



Esitelmä Semantic Web Kick-Off in Finland -tilaisuudessa 2.11.2001, Helsingin yliopisto, Porthania, Helsinki.

Semantic Web

- kohti uutta merkitysten Internetiä

Eero Hyvönen

Helsingin yliopisto ja Helsinki Institute for Information Technology (HIIT)
Tietojenkäsittelytieteen laitos
PL 26, 00014 HELSINGIN YLIOPISTO
eero.hyvonen@cs.helsinki.fi

Nykyinen WWW on kehitetty ihmislukijaa varten. Kone ei verkon sisältöjä juurikaan voi ymmärtää, ainostaan tarjota niitä ihmisen tulkittavaksi. Sisältöjen automaattinen tulkinta on kuitenkin ensiarvoisen tärkeää älykkäiden ja helpokäyttöisten Internet-sovellusten kehittämiseksi. Semantic Web, merkitysten Internet, on visio seuraavan polven älykkästä Internetistä, jonka käyttäjinä ovat ihmisten ohella koneet. Visio on nopeasti muuttunut kansainväliseksi tahtotilaksi sekä tutkimus- ja standardointiohjelmiiksi. Käytännön tasolla Semantic Web konkretisoituu ontologia- ja XML-perustaisiksi "semanttisiksi" standardeiksi ja välineiksi, joiden avulla WWW:ssä voidaan esittää sisältöjä ja toteuttaa seuraavan polven älykkäitä Internet-sovelluksia.

1. Merkityksen ongelma

Nykyisen WWW:n tiedot on pääosin kuvattu siten, että niitä voidaan helposti välittää ja näyttää ihmiselle. Esimerkiksi HTML ja PDF ovat dokumenttien ja sisältöjen ulkoisen muotoilun kieliä. Ulkoasua ja dokumentin rakennetta kuvaavat merkkaukset, esimerkiksi HTML:n otsikkomerkaus <H1>Otsikko</H1>, eivät kerro mitään merkkauksen sisällöstä ilman ihmisen tulkintaa.

Internetin käyttäjinä eivät kuitenkaan ole pelkästään ihmiset, vaan yhä enevässä määrin koneet, erilaiset sovellusohjelmistot, hakukoneet (search robot), sähköisen kaupan agentit (shopbot), verkkomönkijät (web crawler) jne. Niille Internetin sisältöjen hyödyntäminen on vaikeaa, sillä verkossa olevan sekalaisen ja jäsentämättömän (unstructured) tiedon tulkitseminen, "ymmärtäminen", on vaikeaa.

Tiedon haku, tunnistaminen ja louhinta

Ongelmaa on lähestytty mm. tiedon haun (information retrieval), tunnistamisen (information extraction) ja louhinnan (data mining) keinoin.

Tiedon haussa (Korfhage, 1997) on kyse käyttäjän tietotarpeen tyydyttävän tiedon hakemisesta tietomassasta, mikä edellyttää merkitysten tunnistamista. Internetissä tiedonhaku perustuu kahteen päämenetelmään:

- Haluttua tietoa voidaan etsiä assosiativisia linkkejä yksi kerrallaan seuraamalla ("surffailu").
- Hakukoneissa (esimerkiksi www.altavista.com, www.lycos.com, www.google.com) voidaan etsittäviä sivuja luonnehtia avainsanojen perusteella ja tuottaa näin osumalistoja (hit list) tarkempaa analysointia varten.

Sanafrenvenssejä laskemalla ja verkon rakenteen perusteella (esim. Google) voidaan löydettyjä dokumentteja, joita tyypillisesti on hyvin paljon, järjestää prioriteetin mukaiseen järjestykseen.

Tiedon ekstrahoinnissa (Pazienza, 1997) ideana on dokumenttien koneellinen käsittely siten, että niiden merkitysrakenne saadaan selville. Voidaan esimerkiksi tunnistaa luonnollisen kielen sanoja, näiden morfologisia muotoja, lause- ja dokumenttirakenteita ja semanttisia hahmoja. Näiden perusteella on esimerkiksi mahdollista siivilöidä tietomassasta esiin haluttu osa, vaikkapa etsiä uutissähkeiden tai pörssitiedotteiden virrasta eri käyttäjien kiinnostusprofiileihin sopivia uutisia.

