäjän, Tim Berners-Leen mukaan semanttisen webin tavoite on kehittää standardeja ja tekniikoita, jotka auttavat ymmärtämään Webissä olevaa tietoa koneellisesti. Tavoitteena on tuki nykyistä monipuolisemmalle tiedon etsimiselle, datan integroinnille eri lähteistä, navigoinnille ja tehtävien automatisoinnille. [Berners-Lee & al.]

Koivunen & Miller kuvaavat artikkelissaan periaatteita, jonka päälle semanttinen web rakentuu W3C:n tavoitteiden mukaan: [Koivunen & al.]

- **Kaikkeen voidaan viitata URI:lla.** Kuka tahansa voi luoda osoitteen WWW-avaruuteen ja sanoa, että se viittaa johonkin fyysisessä maailmassa. Lisäksi mihin tahansa WWW:ssä olevaan dokumenttiin tai (jos kyseessä XML-dokumentti) sen osaan voidaan viitata URI:lla. Erityisesti myös määrittelyt sanastoille, joilla dokumentin metatietoa kuvataan, ovat URI:lla määriteltyjä.
- **Resursseilla ja linkeillä voi olla tyyppi.** Tämänhetkinen Web koostuu resursseista (esim. WWW-sivut) ja linkeistä niiden välillä. Hakukoneet voivat siirtyä linkkien kautta dokumentista toiseen, mutta eivät pysty päättelemään, miten eri dokumenttien tiedot liittyvät toisiinsa. Tyypitetty resurssit ja linkit mahdollistavat kehittyneemmän tiedonhaun (resurssi- ja linkkityyppien merkitys täytyy tietysti määrittellä erikseen). Tyyppien myötä WWW tulee muistuttamaan

<sup>5</sup>XML-pohjaisia kieliä on luokiteltu esim. sivustolla <http://www.xml.org/xml/registry.jsp>.

semanttista verkkoa, jossa resurssit ovat verkon solmuja ja linkit kaaria.

- **Epätäydellinen tieto sallitaan.** Semanttinen web perustuu nykyisen WWW:n tapaan 1-suuntaisiin linkkeihin. Dokumenttien kirjoittajat voivat vapaasti viitata toistensa dokumentteihin, eikä viitatus dokumentin tekijän tarvitse huolehtia viittaavista dokumenteista. Semanttisessa webissä kuka tahansa voi vastaavasti väittää mitä tahansa mielivaltaisista web-resursseista. Työkalujen on sopeuduttava tilanteeseen, jossa johonkin resurssiin ei saada yhteyttä tai resurssi on siirretty uuteen paikkaan.
- **Ei absoluuttista totuutta.** Tunnetusti osa nykyisen WWW:n tiedosta on virheellistä tai jopa tarkoituksellisesti harhaanjohtavaa. Näin tulee olemaan myös semanttisessa webissä, mutta dokumenttien luotettavuuden varmistamiseksi ollaan luomassa tekniikoita: tiettyjen henkilöiden tai organisaatioiden sivuilla oleva tieto voidaan määrittellä luotettavaksi ja käyttäjät voivat kommentoida (annotoida) sivujen sisältöä.
- **Evoluutio on tuettu.** Semanttisen webin lähtökohtana on muuttuva ja kehittyvä informaatio, jota eri lähteet voivat tuottaa toisistaan riippumatta — jopa eri sanastoilla. Samalla sivustolla voidaan käyttää useita eri ontologioita XML:n nimiavaruus-käsitteen avulla ja niitä voidaan myös limittää.
- **Minimalismi.** W3C:n tavoite on pitää standardointi semanttisen webin suhteen mahdollisimman yksinkertaisena. Lisäksi olemassaolevia standardeja käytetään hyödyksi: yleistä metadataa voidaan esittää Dublin Coren<sup>6</sup> määrittelyihin perustuen ja ontologioita varten luotu OWL-kieli perustuu kahteen olemassaolevaan ontologiakieleen: DAML (*DARPA Agent Markup Language*) ja OIL (*Ontology Inference Layer*).

Ottaen huomioon W3C:n nykyisten standardien kasvavan määrän eri versioineen on kyseenalaista, pääseekö W3C koskaan viimeiseen tavoitteeseen. URI-viitteisiin liittyvää kritiikkiä käsitellään luvussa 4.1.3.

## 3.2 Semanttisen webin perustekniikat

Semanttisen webin tekniikat muodostavat kerrosarkkitehtuurin. Alemmat kerrokset vastaavat nykyistä WWW:tä lisättynä metadatakuvauksilla.

---

<sup>6</sup><http://dublincore.org/>



Merkitys metadatalle saadaan ontologioista, joiden päälle voidaan edelleen määritellä päättelysääntöjä. Hierarkian huipulla ja edellytyksenä kaupallisille palveluille ovat luottamukseen liittyvät tekniikat. Semanttisen webin kerrokset on esitetty taulukossa 2. [Hyvönen, 2002]

Kerros	Tekniikoita
Luottamus	Digitaalinen allekirjoitus, annotoinnit.
Logiikka	KIF, CycL, RuleML.
Ontologia	OWL, DAML+OIL.
Metadata	RDF(S), aihekartat.
Rakenne	XHTML, XML(S), XSL(T).
Internet	Unicode, URI.

Taulukko 2: Semanttisen webin arkkitehtuuri.

Tässä luvussa käsitellään ensisijaisesti W3C:n määrittelemiä RDF(S)- ja OWL -kieliä. Logiikkakerroksella käytettävää standardikieltä ei ole vielä määritelty, mutta taulukossa mainitut kielet soveltuisivat ilmaisukykyensä puolesta logiikkakerrokselle. Aihekartat (*topic maps*) ovat vaihtoehtoinen formalismi metadatan ja yksinkertaisen ontologiatiedon kuvaamiseen. Aihekarttoja käsitellään luvussa 3.3.

### 3.2.1 Metadata ja RDF

Metadata on tietoa tiedosta — tässä tapauksessa WWW:ssä olevista dokumenteista tai laajemmin, *resursseista*. Resurssi tarkoittaa tässä mitä tahansa WWW:n kautta saatavissa olevaa, esim. (X)HTML-sivua tai sen osaa, kuvaa tai sähköpostiosoitetta. Resurssi voi kuvata myös tunnistettavissa olevaa todellisen maailman oliota, jolle ei ole vastinetta WWW:ssä, esim. taideteos. Resurssiin viitataan URI:lla. Metadata ei yleensä näy käyttäjälle, vaan se on tarkoitettu koneen käsiteltäväksi (esim. WWW-hakukoneet).

Hyvönen, Harjula ja Virtanen esittävät eri tapoja metatiedon luokitteluun. Metadata voi olla *implisiittistä*, sisältyen suoraan dokumentin rakenteeseen (Esim. HTML-kielen <h1>-elementti ilmaisee pääotsikkoa) tai *eksplisiittistä*, jota eri sovellukset pystyvät käsittelemään ja tulkitsemaan. Semanttisen webin tekniikat keskittyvät nimenomaan eksplisiittiseen metadataan.

Metadata-arkkitehtuureja voi tarkastella joko resurssin tai asiakasohjelman näkökulmasta. Resurssin näkökulmasta metadata voi olla *ulkoista* (erillinen kuvausdokumentti, esim. RDF-kuvaukset ovat usein ulkoisia) tai *upotettua* (metadata dokumentin sisällä, esim. <meta>-elementit HTML-dokumenteissa). Asiakasohjelman näkökulmasta metadata voi olla *keskitettyä* (Esim. Web-palveluita kuvaava UDDI-rekisteri) tai *hajautettua* (käyttäjien WWW-sivulle tekemät annotoinnit). [Hyvönen & al.]

W3C:n kehittämä RDF-kieli on XML-pohjainen yleiskäyttöinen malli resurssien kuvaamiseen. RDF-kuvaus on yleistys WWW-sivuilla käytetyille `<meta>`-elementeille. RDF mahdollistaa `<meta>`-elementtejä monipuolisemman tiedon liittämisen resursseihin, jopa niin, että tiedon kuvaamisessa käytettävät käsitteet voivat olla eri lähteistä.

RDF-kuvaukset ovat kolmikkoja (*Resurssi, Ominaisuus, Arvo*), joista jokainen voi edelleen olla resurssi. Kielitieteen käsitteillä resurssi voidaan ajatella subjektiksi, ominaisuus predikaatiksi ja arvo objektiksi. Predikaattilogiikassa kolmikko vastaa kaksipaikkaista predikaattia. Resurssi-kuvausten voidaan tulkita myös suunnattuna, tyypitettynä graafina, aivan kuten semanttisilla verkoilla.

Esimerkkinä RDF-kuvauksesta esitetään eräs mahdollinen kuvaus tästä tutkielmasta:

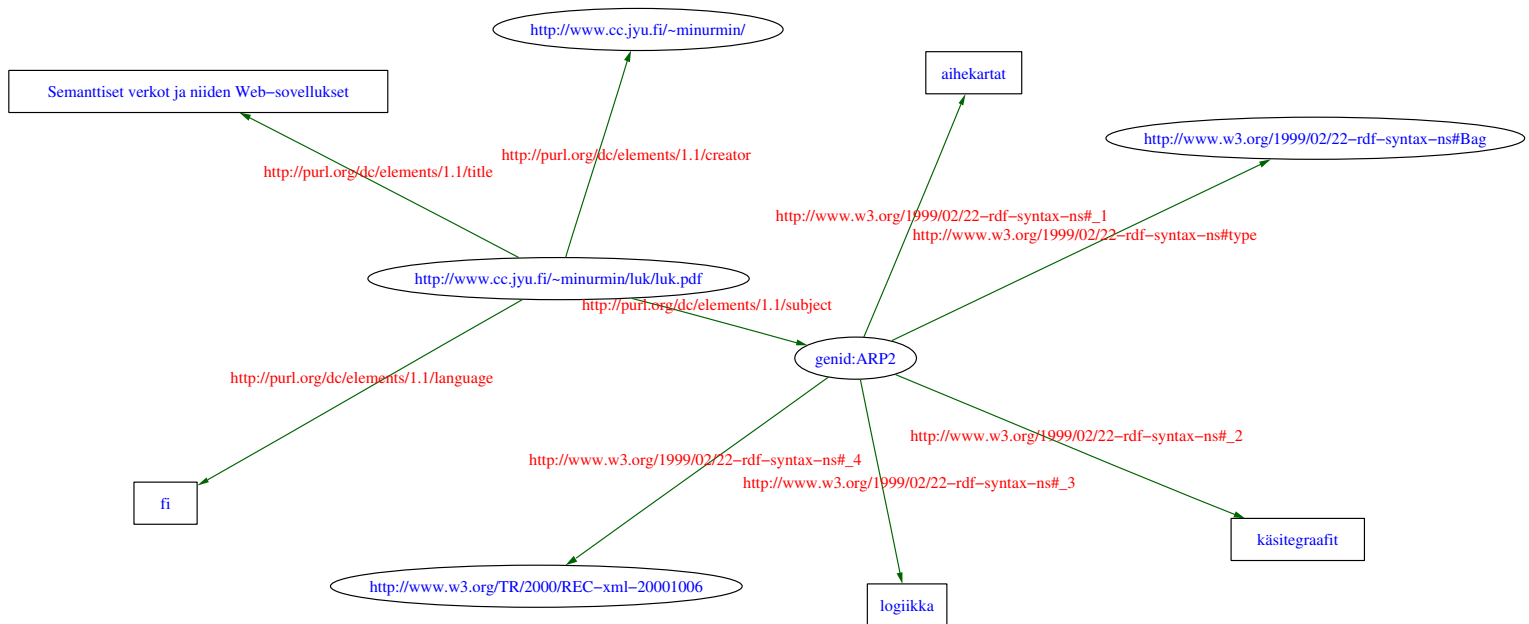
```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.cc.jyu.fi/~minurmin/luk/luk.pdf">
    <dc:title>Semanttiset verkot ja niiden Web-sovellukset</dc:title>
    <dc:creator rdf:resource="http://www.cc.jyu.fi/~minurmin/" />
    <dc:subject>
      <rdf:Bag>
        <rdf:li>aihekartat</rdf:li>
        <rdf:li>käsitegraafit</rdf:li>
        <rdf:li>logiikka</rdf:li>
        <rdf:li rdf:resource="http://www.w3.org/TR/2000/REC-xml-20001006" />
      </rdf:Bag>
    </dc:subject>
    <dc:language>fi</dc:language>
  </rdf:Description>
</rdf:RDF>
```

Kuvauksessa käytetään kahta XML-nimiavaruutta: RDF-kielen elementit on määritelty `rdf`-nimiavaruudessa, `dc`-nimiavaruus on Dublin Core -organisaation määrittelemä ehdotus kaikille dokumenteille yhteisestä metatiedosta.

`about`-attribuutti määrittää resurssin, jota RDF-kuvaus koskee. Dokumenttiin liitettävät suhteet (ominaisuudet) kuvataan `Description`-elementin sisällä. Formaali kuvaus suhdetyypeistä on `dc`-nimiavaruuden määrittämissä. Suhteiden arvot ovat elementtien sisällä olevia arvoja. Kuvauksista huomattakoon `Creator`-elementti, joka sisältää arvonaan resurssilinkin tekijän kotisivulle sekä `Subject`-elementti, jonka arvot ovat RDF-kielen `bag`-tietorakenteessa (joukko tai järjestämätön lista). Subject-listassa oleva XML-avainsana on edelleen resurssilinkki. [RDF]

Kuvassa 7 on esitetty RDF-kuvausta vastaava graafi. Kuva on luotu käyttäen W3C:n RDF-validaattoria.<sup>7</sup>

<sup>7</sup><http://www.w3.org/RDF/validator/>



Kuva 7: RDF-graafi.

Kuvan perusteella RDF-kuvausten yhteys semanttisiin verkkoihin on ilmeinen. RDF-kuvausten ja semanttisten verkkojen yhteyttä pohditaan tarkemmin luvussa 4.1.1.

Toinen merkittävä tekniikka metadatan (ja osittain myös ontologioiden) kuvaamiseen ovat ISO:n kehittämät aihekartat. Aihekarttoja käsitellään luvussa 3.3.

### 3.2.2 RDFS ja ontologiat

RDF tarjoaa tavan yksinkertaisten väitteiden ilmaisemiseen resursseista nimetyillä ominaisuuksilla ja arvoilla. RDF ei kuitenkaan itsessään sisällä tietoa käsitteistä tai ominaisuuksista, niiden määrittelyä varten W3C on kehittänyt RDF Schema -kielen [RDF]. RDFS sisältää mekanismin sovel- lusalueen sanaston määrittelyyn.

RDFS-kieli on rakennettu RDF:n päälle siten, että RDFS-kuvaukset ovat myös RDF-kuvauksia, mutta käyttävät RDFS:n omaa sanastoa, kuten käsitettä `subClassOf`. RDFS-kielillä resurssit voidaan määrittellä kuuluviksi tiettyihin luokkiin (jotka ovat myös resursseja). Luokkiin voi myös liittää uusia resursseja ominaisuuksien muodossa. Luokat ja ominaisuudet muodostavat tyyppihierarkian. Mekanismi muistuttaa olio-ohjelmoinnin rakenteita tai kehysjärjestelmiä. Mekanismia kuvataan tarkemmin luvussa 4.1.1.

RDFS on yksinkertainen, mutta ilmaisukyvyltään rajoitettu tapa ontologioiden kuvaamiseen. Monipuolisempia väitteitä ja käsitteiden mää-

rityksiä varten on kehitetty ontologiakieliä, joilla voidaan ilmaista esim. seuraavia väitteitä:

- Ominaisuuksien määrärajoitteet (esim. henkilöllä oltava tasan yksi biologinen isä).
- Ominaisuuksien transitiivisuus ( $(a \circ b \wedge b \circ c) \Rightarrow a \circ c$ ).
- Tietyn ominaisuuden määrittäminen tunnisteeksi.
- Eri resurssien merkitysten yhdistäminen.
- Mahdollisuus määrittellä uusia luokkia aiempien luokkien yhdisteinä tai leikkauksina.

Ontologiakieliä ovat DAML, OIL ja OWL, jonka W3C on kehittänyt edellisten pohjalta. OWL-kielestä on määritelty kolme eri versiota (OWL Lite, OWL DL, OWL Full), joista OWL DL (Description Logic) tarjoaa kohtuullisen ilmaisuvoiman ja laskennallisesti tehokkaan toteutuksen. OWL DL tukee kaikkia OWL-kielen määrittelyjä, mutta niitä voi käyttää rajoitusti. Esimerkiksi luokka voidaan periä useasta yliluokasta (moniperintä), mutta luokka ei saa olla toisen luokan (metaluokka) ilmentymä, mikä olisi sallittua OWL Full -kielen mukaan. [OWL]

OWL-kielen looginen perusta on kehysjärjestelmien tapaan terminologialogiikoissa. Niitä käsitellään tarkemmin luvussa 2.2.2.