Tiedon louhinnan (Han, Kamber, 2000) tekniikoilla voidaan tietomassasta löytää tai oppia uutta sisäistä rakennetta ja tämän perusteella tuottaa esimerkiksi säännöstöjä uusien tietojen analysointia varten tai jäsentää tietoja samankaltaisiin ryhmiin. WWW:n yhteydessä puhutaan myös verkon louhinnasta (web mining) (Cooley et al., 1997), jolloin oppimisen tai automaattisen tunnistamisen kohteena ovat joko verkon rakenne ja sisällöt (content mining) tai verkossa olevien toimijoiden käyttäytyminen (usage mining). Tämä mahdollistaa esimerkiksi adaptiivisten, käyttäjän tottumuksiin mukautuvien sivustojen kehittämisen.

Näissä ratkaisumalleissa lähestytään merkityksen ongelmaa ihmisen tavoin yrittämällä tulkita verkon sisältöjä sellaisina kuin ne ovat. Tulkitseminen on kuitenkin usein hyvin vaikeaa mm. seuraavista syistä:

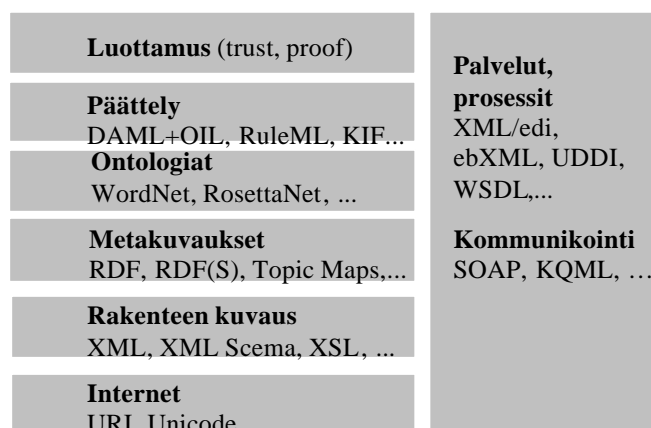
- Dokumentteissa käytetty kieli on vapaamuotoista luonnollista kieltä, joka on muodoiltaan ja merkityksiltään rikasta. Se voi myös sisältää puutteita ja suoranaisia virheitä. Jo tekstin rakenteiden tunnistaminen on laskennallisesti hankalaa merkityksistä puhumattakaan.
- Suuri osa WWW:n sisällöstä ei ole luonteeltaan tekstuaalista, esimerkiksi kuvat, äänet ja musiikki, videot ja ohjelmistokomponentit. Tällaisten ei-tekstuaalisten talletusmuotojen, esimerkiksi bittikartana esitetyn kuvan, koneellinen ymmärtäminen voi olla vieläkin haastavampaa kuin luonnollisen kielen.
- WWW-sivujen sisältöjä ei voida tulkita pelkästään sivuilla olevan tiedon perusteella, vaan ihmismäinen tulkinta perustuu yleensä dokumenttien ulkopuoliseen taustatietoon ja elämäkokemukseen. Tällaisen kontekstitiedon ja yleisen arkitietämyksen (common sense) opettaminen koneelle vähänkään yleisemmässä muodossa on osoittautunut tekoälytutkimuksessa erittäin vaikeaksi (Russel, Norvig, 1995; Hyvönen et al., 1993).

Seurauksena on, että dokumenttien automaattinen tulkinta ja tiedon luohinta on käytännössä mahdollista vain rajoitetusti ja tarkasti määritellyillä sovellusalueilla.

Semantic Web -ratkaisumalli

WWW:n koneellisen tulkinnan perusongelma on, että verkon tieto on alunperin talletettu vaikeasti tulkittavassa vapaassa muodossa. Voimme kehittää yhä älykkäämpiä järjestelmiä sisältöjen ymmärtämistä varten, mutta miksemme toisaalta voisi tulla konetta vastaan ja tallettaa tietoja verkkoon heti sellaisessa muodossa, joka tekee niiden algoritmisen hyödyntämisen helpoksi? Tämä voidaan tehdä paitsi ihmisen toimesta käsityönä myös automaattisesti.

Semantic Web -teknologioiden (Fensel, 1999, 2001; Ahmed et al., 2001) ideana on, että seuraavan polven tietoverkkojen tiedot ja rakenne koodaan siten, että niiden sisältöön (merkitykseen, semantiikkaan) päästään helpommin algoritmisesti käsiksi. Tämä mahdollistaa aiempaa olennaisesti älykkäämmän verkon käytön ja palveluiden kehittämisen (Levy, Weld, 2000) erilaisten ohjelmistoagenttien ja sovellusten kautta. Tuloksena syntyy nykyisen Internetin seuraava sukupolvi: älykäs merkitysten Internet.



Kuva 1. Semantic Web -teknologioita.