### 3.2.3 Päättely ja luottamus

Ontologiat kuvaavat tietyn sovellusalueen käsitteet ja niiden väliset suhteet. Ne eivät sisällä toiminnallisia sääntöjä WWW-palveluiden automaattiseen käyttöön. Toiminnallisuutta varten tarvittavat päättelyyn ja logiikkaan liittyvät standardit ovat vielä kehitteillä, eräs mahdollinen kieli päätelysääntöjen määrittelyyn on RuleML (*Rule Markup Language*).

Semanttisen webin "protokollapinon" huipulla ovat luottamukseen liittyvät standardit, jotka ovat logiikkakerroksen tapaan vielä kehitteillä. Luottamuskerroksen ideana on, että nykyisen WWW:n tapaan kuka tahansa voi sanoa mitä tahansa mistä tahansa, mutta käyttäjä voi päättää, mihin tietolähteisiin hän luottaa. Luottamusta voi delegoida eteenpäin siten, että käyttäjä voi luottaa esim. luotetun organisaation luotettaviksi ilmaisemiin lähteisiin.

Toinen lähestymistapa WWW-sivujen luotettavuuden ilmaisemiselle ovat *annotoinnit*, käyttäjien ilmaisemat kommentit tai arvostelut sivustojen sisällöstä. Sivun luotettavuutta voidaan arvioida kommenttien perusteella. Käyttäjän henkilöllisyys varmistetaan digitaalisen allekirjoituksen avulla. [Hyvönen, 2002]

### 3.3 Aihekartat

Aihekartat ovat ISO:n standardoima<sup>8</sup> kieli WWW:n resurssien kuvaamiseen. Käsitteellisesti ne sijoittuvat semanttisen webin metadatakeroille, mutta niissä on myös joitakin ontologioiden ominaisuuksia. RDF:n tapaan aihekartat voidaan määritellä XML-pohjaisena rakenteena, mutta niiden semantiikka poikkeaa jonkin verran RDF:n semantiikasta.

RDF:n perusideana on metadatan liittäminen resursseihin, kun taas aihekartat ovat laajennus kirjastoissa käytetyille aiheiden luokittelulle ja kirjojen hakemistoille. W3C:n määrykset pohjautuvat lähinnä tekoälyyn ja matemaattis-luonnontieteelliseen tutkimukseen, aihekarttojen määrittelyyn pohjalla on hakemistoihin ja sanastoihin liittyvä standardointi. Erilaisista lähtökohdista huolimatta määryksillä on sama tavoite: WWW:n tietomassan hallitseminen metadatan avulla. [Hyvönen, 2002]

Aihekarttatiedon pohjalta on helppoa luoda hakemistojen tapaisia sivustoja automaattisesti (esimerkkinä Omnigator-selain, ks. luku 4.2.3). Aihekartat muistuttavat myös epäformaaleja semanttisia verkkoja.

#### 3.3.1 Aiheet, ilmentymät ja assosiaatiot

Aihekartoilla voidaan kuvata käsitteellisiä rakenteita ja liittää niitä tietolähteisiin. Aihekartat ovat toisaalta yleistys kirjojen hakemistoille (hakemisto on aiheista, esiintyminä sivunumerot), toisaalta ne voidaan tulkita aiheista ja assosiaatioista muodostettuina semanttisina verkkoina. Standardit aihekartat ovat SGML-kielisiä dokumentteja, mutta suurin osa nykyisistä aihekartoista käyttää standardin XML-kielelle sovitettua muunnosta (*XML Topic Maps, XTM*). Aihekarttojen peruskäsitteet ovat aiheet, esiintymät ja assosiaatiot.

Aihe (*topic*) voi olla periaatteessa mitä tahansa — henkilö, sovellusalueen olio, käsite, ontologian kategoria. Myös osa aihekarttojen määrittelyssä käytettävistä käsitteistä määritellään aiheina. Aiheet ovat tyypitettyjä (vrt. semanttisten verkkojen *instance-of-operaattori*) ja niille voidaan määritellä useita aiheen määrittelystä riippumattomia nimiä (esim. synonyymit, aiheen erikieliset esitykset). Aihe voidaan tunnistaa yksikäsitteisesti subjektitunnisteen (*subject identifier*) avulla, joka on jokin URI.

Esiintymä (*occurrence*) on linkki käsiteltävään aiheeseen liittyvään resurssiin. Aihekarttojen XML-versiossa linkit esitetään XPointer-määryksellä. Aiheella voi olla useita esiintymiä, jotka voivat olla aiheiden tapaan tyypitettyjä. Esiintymätyyppejä ovat esim. WWW-linkki, kuvalinkki tai kommentti.

Assosiaatioilla voidaan liittää aiheita toisiinsa. Assosiaatiot ovat tyypitettyjä (esim. aiemmin mainittu *aiheen tyyppi* voidaan tulkita assosiaatioksi, joka liittää aiheen aiheyyppiin). Assosiaatioon osallistuvilla aiheil-

---

<sup>8</sup>ISO/IEC 13250

la on lisäksi nimetyt roolit (esim. *työsuhde*-assosiaation roolit ovat *työntekijä* ja *työnantaja*). Mekanismi muistuttaa semanttisen verkkojen suhteita sillä erotuksella, että assosiaatiot ovat aina 2-suuntaisia. Assosiaation "suunta" päätellään sen rooleista käsiteltävästä kysymyksestä riippuen.

Aihekarttastandardi sisältää edellisten lisäksi tuen eri aihekarttaelementtien samastamiseen subjektitunnisteiden avulla. Tämä mahdollistaa eri aihekarttojen yhdistämisen. Aiheita voidaan kerätä konteksteihin (*scope*), jotka määrittävät alueen, jolla tiettyä aiheetta voidaan käyttää päätelyssä. Kontekstien avulla voidaan myös poistaa aihekartoista moniselitteisyyksiä. [[Pepper](#)]

### 3.3.2 RDF:n ja aihekarttojen esityskyvystä ja muunnoksista

RDF on tietomallina selvästi aihekarttoja yksinkertaisempi. Tietomalli perustuu (*Resurssi, Ominaisuus, Arvo*)-kolmikolle minimaalisilla semanttisilla määrittäyksillä. Aihekartoilla voidaan kuvata muutamia käsitteitä, joille ei ole suoraa tukea RDF-kielissä. Näitä ovat esimerkiksi kontekstit ja n-paikkaiset assosiaatiot.

Eräs molempien mallien ilmaisuvoimaisimpia piirteitä on väitteiden ilmaiseminen toisista väitteistä. Tämä onnistuu aihekartoilla yksinkertaisesti määrittelemällä uusi aihe, jonka tunnisteeksi merkitään toinen aihe tai assosiaatio. RDF:ssä vastaava toiminto (*reification*) vaatii olemassaolevan metatietorakenteen muuttamista. Väitteen kohteena olevaa väitettä kuvaamaan on luotava `statement`-tyyppinen tyhjä solmu.

Verrattaessa RDF-mallia ja aihekarttoja on muistettava niiden erilaiset tavoitteet. Aihekartat ovat yleistys kirjojen hakemistoille ja niitä voidaan käyttää yksinkertaisena ontologian määrittäyskielenä sellaisenaan (esim. aiheet voidaan järjestää käsittehierarkiaan). RDF on tarkoituksellisesti yksinkertainen matalan tason määrittäminen, jonka päälle voidaan rakentaa ilmaisuvoimaisempia kieliä. Useimmat aihekarttojen ominaisuudet voidaan ilmaista RDFS- tai OWL-kielillä. Lisäksi OWL on yleiskäyttöisenä ontologiakielenä selvästi aihekarttoja ilmaisuvoimaisempi.

Semanttisen webin tavoitteiden kannalta ei ole toivottavaa, että WWW jakautuu kilpaileviin metatietoalueisiin. RDF on W3C:n suositus, aihekartat on ISO:n standardi. Eri kehitysorganisaatioista johtuen määrittäysten yhdistäminen ei lyhyellä aikavälillä ole realistinen tavoite. Koska molemmilla malleilla on jo käyttäjäkuntaa ja sovelluksia, muunnokset RDF- ja aihekarttapohjaisen tiedon välillä ovat tärkeitä.

RDF:n ja aihekarttojen välillä on esitetty seuraavia muunnostapoja:

- RDF-käsitteiden esittäminen aihekarttana (tai päinvastoin).
- Yleinen semanttinen kuvaus määrittäysten välillä.
- Ontologiariippuvainen kuvaus määrittäysten välillä.