2. Semantic Web: käsitteelliset tasot

Semantic Web ei ole vain visio tulevasta vaan jo tänään joukko konkreettisia teknologioita, joiden avulla

visiota ollaan ryhdytty toteuttamaan. Näitä voidaan ryhmitellä eri tasoille kuvan 1 mukaisesti. Teknisenä perustana toimivat WWW ja sen XML-perustaiset teknologiat rakenteisille dokumenteille. Näiden avulla on luotu ylempitasoisia semanttisia metakuvauskieliä verkkojen resurssien kuvaamiseksi. Semantic Web -konseptin ydintä ovat myös ontologiat, eri sovellusalueilla tarvittavat terminologiset määrittelyt. Käsitteistöjen täsmällistä ilmaisua varten on kehitetty tietämysteknisiä esityskieliä ja välineitä. Ontologioiden avulla esitettyjen olioiden hyödyntämiseksi voidaan käyttää päättelytason standardeja ja järjestelmiä. Tätäkin korkempaa käsitteellistä tasoa edustaa luottamus verkon resursseihin. Oman alueensa muodostavat järjestelmien välisessä kommunikoinnissa sekä palveluiden ja prosessien kuvaamisessa tarvittavat menetelmät.

Semantic Web -välineet ja -standardit ovat tietysti vain teknologiaperusta, joka ei ratkaise merkityksen esittämisen ja hyödyntämisen ongelmaa niinkuin vasarakaan ei ratkaise talon rakentamisen ongelmaa. Kokemuksia semanttisen tietojenkäsittelyn vaikeuksista on saatu mm. 80-luvulla laajoista tekoäly- ja asiantuntijajärjestelmien rakentamishankkeista, joiden tuloksena ei syntynytäkään yliälykkäitä koneita. Semantic Web -konseptissa ei kuitenkaan ole tavoitteena feigenbaumilainen unelma ihmistä etevämmistä asiantuntijajärjestelmistä, vaan nöyrästi siitä, miten saada nykyinen tiedon visualisointiin (HTML) keskittyvä WWW älykkäämmäksi lisäämällä siihen eritasoisia sisällöllisiä kuvauksia, semantiikkaa.

Tarkastelemme seuraavassa Semantic Webin ideaa teknologiaalähtöisesti sen eri tasoilla kuvan 1 perusteella.

3. Rakenteen kuvaaminen

Resurssien verkko

Internetin keskeisimpiä käsitteitä on URI-tunniste (Universal Resource Identifier), joilla voidaan yksikäsitteisesti viitata mihin tahansa maailman verkon *resurssiin* (resource). WWW on URI-tunnisteiden yksilöimien resurssien verkko. URI-järjestelmään kuuluvat keskeisenä osana perinteiset WWW-linkeissä käytettävät URL-tunnisteet (Universal Resource Locator) (esimerkiksi <http://www.cs.helsinki.fi>), joilla voidaan viitata verkon tiedostoihin annetulla protokollalla, kuten HTTP tai FTP. Ted Nelsonin, hypertextin isän, idean mukaisesti resurssi voidaan ottaa käyttöön viittamalla siihen tarpeen mukaan, jolloin resursseja ei tarvitse kopioida eikä hallita keskusjohtoisesti. Hallinta säilyy resurssin alkuperäisellä toimittajalla ja ajantasaisen tiedon hajautettu ylläpito ja hyödyntäminen verkon yli käy mahdolliseksi.

XML-kielet

XML (eXtensible Markup Language) (Bradley, 2000; Deitel et al., 2001) on WWW:n uusi perusta. Siinä käytettävä Unicode-merkistö kattaa kaikki maailman kielet. XML ei ole HTML:n kaltainen kieli vaan metakieli, jonka avulla kukin voi määrittellä itselleen sovelluskohtaisia merkkäuskieliä tietojen esittämiseksi. W3C:n toimesta on esimerkiksi otettu käyttöön HTML:stä uusi XML-perustainen XHTML-määrittely, jonka etuna on mm. kielen muotojen laajentamismahdollisuus.

Esimerkiksi osoitetiedon esittämiseen voitaisiin määrittellä suomenkielinen XML-syntaksi, ja esittää osoitteita seuraavaan tapaan:

```
<OSOITE>
  <NIMI>Onni Ohjelmoija</NIMI>
  <PUHELIN> 123 456 </PUHELIN>
</OSOITE>
```

XSL-transformaatioiden (XML Stylesheet Language) avulla osoitetiedoista voitaisiin automaattisesti tuottaa selaimella luettava normaali HTML-sivu, painettu osoitekirja tms. ulkoinen esitysmuoto. Ideana on dokumentin rakenteen ja ilmiäsuun erottaminen toisistaan.

XML-merkkäukseen ei sisälly mitään semantiikkaa koneen kannalta, vaan merkityksen kielen eri ilmauksille kuten "OSOITE" antaa ihminen. Koneen kannalta yllä oleva osoite tieto on täsmälleen samanarvoinen kuin sen alla oleva kreikkalaisilla merkeillä uudelleen kirjoitettu versio. XML kuvaa vain tieton rakenteen, syntaksin.

```

vXQXTwE {
  vφTφT {Xφφx XθωΔx YωxωYvHφTφT {
  vχAcEvTφ { θθI iKk vHχAcEvTφ {
vHXQXTwE {

```

Filosofian sanakirjan (Angeles, 1981) mukaan "semantiikka" (kreikaksi *semainein*, to signify, antaa merkitys) on oppi, joka tutkii symbolien viittaamista toisiin olioihin. Tässä yhteydessä voidaan olla kiinnostuneita erityisesti symbolien merkityksestä (intensio) tai viittauksen kohteista (ekstensio). Vaikka XML mielletään merkitysten esittämisen kielenä, se ei ole "semanttinen" edellä kuvatussa mielessä. XML-kielen symbolit, esimerkiksi OSOITE yllä, eivät viittaa mihinkään vaan viittaus muodostuu ihmislukijan mielikuvituksessa (kreikkalaisia aakkosia käytettäessä viittausta ei muodostu).

XML:n avulla voidaan kehittää eri toimialoille yhteisiä syntaktisia (kreikaksi *syntaxis*, järjestys, rakenne) kieliä tietojen esittämiseen. Parhaillaan kokoontuu suuri määrä erilaisia kansainvälisiä työryhmiä suunnittelemassa omille aloilleen XML-standardeja (ks. esimerkiksi <http://www.xml.org>). Yhteinen syntaksi on edellytys järjestelmien yhteiskäyttöisyydelle ja kommunikaatiolle. Yhteisen kielen avulla myös vähenee eri esitysmuotojen välillä tarvittavien muunnosten määrä. Jos halutaan määritellä muunnokset esimerkiksi n:stä eri osoiteformaateista m:ään toiseen muotoon, tarvitaan n^*m muunnosta. Jos sensijaan käytetään yhteistä kieltä, tarvitaan vain $n+m$ muunnosta: ensin n muunnosta välikieleen ja sitten m muunnosta siitä kohdekieliin.

Näköpiirissä on XML-muotoisen tiedon nopea lisääntyminen WWW:ssä. Yhdeksi käytännölliseksi ongelmaksi muodostuu tällöin laajojen XML-dokumenttivarastojen tiedonhallinta tietokantatyyppeillä ratkaisulla (Halevy, 2001).

4. Semanttiset metakuvaukset

Jotta kielellä olisi koneellisesti tulkittavissa oleva merkitys, täytyy sen symbolien ja rakenteiden viitata johonkin alla olevaan malliin. Merkitys on olemassa vain suhteessa johonkin allaolevaan. Esimerkiksi logiikassa merkitys muodostuu malliteorian joukko-opillisten rakenteiden kautta, jotka kuvaavat mahdollisia maailmantiloja. Semantic Web –konseptin keskeinen oivallus kielelliseltä kannalta on tämän käyttökelpoiseksi osoittautuneen idean tuominen WWW-maailmaan.

RDF, Resource Description Framework

Tällä hetkellä ehkä tunnetuin "semanttinen" WWW-kieli on W3C:n RDF (Resource Description Framework) (Lassila, Swick, 1999; Hjelm, 2001) ja siihen liittyvä laajennus RDF Schema (RDFS) (Brickley, Guha, 2000). RDF:llä kuvataan URI-viittausten avulla osoitettujen verkkosisäلتöjen, resurssien (resource), ja symbolien merkityksiä. RDF-kuvauksella voidaan esimerkiksi kertoa, että "<http://www.cs.helsinki.fi>" on Helsingin yliopiston tietojenkäsittelytieteen laitoksen WWW-palvelin. Kielen "semanttisuus" syntyy viittauksista allaoleviin resursseihin.

RDF(S):ää voidaan tarkastella kolmella eri tasolla. Syntaksin tasolla RDF-kuvaukset esitetään (yleensä) XML-dokumentteina. Rakenteellisesti kuvaukset koostuvat kolmikoista. Semanttisesti kuvaukset ovat suunnattuja verkkoja, joille on määritelty osittainen semantiikka.