Suora käsitteiden esittäminen on teknisesti yksinkertaista. RDF:n käsitteet (esim. resurssi, ominaisuus, tyyppi) määritetään aiheiksi, joita käytetään RDF-graafista muunnetussa aihekartassa. RDF-graafin solmut ovat aihekartan aiheiden tunnisteita tai esiintymiä. Muunnos säilyttää tiedon mallista toiseen siirryttäessä, mutta on tiedonhaun ja -organisoinnin kannalta tehoton. Muunnos ei myöskään hyödynnä mallien semantiikan yhtäläisyyksiä. Tulkittaessa aihekartaksi muunnettua RDF-kuvausta on ymmärrettävä RDF-mallin toiminta.

Semanttisen kuvauksen tarkoitus on olla yleiskäyttöinen kuvaus mieli-valtaisten RDF-mallien ja aihekarttojen välillä. Kuvauksen pitäisi säilyttää mallien semantiikka luonnollisella tavalla. Lähestymistavan hankalauute-na on, että kuvaus ei ole bijektio eikä edes täysi funktio: RDF-lauseista ei voi päätellä, ovatko ne aihekarttojen kannalta esiintymiä vai assosiaatioita. Toisaalta aihekarttojen korkeamman tason ominaisuuksia ei voi esittää pelkällä RDF:llä.

Ontologiariippuvainen kuvaus vastaa edellistä kohtaa, mutta nyt kuvauksessa otetaan huomioon yksittäisen mallin käsitteiden semantiikka. Lähestymistapa on selvästi toteuttamiskelpoisempi edelliseen verrattuna, mutta työläs. Jokainen muunnettava ontologia vaatii oman muunnos-funktionsa. Muunnoksen etuna on kuvauksen luonnollisuus: aihekarttois-sa ja RDF-mallissa voitaisiin käyttää samaa ontologiaa, jonka vaihtoehtoi-set esitystavat ovat RDF tai aihekartat. [[Garshol](#)]

## 4 Semanttiset verkot webin osana

Luvussa tutkitaan perinteisten semanttisten verkkojen suhdetta semanttiseen webiin. Tarkemmassa käsittelyssä ovat aihekarttojen ja epäformaalien verkkojen sekä RDFS-skeemojen ja kehysjärjestelmien suhteet.

Semanttista webiä voidaan perustellusti sanoa valtavaksi semanttiseksi verkoksi, joka käsittää potentiaalisesti kaiken WWW:ssä olevan julkisen tiedon. Tämä laajuus tuo aivan omat haasteensa ontologioiden käyttöön, päättelyyn ja luotettavuuden arviointiin. Semanttinen web sisältää valtavan lupauksen, mutta myös mittavasti riskejä.

Luvun lopussa käsitellään esimerkinomaisesti semanttista webiä ja ontologioita sivuavia sovelluksia.

### 4.1 Tietämyksen esittäminen Web-sovelluksissa

RDF-tietomalli ja aihekartat muistuttavat rakenteeltaan semanttisia verkkoja. RDFS-skeemoilla RDF:n päälle voidaan rakentaa kehysjärjestelmän kaltainen tietämyskanta, OWL-kieli lisää edelleen RDF-mallien ilmaisuvoimaa. Aihekartat muistuttavat nykyisessä muodossaan lähinnä puoliformaaleja verkkoja, mutta standardia ollaan formalisoimassa.

Semanttisen webin merkittävin eroavuus perinteisiin tietämysjärjestelmiin nähden on tiedon monimuotoisuus ja mahdollisuus viitata mieltävaltaisiin resursseihin. Ontologioiden standardointiyrityksistä huolimatta WWW tulee jatkossakin sisältämään ontologioita, jotka ovat osittain päällekkäisiä tai eri tavoin määriteltyjä.

Useiden rinnakkaisten standardien olemassaolo ja toisaalta standardointityön keskeneräisyys ovat riskejä semanttisen webin toteutumiselle. Metadatan järjestelmällinen liittäminen dokumentteihin edellyttää hyvin määriteltyjä ontologioita, sääntöjärjestelmiä ja keinoja resurssien luotettavuuden arviointiin. Standardien käyttöönotto vaatii helppokäyttöisiä työkaluja ja aidosti hyödyllisiä sovelluksia.

#### 4.1.1 RDF(S) ja aihekartat semanttisina verkkoina

Luvussa 3.2.1 todettiin RDF-mallien muistuttavan ainakin ulkoisesti semanttisia verkkoja. Tämä on tarkoituksellista: RDF-tietomalli on laadittu samojen periaatteiden mukaisesti kuin oliojärjestelmät, terminologialogiikat ja kehysjärjestelmät — tosin rajoittuneemmassa muodossa. Kehysten etuna on ymmärrettävyys ja kohtuullinen ilmaisuvoima, joka on formalisoitu terminologialogiikoina. RDF:n resurssit vastaavat kehysten ilmentymiä, ominaisuudet attribuutteja ja tyypit luokkia. RDFS-kieli lisää malliin luokkahierarkiat ja rajoitteet ominaisuuksien luokille. [Lassila & al.]

Erona perinteisiin kehysjärjestelmiin ja olio-ohjelmointikieliin luokalle ei voi antaa suljettua määritystä RDF-kielellä. Semanttisen webin periaat-



teiden mukaan kuka tahansa voi sanoa mitä tahansa mistä tahansa. RDF-mallin kannalta tämä tarkoittaa, että luokan määrittäminen koostuu yksittäisistä ominaisuuksikuvauksista, jotka voivat sijaita fyysisesti eri paikoissa. Luokkamäärittämiä voidaan laajentaa uusia RDF-kolmikoita lisäämällä. OWL-ontologiakieli (katso luku 3.2.2) nostaa RDF(S):n ilmaisuvoiman terminologialogiikoiden tasolle, jolloin semanttista webiä voidaan toteutuksessaan pitää aitona kehysjärjestelmänä.

Vaihtoehtoinen tapa RDF(S)-kuvausten ymmärtämiseen on tulkita ne käsitegraafina (katso luku 2.2.3). Tulkinnan etuna on käsitegraafien RDF-kieltä suurempi ilmaisuvoima, joka mahdollistaa kehittyneen päättelyn ja monipuoliset hakutoiminnot. Muunnos on mahdollista tehdä läpinäkyvästi, jolloin WWW-sivun tuottaja voi tehdä metakuvauksen RDF-kielillä, mutta hakumoottori käyttäisi sisäisesti muunnettua käsitegraafiesitystä [Corby & al.]. Esimerkkinä muunnetusta kuvauksesta on kuvan 7 RDF-malli käsitegraafina:

```
[Resource : http://www.cc.jyu.fi/~minurmin/luk/luk.pdf] – {
  → (Title) → [Literal : Semanttiset verkot ja niiden web-sovellukset]
  → (Creator) → [Resource : http://www.cc.jyu.fi/~minurmin/]
  → (Subject) → [Bag] – {
    → (Li) → [Literal : aihekartat]
    → (Li) → [Literal : käsitegraafit]
    → (Li) → [Literal : logiikka]
    → (Li) → [Resource : http://www.w3.org/TR/2000/REC-xml-20001006]}
  → (Language) → [Literal : fi]}
```

RDF-mallin tapaan myös aihekartat muistuttavat semanttisia verkkoja: tieto tallennetaan verkkorakenteeseen, solmujen väliset suhteet kuvataan linkeillä (assosiaatioilla) ja semantiikkaa solmuihin ja linkkeihin saadaan tyypityksen avulla.

Aihekarttojen ero semanttisiin verkkoihin on verkon solmuihin (aiheisiin) keskittyminen linkkien kustannuksella. Tämä ilmenee siten, että assosiaatioille ei voi määrittellä esiintymiä eikä assosiaatiotyyppien semantiikkaa ole määritelty tarkasti muille kuin subclass-of ja instance-of -assosiaatiolle [Freese]. Erityisesti aiheille tai assosiaatioille ei voi määrittellä attribuutteja RDFS:n tapaan (tosin assosiaatiot voidaan tulkita eräänlaisina attribuutteina). Uusia ominaisuuksia semanttisiin verkkoihin nähden ovat kontekstit, esiintymän käsite ja tarkka määrittely eri aihekarttojen yhdistämiseen [Rath].