RDF on ytimeltään yksinkertainen relaatiomalli. Sillä esitetty tieto koostuu joukosta (*objekti, ominaisuus, arvo*) -kolmikoita, joissa kukin jäsen voi olla joko vakio symboli (literal) tai verkon resurssi (URI). Esimerkiksi kolmikot

(<http://www.cs.helsinki.fi/>, onOsa, <http://www.helsinki.fi/index.html>)

(<http://www.cs.helsinki.fi/index.html>, sijainti, <http://www.hel.fi>)

voisivat kertoa, että tietojenkäsittelytieteen laitos on osa Helsingin yliopistoa, joka sijaitsee Helsingissä.

RDF:n syntaksin spesifikaatio perustuu XML-määrittelyyn. Oleellista on kuitenkin huomata, että RDF:n tietomallina ei ole XML:n tavoin ole sarjallistettu (serialize) syntaktinen puu, jonka haarat pitää esittää XML-skeeman mukaisessa järjestyksessä. Tietomalli on esitysmuodosta erillinen kolmikkojoukko, jonka voi ajatella graafisesti muodostavan suunnatun verkon. RDF-kuvauksen esittämiselle voidaan siksi kehittää muunkinlaisia kuin W3C:n suosituksen mukaisia sarjallistettuja XML-muotoja. Eräs tällainen on suoraan kolmikoihin perustuva helpopolukuinen N3 (Notation3). Jo W3C:n syntaksisuosituksen käyttö mahdollistaa saman RDF-mallin kirjoittamisen useassa eri muodossa.

RDF:ssä viitattavina URI-resursseina voivat olla myös toiset RDF-ilmaukset. Tämä merkitsee sitä, että

RDF-verkko voi olla hierarkkinen. Voidaan esimerkiksi ilmaista luottamusta toisen tahon RDF-kuvaukseen. Tällaisesta korkeammanasteisesta viittauksesta käytetään nimitystä "reification".

RDF Schema

RDF ei tarjoa välineitä kuvauksissa käytettävän sanaston määrittelemiseksi (esimerkiksi ominaisuudet "onOsa" ja "sijainti" yllä olevassa esimerkissä). RDF Schema (RDFS) (Brickley, Guha, 2000) on kieli, jonka avulla voidaan:

- Kuvata RDF-sovelluksessa käytävät termit.
- Asettaa tyyppirajoituksia (constraint) olioille ja arvoille eri ominaisuuksiin liittyen.

Termien kuvaamisen RDFS tarjoaa tähän joukon primitiivejä. Class, subclass ja type -ilmausten avulla voidaan määritellä käsittehierarkioita hieman samaan tyyliin kuin olio-ohjelmoinnissa. domain ja range -ilmaukset puolestaan kertovat millaiselle oliolle (domain) tietyn ominaisuuden arvo (range) voidaan antaa.

RDFS-kuvauksissa käytetään normaalia RDF-kieltä. Etukäteen sovittujen Class, subclass, type ym. primitiivien avulla vain annetaan sovelluksen RDF-ilmauksille korkeammanasteinen semanttinen tulkinta: kerrotaan miten käytetyt käsitteet liittyvät hierarkkisesti toisiinsa ja millaisia rajoituksia ominaisuuksien esittämiseksi on annettu. Primitiivejä on vähän eikä niihin liity minkään tietyn sovellusalan semantiikkaa, joten RDFS on käyttökelpoinen eri sovelluksiin samassa mielessä kuin esimerkiksi olioajattelu ohjelmoinnissa.

RDF(S) tarjoaa perustan, jolle sovellusten kehittäjät voivat toteuttaa varsinaiset sovelluksensa. Älykkyys syntyy vasta sovellustasolla, jossa kuvaukset luetaan RDF-jäsentimellä (parser) koneen keskusmuistiin ja jossa niille sovelluksen toimesta annetaan operationaalinen tulkinta. XML:n nimiavaruusmekanismin (name space) avulla kuvauksissa voidaan hyödyntää eri standardointiryhmien laatimia sanastoja. Esimerkiksi Dublin Core (<http://www.dublincore.org>) on kirjastoalalta lähtöisin oleva standardi, jossa viidentoista ominaisuuden avulla kuvataan dokumentteihin liittyvä yleinen metatieto, kuten dokumentin laatija, laatimisaika jne.

RDF(S):n XML-syntaksi on ihmisen kannalta kömpelö tietämyksen esittämiskieli. Semantic Web -visio ei voi toteutua, jos esimerkiksi WWW-sivujen tekijöiden pitäisi koodata sillä merkityksiä. Metakuvauksia voidaan kuitenkin tuottaa myös ohjelmallisesti. Esimerkiksi W3C:n kehittämän RDFPic Java-appletin avulla voidaan JPEG-kuvien tiedostoihin upottaa tekstimuotoista metadataa RDF-muodossa. Adobe'n PDF-tulostimen uusi versio käyttää dokumenttien metakuvauksissa RDF-standardia ilman, että käyttäjä sitä edes huomaa (Dublin Core -standardin mukaiset dokumentin perustiedot). Tätä kirjoitettaessa yleiskäyttöiset hakukoneet (Google, Altavista, Lycos jne.) eivät osaa hyödyntää RDF(S) metakuvauksia.

RDF(S) mahdollistaa myös verkon eri paikassa esitettyjen metakuvausten yhdistämistä eli tietojen yhteiskäyttöä. Samaan resurssiin (URI) liittyvät erilliset kuvaukset, esimerkiksi musiikkikappaleeseen liittyvät tiedot, voidaan koota yhteen RDF-verkkoon. RDF(S) tarjoaa myös abstraktin tason, jonka kautta erilaisten XML-kielien muuntaminen yhteismitalliseksi näyttäisi olevan helpompaa kuin kirjoittamalla muunnoksia suoraan XML-tasolla (Omelayenko, Fensel, 2001).

WWW:n RDF(S)-dokumenttien määrä tulee jatkossa lisääntymään. RDF-muotoisen tiedon hallitsemiseksi on sille ryhdytty kehittämään tietokantaratkaisuja (esimerkiksi Redland, <http://www.redland.opensource.ac.uk/>) vastaavaan tapaan kuin XML-kielillekin on olemassa tietokantoja ja kyselykieliä (esimerkiksi XQuery). Myös RDFS:lle on toteutettu ensimmäinen tietokantaratkaisu ja RQL-kyselykieli, Sesame (<http://sesame.aidadministrator.nl/>).

Aihekartat - Topic Maps

ISO:n (International Organization for Standardization) vuonna 2000 standardoima Topic Maps (TM) (ISO/IEC 13250) on RDF(S):ää muistuttava kieli verkon resurssien kuvaamiseen (Pepper, 2000a). Molemmilla lähestymistavoilla on sama tavoite: infoähkyn hoitaminen WWW:n resurssien metakuvauksilla. Kielet myös perustuvat syntaksiltaan XML:ään ja ovat yksinkertaisuudestaan huolimatta ilmaisuvoimaisia mm. reifikaation takia.

TM:n idea syntyi jo ennen WWW:n yleistymistä 90-luvun alkupuolella kirjojen hakemistoihin ja sanastoihin liittyvän standardointityön yhteydessä. Kirjan hakemiston ideana on indeksoida

dokumentissa esiintyvät aiheet, jolloin lukija voi löytää teoksessa käsiteltävät asiat hakusanojen perusteella. TM-konsepti laajentaa hakemiston ideaa semanttisessa mielessä ja tarjoaa XML-perustaisen kielen XTM (XML Topic Maps) (<http://www.topicmaps.net>) aihekarttojen kuvaamiseksi.

TM-aihekartan ydin koostuu aiheista (topic), assosiaatioista (association) ja ilmentymistä (occurrence). Aihe voi olla periaatteessa mikä tahansa asia, käsite, henkilö tms., johon voidaan liittää tietoa. Esimerkiksi Väinämöinen, Sampo tai Tuonelan joutsen. Aiheita vastaa perinteisen kirjan hakemiston hakusanat.

Aiheet voivat kuulua yhteen tai useampaa aihetyyppiin (topic type). Esimerkiksi Sampo voi olla kalevalainen taikakalu tai vakuutusyhtiö. Tulkinta määräytyy aihekartan avulla ilmaistavissa olevan näkökulman (scope) avulla. Aihetta karakterisoi sille annetut nimet (aihe tosin voi olla myös nimetön), ilmentymät ja assosiaatiot. Ilmentymiä ovat aiheen erilaiset inkarnaatiot: esimerkiksi Tuonelan joutsen -aihe voi konkretisoidua maalauksena, sävellyksenä tai aihetta käsittelevänä artikkelina (roolit). Nimetyt assosiaatiot liittävät aiheita toisiinsa. Esimerkiksi "säveltäjä" -assosiaatio voi liittää Jean Sibeliuksen Tuonelan joutseneen. Assosiaatiot voivat olla myös omia aiheitaan.

Aihekartat ovat fyysisesti erillisiä metakuvauksia niiden kuvaamista dokumenteista. Karttoja voidaan kehittää kohdedokumentteja muokkaamatta ja eri karttojen yhdistely laajemmaksi kokonaisuudeksi on mahdollista. Esimerkkisovelluksena voisi olla "Semantic Web Kalevala", joka liittää toisiinsa kansalliseepoksemme teemat, henkilöt, näiden innoittamat sävellykset, maalaukset, tutkimukset, teoksen historian jne. Aihepiiriin liittyvä tiedonhaku tapahtuisi aiheesta ja sen ilmentymistä toiseen siirtymällä assosiaatioita seuraamalla.