Aihekartat muistuttavat nykyisellään lähinnä puoliformaaleja verkkoja, mutta nykyisen standardin päälle on kehitteillä rajoitekieli TMCL *Topic Map Constraint Language*, joka lisää standardin ilmaisuvoimaa ja for-

maaliuden astetta. Garshol huomauttaa, että tulevaisuudessa aihekarttoja voitaisiin periaatteessa määrittellä myös OWL-kielellä, jos W3C:n OWL-suositukseen saataisiin mukaan aihekarttoja tukevia käsitteitä. [Garshol]

#### 4.1.2 Ontologioiden jakaminen Webissä

Semanttisen webin tekniikoiden — sekä RDF- että aihekarttapohjaisten määritysten — tärkeänä tavoitteena on mahdollistaa eri ontologioiden liittämisen ja jakaminen. Tämä mahdollistaisi saman sovellusalueen palvelujen toimivuuden, vaikka ne käyttäisivät eri sanastoa toiminnassaan. RDF-mallissa liittämisen tapahtuu yksinkertaisesti URI:en avulla. Jos eri malleissa esiintyy solmuja, jotka viittavat samaan resurssiin, resurssi ja siihen viittaavat solmut voidaan tulkita kuuluvaksi samaan malliin [Koivunen & al.]. Aihekarttoilla on vastaava aiheiden tunnistettiin perustuva samastusmekanismi [Pepper].

Ontologioiden yhdistämisessä ja jakamisessa on RDF-kielen tuesta huolimatta monia haasteita. Ontologioita voidaan määrittellä eri kielillä, joiden terminologia ja semantiikka voivat poiketa toisistaan huomattavasti. Esimerkiksi DARPA-organisaation ylläpitämän ontologiakirjaston<sup>9</sup> ontologiat on kuvattu DAML- OWL- tai RDFS-kielillä (kehyksiin perustuva kuvaus). Ontolingua<sup>10</sup>-portaalin ontologiat on kuvattu KIF-kielen variantilla (predikaattilogiikkaan perustuva kuvaus) [Fensel, sivut 66-69]. OWL-kieli selkeyttäneen ontologioiden määrittelyä WWW:ssä olevien resurssien suhteen, mutta muunnokset kielestä toiseen ovat tarpeen käsiteltäessä olemassaolevia ontologioita.

Reed ja Lenat kuvaavat ontologioiden erojen asteita Cyc-järjestelmään liitettyjen ontologioiden pohjalta seuraavasti: [Reed & al.]

1. **Terminologiset erot.** Eri nimet samoille asioille, kuvauksen perustana attribuutit / funktiot / predikaatit.
2. **Yksinkertaiset rakenteelliset erot.** Samanlaiset erillään olevat ontologiat, eri tavoin organisoidut ontologiat.
3. **Mutkikkaat rakenteelliset erot.** Toimintapredikaatit / konkretisoidut tapahtumat.
4. **Perustavanlaatuisesti erilaiset esitykset.** Totuusarvoinen logiikka / sumea logiikka.

RDF-kielen tai aihekarttojen samastuskeinot riittävät lähinnä terminologisten erojen häivyttämiseen. OWL-kielellä voidaan yhtenäistää myös yksinkertaisia rakenteellisia eroja [Garshol].

<sup>9</sup><http://www.daml.org/ontologies/>

<sup>10</sup><http://ontolingua.stanford.edu/>

Käyttökelpoisten ontologioiden määrittely on vaikeaa. On mahdotonta kehittää kaikille sovellusalueille sopivaa ontologiaa, erityisesti WWW:n kaltaisessa heterogeenisessä ympäristössä. Fensel uskoo sovellusaluekohtaisten "ontologiaryhmien" muodostumiseen. Etäisesti nykyisten uutisryhmien tapaan ontologiaryhmä koostuisi joukosta käyttäjiä, jotka ovat sopineet yhteisen ontologian käyttämisestä dokumentteihin ja niiden annotointeihin. Ryhmälle kuuluvien dokumenttien jakamisesta huolehtisi välittäjäpalvelin (*broker*) [Fensel, sivut 81-82].

Toivonen esittää vastakkaisen näkemyksen ontologioiden jakamisen suhteen. Hänen mielestään RDF(S)-kieli ei sellaisenaan sovellu laajojen ontologioiden kuvaukseen. Syynä on toisaalta kielen ilmaisuvoiman puute, toisaalta RDFS-luokkien ja ominaisuuksien käsitteellinen erottelu. Sovellusalueesta riippuen samat resurssit saattavat olla loogisesti luokan tai ominaisuuden roolissa, mikä johtaa ylimääräisiin RDF-määrittelyihin ja vaikeuttaa eri sovellusalueiden ontologioiden limittämistä. Ylemmän tason määrittelyt (esim. OWL) eivät poista ongelmaa, koska ne liittyvät suoraan RDF-tietomalliin.

Toivonen ehdottaa, että RDF(S)-kieltä käytettäisiin määrittelemään eri näkymiä toisaalla määriteltyyn laajaan ontologiaan. Näkymät ovat sovellusaluekohtaisia ja alkuperäistä ontologiaa yksinkertaisempia [Toivonen]. Ehdotus tukisi standardoidun ylitason ontologian käyttöönottoa. Perustana tälle voisivat olla esim. Sowa:n kategoriat (luku 2.1.3), Cyc-ontologia (luku 4.2.1) tai IEEE:n *Standard Upper Ontology*<sup>11</sup>.

#### 4.1.3 Semanttisen webin mahdollisuuksista ja riskeistä

Semanttinen web on käsitteenä uusi ja siihen liittyvät standardit ovat osittain vielä valmisteilla. Visio "älykkäästä WWW:stä" on innoittanut runsaasti tutkimusta ja sovelluksia ja luonut suuria odotuksia liiketoiminnan suhteen. Odotusten kärjessä ovat mm. älykkäät hakukoneet, WWW:ssä toimivat autonomiset agentit, sisällönhallinta sekä Web-palvelut (*Web services*).

Semanttiseen webiin liittyvä innostus muistuttaa monessa suhteessa 80-luvulla ja aiemmin vallinnutta tekoälyhuumaa. Aluksi asiantuntijajärjestelmiä rakennettiin semanttisten verkkojen ja kehysten pohjalta, sitten niihin liitettiin sääntöjä. Kun huomattiin, että symboliset rakenteet eivät sovellu epätäsmällisen tiedon kuvaamiseen, kehitettiin tilastollisiin malleihin ja sumeaan logiikkaan perustuvia järjestelmiä. Lopulta neuraaliverkot ja muut laskennallisesti älykkäät järjestelmät korvasivat symboliset esitykset osassa sovelluksista. Hyötyniemi arvelee, että vastaava kehityskulku toistuu myös semanttisen webin kohdalla ja tulevaisuudessa hämmöittää "neuraaliweb".

---

<sup>11</sup><http://suo.ieee.org/>

Merkittävä ero klassisen tekoälytutkimuksen ja semanttisen webin välillä on WWW:n avoin ympäristö. Koska asiantuntijajärjestelmät rakennettiin yksinään toimiviksi sovelluksiksi, jokainen järjestelmä piti rakentaa alusta alkaen itse. WWW:ssä olevien palvelujen ja ontologioiden käyttö vähentää työtaakkaa uusien järjestelmien kehittäessä. [Hyötyniemi]

RDF-mallia on kritisoitu joidenkin sen piirteiden osalta. Yksi ongelma liittyy URI:en käyttämiseen resurssien tunnistamisessa. URI voidaan toisaalta määrittellä reaalimaailman olioiden tunnistukseksi, toisaalta resurssiksi, joka kuvaa jotain oliota. RDF-malli ei erottele näitä rooleja, jolloin niitä saatetaan käyttää rinnakkain metadatakuvauksessa [Garshol]. Esimerkiksi URI <http://www.cc.jyu.fi/~minurmin/> sisältää tietoja kirjoittajasta, mutta se voidaan tulkita myös kirjoittajan identiteetiksi WWW-ympäristössä. Kuvassa 7 URI:a on käytetty jälkimmäisessä merkityksessä.

Toinen URI-osoitteisiin liittyvä ongelma on, että ne eivät ole aina pysyviä eivätkä yksikäsitteisiä. Sama resurssi (tai pikemminkin resurssin kopio) voi sijaita monessa eri paikassa verkkoa. Miten RDF-malli suhtautuu resurssiin, jolla on potentiaalisesti monta tunnistetta? URI:en pysyvyyttä kuvastanee jokaiselle WWW:n käyttäjälle tuttu tilanne, jossa tuttu WWW-sivusto on siirretty uuteen paikkaan tai lopetettu. Joskus vanhalla sivulla on ilmoitus tai linkki uusille sivuille, usein tuloksena on ns. 404-virhe, eli sivua ei löydy. W3C:n periaatteinen [Koivunen & al.] mukaan WWW:ssä tietojen yhtenäisyydestä on luovuttu rajattoman skaalauvuuden hyväksi — sama linja jatkuu myös semanttisessa webissä.