W3C:n RDF- ja ISO:n Topic Maps -standardien välinen suhde on herättänyt kiivastakin keskustelua (Pepper, 2000b). TM on käsitteellisesti korkeamman tasoisen ja sovellusläheisempi sisältöjen semanttisten rakenteiden kuvaamisen formalismi, mikä näkyy mm. käsitteiden "aihe", "assosiaatio" ja "ilmentymä" käytöstä. RDF on jo nimensäkin perusteella geneerinen matalan tason "kehys" (frame work) ja verkkotietomalli, jonka päälle sovellusläheisempiä esityskieliä voidaan laatia. Ei ole selvää miten RDF:llä voisi helposti esittää TM:n näkökulmia (scope). Toisaalta TM:sta puuttuu RDFS:n kaltainen ontologioiden määrittelymekanismi. Keskustelu eri lähestymistapojen eduista, haitoista ja mahdollisesta yhdistämisestä tulee jatkumaan.

5. Ontologiat

Ontologian käsite (Sowa, 2000) on Semantic Web –konseptin keskeisimpiä (Fensel, 2001a). Ontologia voidaan määrittellä (Gruber, 1993) seuraavasti:

Ontologia on formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä
(An ontology is a formal, explicit specification of a shared conceptualization)

"Formaalisuus" ja "eksplisiittisyys" mahdollistavat käsitteistön koneellisen tulkinnan, "yhteisyys" tietämyksen jakamisen, yhdistämisen ja yhteiskäytön, mikä erityisen olennaista WWW-ympäristössä. Käytännössä ontologiat ovat eri sovellusalojen terminologisia käsittehierarkioita, joissa määritellään alalla käytettävät termit ja käsitteet ja näiden välisiä suhteita.

Ontologioiden käyttötarkoitukset, lähestymistavat ja tiedoesitysmekanismit poikkeavat huomattavasti toisistaan (Noy, Hafner, 1997; Fensel, 2001a). Ontologian avulla voidaan esittää jonkin erityisalan ammattikäsitteitä ja -tietämystä (esim. biologia, elektroniikka tms.), metadataa (esimerkiksi tietolähteiden tai kuvien julkaisutiedot), yleistä arkitietämystä, käsitteistöjä (esimerkiksi ontologioiden kuvaamisessa käytettävät metakäsitteet), tehtäviä, prosesseja ja palveluita. Tunnettuja laajoja ontologioita ovat mm. suomenkielen Yleistä asiasanastoa (<http://vesa.lib.helsinki.fi/>) muistuttava WordNet (<http://www.cogsci.princeton.edu/~wn/>), joka sisältää yli 100.000 englannin kielen käsitettä, IT- ja elektroniikkateollisuuden RosettaNet (<http://www.rosettanet.org>), CYC (<http://www.cyc.com/>) ja vireillä oleva IEEE:n SUO (Standard Upper Ontology) standardointihanke ontologisista yläkäsitteistä sopimiseksi (<http://suo.ieee.org/>).

Ontologian luomisen keskeiset välineet ovat: (1) Ontologiakieli, jolla käsitteet ja niiden väliset suhteet määritellään. (2) Ontologiaeditori, joilla ontologiset kuvaukset käytännössä laaditaan. Ontologiakielten osalta suurimman huomion kohteeksi ovat viime aikoina joutuneet eurooppalainen OIL (Ontology Inference Layer) (<http://www.ontoknowledge.org/oil/>) ja amerikkalainen DAML (DARPA Agent Markup

Language) (<http://www.daml.org>). Näiden pohjalta on tekeillä yhdistelmä DAML+OIL. Tavoitteena on saada lopulta aikaan yhteinen W3C:n standardisuositus. DAML+OIL on ihmisläheinen loogisten käsittehierarkioiden määrittelykieli. Sen puitteissa terminologialogiikoille keskeinen käsitteeseen sisältyvyyden (subsumtion) ongelma on ratkeava (decidable) ja kielelle on voitu toteuttaa tehokas FaCT-niminen päättelykone.

DAML+OIL:n innovaationa on yhdistää tietämystekniikan piirissä kehitetyt ns. terminologialogiikat (description logic) (<http://dl.kr.org>) kehys- ja olioajatteluun (frame, object) ja WWW-ohjelmoinnin välineistöön, erityisesti XML:ään ja RDF(S):ään. Nämä tarjoavat perustan ontologioiden esittämiseen ja jaettuun käyttöön Internetissä. DAML+OIL:n kaltaisen korkean tason loogisen ontologiakielen avulla voidaan laatia WWW:n resursseista metakuvauksia ja jättää niiden muuttaminen alemman taso RDF(S)- ja XML-ilmauksiksi koneen murheeksi.