Patel-Schneider ja Siméon pitävät ongelmana epäyhtenäisyyttä semanttisen webin eri kerrosten välillä. Tämänhetkisen WWW:n perustana on XML. Sen tietomalli perustuu puurakenteeseen, jossa kaaret ovat järjestettyjä ja nimettömiä. RDF:n tietomalli on suunnattu graafi, jossa kaaret on nimetty, mutta niillä ei ole järjestystä. XML-kieltä tosin käytetään RDF-mallin esittämiseen, mutta kielessä itsessään olevaa (alkeellista) semantiikkaa ei hyödynnetä. XML-dokumenttien käsittelyyn on olemassa kyselykieli XQuery ja syntaksin määrittelykieli XML Schema, mutta RDF-malleja käsiteltäessä näitä voidaan käyttää vain rajoitetusti. [Patel-Schneider & al.]

Lisää epäyhtenäisyyksiä on ilmennyt määriteltäessä OWL-kieltä RDF-mallin ja RDFS-kielen päälle. RDF-malli tukee syntaktisesti korkeamman kertaluvun määrittelyä (väittämien käyttäminen ominaisuutena), mutta OWL-kielen DL-version päättely kattaa vain 1. kertaluvun logiikan osajoukon. OWL-kielen Full-versio tukee periaatteessa kaikkia RDF-mallin väitteitä, mutta kielen kehittäjät pitivät epätodennäköisenä, että mikään päättelykone pystyisi tukemaan kaikkia OWL Full -kielen ominaisuuksia [OWL].

Eräs ratkaisu olisi laajentaa XML-kieltä niin, että RDF-mallin piirteet pystyttäisiin kuvaamaan sillä luonnollisesti. RDF-mallia yksinkertaistettaisiin luopumalla korkeamman kertaluvun väitteistä ja korvaamalla ni-

metyt kaaret ylimääräisillä nimisolmuilla resurssisolmujen ja arvosolmujen välillä. Verkkorakennetta vastaava esitys saadaan lisäämällä solmuihin tarvittaessa ID-attribuutteja. RDFS-kielen ja OWL-kielen voisi yhdistää yhdeksi ontologiakieleksi, joka tukee myös XML Schema -kielen ominaisuuksia. [Patel-Schneider & al.]

Hyvönen [Hyvönen, 2002] muistuttaa, että lukuiset organisaatiot ovat kehittämässä osittain limittäisiä semanttista webiä koskevia standardeja (Esim RDF / aihekartat, OWL / muut ontologiakielet), vaikka toimivista sovelluksista ei ole vielä kokemuksia. Liian pitkälle menevä standardointi voi kääntyä itseään vastaan: kahden osittain päällekkäisen ”standardin” olemassaolo merkitsee, että todellista standardia ei ole.

## 4.2 Sovelluksia

Fensel luettelee semanttisen webin keskeisinä sovellusalueina tiedonhaun, tietämyksen hallinnan organisaatioiden sisällä ja välillä, verkko-kaupan (B2C-sovellukset) ja elektronisen liiketoiminnan (B2B-sovellukset, Web-palvelut) [Fensel, sivut 1-7]. Semanttinen web yksinkertaistaa myös autonomisten agenttien toimintaa [Toivonen].

Käsiteltävistä sovelluksista Cyc ja WordNet ovat molemmat sovelluksia ontologioiden esittämiseen. Cyc pyrkii formalisoimaan ihmisen jokapäiväisessä päättelyssään käyttämän arkitietämyksen, WordNet on vapaamuotoisempi tietokanta englannin kielen sanoista ja niiden suhteista. Molemmista on WWW-pohjainen käyttöliittymä saatavilla.

Ontopian tuottama Omnigator on WWW-pohjainen työkalu aihekarttojen selaamiseen ja tiedonhakuun. Selaimelle voi syöttää mielivaltaisen aihekarttatiedoston, jonka perusteella sitä voi selata WWW-sivuston tapaan.

Ontologioihin liittyviä sovelluksia on esitelty laajemmin Fenselin kirjassa ”Ontologies” [Fensel]. Semanttiseen Webiin liittyviä sovelluksia käydään läpi mm. kirjassa ”Semantic Web Kick-off in Finland” (ks. Lähteet).

### 4.2.1 Cyc

Cyc on yleiskäyttöinen ontologia, jonka kunnianhimoinen tavoite on formalisoida kaikki inhimillinen arkitietämys. Cyc-järjestelmän tietämuskanta koostuu yli 100 000 aksiomaattisesti määritellystä käsitteestä. Järjestelmää on kehitetty vuodesta 1990 alkaen ja kehitys jatkuu edelleen. Projektin laajuuden takia järjestelmän kehittäjät ovat joutuneet ottamaan kantaa useimpiin tietämyksen esittämiseen liittyviin ongelmiin [Sowa, 2001, sivut 51-55].

Cyc-tietämuskanta määritellään CycL-kielellä, joka on predikaattilogiikan laajennus. Predikaattilogiikkaa on laajennettu tyyppityksellä, uusilla kvanttoireilla (esim. *onOlemassaVähintään*), oletusarvoilla ja korkeamman

kertaluvun lausekkeilla. Lausekkeiden totuusarvot määritellään konteksteittain, joita kutsutaan CycL:ssä mikroteorioiksi [Reed & al.]. CycL-kieli on semanttisesti lähellä KIF-formalismia [Fensel, sivut 62-66].

[Individual] [Collection]

**Hacking** (Show Hacking in normal browser)

Defined in UniversalVocabularyMt BaseKB

"A specialization of ProgrammingAComputer, comprising particularly clever or expert programming feats. Contrary to popular usage of the word "hacking", this constant is not intended to refer to acts of breaking into computer systems (for that, see PenetratingAComputerNetwork and AttackByComputerOperation)."

* isa:	in Mt	Unassert
ProposedPublicConstant	BaseKB	<input type="checkbox"/>
TemporalStuffType	UniversalVocabularyMt	<input type="checkbox"/>
DefaultDisjointScriptType	UniversalVocabularyMt	<input type="checkbox"/>
ProposedPublicConstant-CommentOK	BaseKB	<input type="checkbox"/>
ProposedPublicConstant-DefinitionalGAFsOK	BaseKB	<input type="checkbox"/>

* genls:	in Mt	Unassert
Individual	UniversalVocabularyMt	<input type="checkbox"/>
ProgrammingAComputer	InformationGVocabularyMt	<input type="checkbox"/>
UsingADevice	HumanActivitiesMt	<input type="checkbox"/>
LearnedActivity	HumanActivitiesMt	<input type="checkbox"/>

\* Change current Mt from CyclistsMt Go

New Mt name: Complete

Copyright © 1995 - 2002 Cycorp. All rights reserved.

Kuva 8: Cyc-selain.

Kuvassa 8 on esimerkki Cyc-järjestelmän WWW-käyttöliittymästä, jonka avulla tietämyskantaa voi selata ja muokata. Käyttöliittymä ja Cyc:n ylemmän tason ontologia (n. 3000 käsitettä) ovat saatavilla WWW:ssä<sup>12</sup>.

#### 4.2.2 WordNet

WordNet<sup>13</sup> on ilmainen, WWW:ssä selattavissa oleva tietokanta englannin kielen sanojen merkityksistä. WordNetin taustalla ovat psykokieli-tieteelliset teoriat ihmisen kielellisestä muistista. WordNet on kehitetty Princetonnin yliopiston kognitiotieteen laboratoriossa.

WordNetin tietokantaan on määriteltä noin 100 000 käsitettä organisoituna sanaluokkiin ja kielellisiin suhteisiin. Näitä ovat esimerkiksi synonyymit, ylä- ja alakäsitteet sekä jaottelu kokonaisuuksiin ja osiin. WordNetia voidaan pitää sovellusalue-riippumattomana ontologiana.

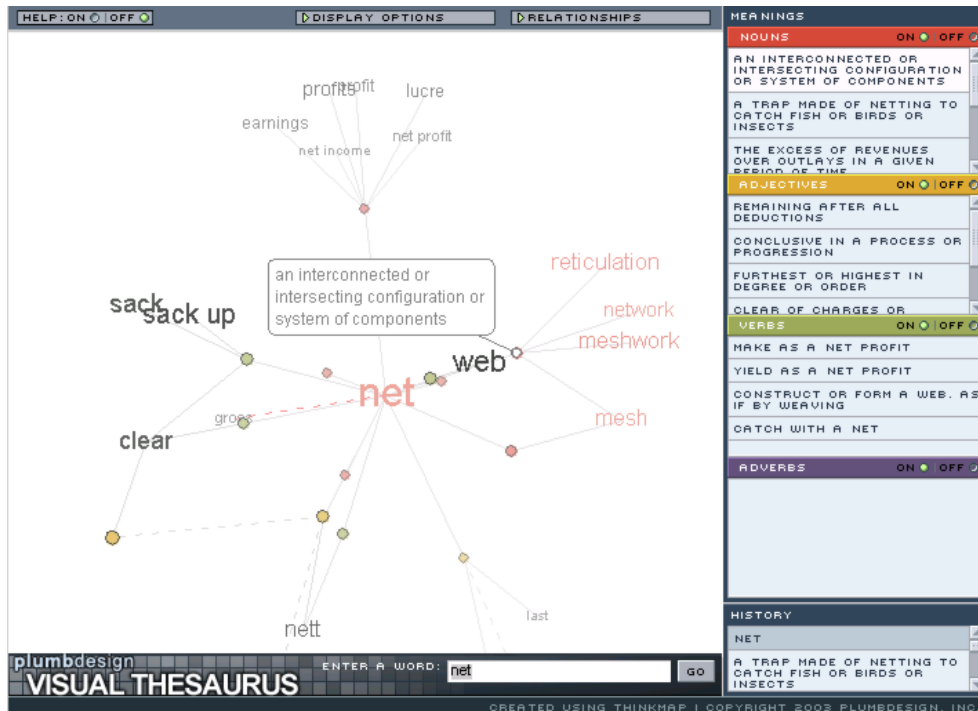
WordNetille tunnusomaista on epäformaali määrittely. Käsitteiden ja kielellisten suhteiden merkitykset on annettu luonnollisella kielellä, mikä rajoittaa automaattisen päättelyn tekoa. Syy epäformaaliuteen on järjestelmän kielitieteellinen — ei matemaattinen — perusta. [Fensel]

Plumb design -yhtiön tuottama Visual Thesaurus<sup>14</sup> on Java Applet -pohjainen graafinen käyttöliittymä WordNet-ontologian selaamiseen. Esimerkki sovelman toiminnasta on esitetty kuvassa 9.

<sup>12</sup><http://www.opencyc.org/>

<sup>13</sup><http://www.cogsci.princeton.edu/~wn/>

<sup>14</sup><http://www.visualthesaurus.com/index.jsp>



Kuva 9: Plumb design Visual Thesaurus.

### 4.2.3 Omnigator

Ontopia-yhtiön<sup>15</sup> tuottama Omnigator-selain on WWW-pohjainen sovel- lus aihekarttojen selaamiseen, hallintaan ja tiedonhakuun. Selain sisältyy *Ontopia Knowledge Suite* -työkalupakettiin, joka sisältää apuohjelmia aihe- karttojen tuottamiseen. Esimerkki Omnigator-selaimen käyttöliittymästä on esitetty kuvassa 10.

Omnigator on toteutettu Java-kielillä ja sen käyttöliittymä koostuu JSP-sivuista. Selaimen voi ladata minkä tahansa SGML- XML- tai lineaarisella notaatiolla kirjoitetun aihekartan. Lineaarinen notaatio on Onto- pian kehittämä (ei-standardi) formaatti aihekarttojen kirjoittamiseen. Se muistuttaa käsittegraafien lineaarisesta notaatiosta. OKS:n kaupallisissa ver- sioissa on myös tuki aihekarttojen käsittelyyn tietokantapalvelimen kaut- ta.

Omnigator luo aihekartan pohjalta web-sivuston, jossa aiheita niiden suhteita ja esiintymiä voi tarkastella ja selata. Aihekarttoihin voi myös kohdistaa hakuja, niitä voi yhdistellä ja viedä sovelluksesta SGML- tai XML-muodossa. [OKS]

<sup>15</sup><http://www.ontopia.net/>


**omnigator VI**  
The free topic map navigator from Ontopia. Powered by the Ontopia Knowledge Suite.

Welcome | Scripts and languages | Manage | Customise | Filter | Export | Merge | Statistics | Feedback | No schema |  
 | Query

---

**Tengwar** Type(s): Abugida

<p><b>Names (1)</b></p> <ul style="list-style-type: none"> <li>• Tengwar</li> </ul>	<p><b>Subject indicators (1)</b></p> <ul style="list-style-type: none"> <li>• <a href="http://psi.ontopia.net/iso/15924.xtm#teng">http://psi.ontopia.net/iso/15924.xtm#teng</a></li> </ul>
<p><b>Related subjects (3)</b></p> <ul style="list-style-type: none"> <li>• <b>Belongs to</b> <ul style="list-style-type: none"> <li>◦ Artificial scripts</li> </ul> </li> <li>• <b>Created by</b> <ul style="list-style-type: none"> <li>◦ J.R.R. Tolkien</li> </ul> </li> <li>• <b>Supported by</b> <ul style="list-style-type: none"> <li>◦ Unicode</li> </ul> </li> </ul>	<p><b>External resources (2)</b></p> <ul style="list-style-type: none"> <li>• <b>Article</b> <ul style="list-style-type: none"> <li>◦ <a href="http://fan.theonering.net/rolozo/tengwar/tengwar/index.htm">http://fan.theonering.net/rolozo/tengwar/tengwar/index.htm</a></li> <li>◦ <a href="http://hem.passagen.se/mansb/at/tengwar.htm">http://hem.passagen.se/mansb/at/tengwar.htm</a></li> <li>◦ <a href="http://www.egt.ie/standards/csur/tengwar.html">http://www.egt.ie/standards/csur/tengwar.html</a></li> <li>◦ <a href="http://www.geocities.com/TimesSquare/4948/tengwar/">http://www.geocities.com/TimesSquare/4948/tengwar/</a></li> <li>◦ <a href="http://www.tolkienonline.de/tengwar.html">http://www.tolkienonline.de/tengwar.html</a></li> </ul> </li> <li>• <b>Site using</b> <ul style="list-style-type: none"> <li>◦ <a href="http://www.forodrim.org/daeron/md_teng_primers.html">http://www.forodrim.org/daeron/md_teng_primers.html</a></li> </ul> </li> </ul>

 Put the mouse over this text to view the source locator(s) of the current topic.

(C) Copyright 2001-2002 Ontopia. All rights reserved. Download the latest version when this one expires!

Kuva 10: Ontopia Omnigator.

## 5 Yhteenveto

Semanttiset verkot (semantic networks) ovat joukko toisiaan lähellä olevia formalismeja tietämyksen esittämiseen. Semanttisia verkkoja on sovellettu asiantuntijajärjestelmissä, luonnollisen kielen käsittelyssä ja muissa tekoälyä hyödyntävissä sovelluksissa. Tieto kuvataan verkkorakenteena, jonka solmut ja kaaret ovat yleensä nimettyjä.

Semanttiset verkot eivät saaneet varauksetonta kannatusta tekoälytutkimuksessa. Tietämyksen esittämisessä oli pitkään vallalla kaksi koulukuntaa, verkkorakenteiden ja logiikalla ilmaistavien sääntöjen kannattajat. Myöhemmin on ymmärretty, että oikein formalisoituna semanttiset verkot ja 1. kertaluvun logiikka (mahdollisine laajennuksineen) ovat isomorfisia. Molemmilla kuvaustavoilla on etunsa, joskin verkkomainen esitys on ihmiselle luettavampi.

Semanttinen web on joukko teknologioita, joiden tarkoitus on saada WWW:n sisällön merkitykset tietokoneen ymmärrettävään muotoon. Semanttinen web koostuu joukosta määrittämiä, joilla kuvataan eri tasoilla verkon resursseja ja niiden suhteita toisiinsa. Semanttista webiä voidaan perustellusti sanoa valtavaksi semanttiseksi verkoksi, joka käsittää potentiaalisesti kaiken WWW:ssä olevan julkisen tiedon. Tämä laajuus tuo aivan omat haasteensa ontologioiden käyttöön ja jakamiseen, päättelyyn sekä luotettavuuden arviointiin verrattuna yksittäisen tekoälyjärjestelmän



tietämyskantaan.

Semanttinen web sisältää valtavan lupauksen, mutta myös mittavasti riskejä. Tässä(kin) suhteessa on havaittavissa yhtenevyyksiä klassiselle tekoälylle asetettuihin lupauksiin nähden. Tällä hetkellä tutkimuksessa on selvä innostus semanttista webiä kohtaan, mutta kuinka kauan? Semanttinen web ei ehkä pysty lunastamaan ylimitoitettuja lupauksia. WWW:n sekavaa informaatioavaruutta ei W3C:n suosituksilla hallita — erityisesti, kun sekä metadatan että ontologioiden esittämiseen on jo käytössä useita osittain päällekkäisiä ”standardeja”.

Kirjoittaja pitää epätodennäköisenä semanttisen webin toteutumista koko WWW:n laajuisena. Metadatan kirjoittaminen ilman tehokkaita työkaluja on työlästä, päättelyn tekeminen eri sivustojen elementtien välillä (toiminnasta puhumattakaan) vaatii standardoituja ontologioita, joiden kehittäminen on hyvin vaikeaa. Todennäköisempää on semanttisen webin tekniikoiden yleistyminen yksittäisten organisaatioiden sisällä, sovel-lusaluekohtaisia ontologioita hyödyntäen ja rajoitettujen sovellusten sisäl-lä (esim. www-hakusivuston generointi aihekarttamallin pohjalta). Tällai-senaankin semanttinen web tulee selkeyttämään WWW-avaruutta ja tar-joaa uusia mahdollisuuksia tietämyksen esittämiseen ja hallintaan.

## Lähteet

- [Berners-Lee & al.] Berners-Lee T., Hendler J. and Lassila O., *The Semantic Web*, Scientific American, Toukokuu 2001 <URL: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>.
- [Brachman & al.] Brachman R. J., McGuinness D. L., Patel-Schneider P. F., Resnick L. A., Borgida A., *Living With Classic: When and How to Use a KL-ONE-like Language* kirjassa "Principles of Semantic Networks", Morgan Kaufmann Publishers, 1991.
- [Corby & al.] Corby O., Dieng R, and Hébert C., *A Conceptual Graph Model for W3C Resource Description Framework* kirjassa "Conceptual Structures: Logical, Linguistic, and Computational Issues", Springer, 2000, <URL: <http://citeseer.nj.nec.com/dieng00conceptual.html>>.
- [Elmasri & al.] Elmasri R. A., and Navathe S., "Fundamentals of Database Systems", Addison-Wesley, 2000.
- [Fensel] Fensel D., "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce", Springer-Verlag, 2000, <URL: <http://citeseer.nj.nec.com/413498.html>>.
- [Freese] Freese E., *Using Topic Maps for the representation, management & discovery of knowledge*, XML Europe 2000, <URL: <http://www.gca.org/papers/xml europe2000/papers/s22-01.html>>.
- [Garshol] Garshol, L. M. *Living with topic maps and RDF*, XML Europe, 2003, <URL: <http://www.ontopia.net/topicmaps/materials/tmrdf.html>>.
- [Hyvönen, 1988] Hyvönen Eero, "Asiantuntijajärjestelmien tietämystekniikka", Knowledge Engineering, 1988.
- [Hyvönen, 2002] Hyvönen E., *The Semantic Web — The new Internet of Meanings* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002, <URL: <http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>>.
- [Hyvönen & al.] Hyvönen E., Harjula P. and Viljanen K., *Representing metadata about Web Resources* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002.
- [Hyötyniemi] Hyötyniemi H., *Reality and Truth in the Semantic Web* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002.

- [Kivelä & al.] Kivelä A. and Hyvönen E., *Ontological Theories for the Semantic Web* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002.
- [Koivunen & al.] Koivunen M.-R. and Miller E., *W3C Semantic Web Activity* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002.
- [Lassila & al.] Lassila O. and McGuinness D. L., *The Role of Frame-Based Representation on the Semantic Web*. Tekninen raportti, KSL, 2001. <URL: <http://www.ksl.stanford.edu/people/dlm/etai/etai-abstract.html>>
- [OKS] *The Ontopia Knowledge Suite — An Introduction*, Ontopia, 2003, <URL: <http://www.ontopia.net/ontopia/texts/product-wp.html>>.
- [OWL] *OWL Web Ontology Language Overview*, W3C:n työluonnos, 31.3.2003, <URL: <http://www.w3.org/TR/owl-features/>>.
- [Patel-Schneider & al.] Patel-Schneider P. F. and Siméon J., *Building the Semantic Web on XML*, International Semantic Web Conference, 2002, <URL: <http://www-db.research.bell-labs.com/user/pfps/papers/building.ps>>.
- [Pepper] Pepper S., *The TAO of Topic Maps — Finding the way in the age of infoglut*, XML Europe, 2000, <URL: <http://www.ontopia.net/topicmaps/materials/tao.html>>.
- [Rath] Rath H., *Semantic Resource Exploitation with Topic Maps* kirjassa "Sprach- und Texttechnologie in digitalen Medien - Proceedings der GLDV-Frühjahrstagung 2001", Norderstedt: Books on Demand, 2001, <URL: <http://www.uni-giessen.de/fb09/ascl/gldv2001/proceedings/pdf/GLDV2001-rath.pdf>>.
- [RDF] *RDF Primer*, W3C:n työluonnos, 23.1.2003, <URL: <http://www.w3.org/TR/rdf-primer/>>.
- [Reed & al.] Reed S. L. and Lenat D. B., *Mapping Ontologies into Cyc*, Cycorp, Inc, 2002, <URL: [http://www.cyc.com/doc/white\\_papers/mapping-ontologies-into-cyc\\_v31.pdf](http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf)>.
- [Rich] Rich E., "Artificial Intelligence", McGraw-Hill, 1983.
- [Russell & al.] Russell S. and Norvig P., "Artificial Intelligence", Prentice-Hall, 1995.

- [Schubert] Schubert L. K., *Semantic Nets Are in the Eye of the Beholder* kirjassa "Principles of Semantic Networks", Morgan Kaufmann Publishers, 1991.
- [Shastri] Shastri L., *Why Semantic Networks?* kirjassa "Principles of Semantic Networks", Morgan Kaufmann Publishers, 1991.
- [Sowa, 1991] Sowa J. F., *Toward the Expressive power of Natural Language* kirjassa "Principles of Semantic Networks", Morgan Kaufmann Publishers, 1991.
- [Sowa, 2001] Sowa J. F., "Knowledge Representation — Logical, Philosophical and Computational foundations", Brooks/Cole, 2001.
- [Toivonen] Toivonen S., *Using RDF(S) for Multiple Views into a Single Ontology* kirjassa "Semantic Web Kick-Off in Finland", HIIT Publications, 2002.
- [Woods, 1991] Woods W. A., *Understanding Subsumption and Taxonomy: A Framework for Progress* kirjassa "Principles of Semantic Networks", Morgan Kaufmann Publishers, 1991.

## Lisämateriaalia

- Hyvönen E., Karanta I. ja Syrjänen M. (Toim), "Tekoälyn ensyklopedia", Gaudeamus, Hämeenlinna, 1993.
- Digitaalisen median erityiskysymyksiä. Seminaarin aihepiiri: Semanttinen Web. Jyväskylän yliopisto, tietojenkäsittelytieteiden laitos, 2002  
<URL: <http://www.cs.jyu.fi/~airi/opetus/SemanttinenWeb.html>>.
- Käsitteellisen mallintamisen kurssi, Joensuun yliopisto, tietojenkäsittelytieteen laitos, 2002, <URL: <http://cs.joensuu.fi/pages/marjomaa/CMkurssi/CM.html>>.
- Semantic Web In Finland — Älykäs WWW Suomessa. Helsingin yliopisto, tietojenkäsittelytieteen laitos, 2002, <URL: <http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/>>.
- Sowa John F., *Semantic Networks*, yleisesittely, 2002,  
<URL: <http://www.jfsowa.com/pubs/semnet.htm>>.
- Tietämystekniikan seminaari, syksy 2001: Tietämyksen esittäminen, ontologiat ja Semantic Web. Teknillinen korkeakoulu, tietojenkäsittelyopin laboratorio,  
<URL: <http://www.cs.hut.fi/~sto/seminaarit/SW/>>.
- Semantic Web. Portaali, W3C, 2001,  
<URL: <http://www.w3.org/2001/sw/>>.

Lähteiden ja lisämateriaalin linkit viitattu 21.5.2003.