Ontologioiden laatimisen avuksi on kehitetty lukuisia editoreita, kuten Protégé (<http://protege.semanticweb.org/>) ja OilEd (<http://img.cs.man.ac.uk/oil/>). Nämä tarjoavat helppokäyttöisen graafisen käyttöliittymän ohella mm. apuvälineitä ontologian sisäisen konsenssin tarkistamiseen ja eri ontologioiden yhdistämiseen. Protégé-editoriin on mahdollista liittää back-end-generaattoreita, joilla voi tuottaa yleisestä ontologiakuvauksesta toteutuksia eri kielille, esimerkiksi generoida RDFS-dokumentteja.

Ontologioiden kehittämiseen liittyy monia käytännön vaikeuksia. Terminologian standardointi on ylipäänsä erittäin hankalaa johtuen eri tahojen ja intressien erilaisista tarpeista ja mieltymyksistä. Ontologioilla on taipumusta muodostua laajoiksi. Lisäksi ontologiat muuttuvat ajan kuluessa ja niiden hallinta voi olla vaikeaa. Esimerkiksi Tšekkoslovakia oli vielä vähän aikaa sitten mukana valtioiden ontologiassa ja termin avulla on indeksoitu paljon tietoja. Nykyisen valtiotaksonomian kannalta Tšekkoslovakiaa ei kuitenkaan enää ole. Ontologioiden kehittämistekniikat (Noy, McGuinness, 2000), ontologioiden yhteiskäyttö ja yhdistäminen sekä ontologioiden hallinta muodostavat suuren haasteen Semantic Web -vision toteutumiselle (Stuckenschmidt, 2001).

6. Logiikka ja luottamus

Ontologioiden avulla voidaan määrittellä ne käsitteet ja oliot, joita sovellusalaan liittyy. Käsitteet ja niiden ilmentymät eivät kuitenkaan vielä kerro, miten niitä voitaisiin käyttää hyväksi. Esimerkiksi sähköisen ostoagentin pitäisi voida päätellä, että jos tavoitteena on kymmenen PC:n hankkiminen ja se on tilannut yrityksestä A neljä konetta ja yrityksestä B kuusi, voidaan ilmoittaa ostomääräyksen antajalle, että tehtävä on suoritettu. Tällainen tietämys ei ole luonteeltaan ontologista, vaan toimintaa ja loogista päättelyä ontologian puitteissa ilmaistujen olioiden välisistä suhteista ja prosesseista.

W3C:n piirissä ollaankin parhallaan tutkimassa ja valmistelemassa loogisen tason standardeja, jotka sijoittuvat käsitteellisesti ontologiatason yläpuolelle. Työn alla on mm. päättelysääntöjen merkkäuskieli RuleML (Rule Markup Language).

WWW:n voima perustuu paljolti siihen, että kuka tahansa voi julkaista verkossa avoimesti tietoa ja mielipiteitä ja vastaavasti päästä toisten tietoihin käsiksi. Tiedon salauksella voidaan taata, etteivät väärät tahot pääse käsiksi luottamuksellisiin tietoihin. Toinen luottamukseen liittyvä kysymys on, mihin verkosta löytyviin sinänsä julkisiin tietoihin voi luottaa. Esimerkiksi tunnetun henkilön, lehden tai yrityksen antamiin metakuvauksiin, annotointeihin, luotamme mielummin kuin satunnaiselta tuntemattomalta sivulta löydettyihin.

WWW:n resurssien annotoinnin perusmekanismeja kehitetään mm. W3C:n Annotea-projektissa (<http://www.w3.org/2001/Annotea/>). Ideana on tarjota WWW:n käyttäjille välineet sivustojen arvosteluun siten, että toiset käyttäjät voisivat paremmin arvioida verkon resurssien luotettavuutta. Digitaalisten allekirjoitusten (digital signature) avulla voidaan vakuuttaa kuvausten toimittaneen tahon identiteetistä ja näin arvioida annotointien todenperäisyyttä ja arvoa.

7. Sovellusalueita

Metakuvausten ja ontologiatekniikoiden tärkeitä sovellusalueita ovat mm.:

- Informaation haku (information search/retrieval).
- Tietämyksen hallinta (knowledge management).
- Verkkokauppa (web commerce) B2C sektorilla.
- Sähköinen liiketoiminta (electronic business) B2B sektorilla.

WWW:n keskeisimpiä ongelmia on käyttäjän tarvetta vastaavan tiedon tai palvelun löytäminen

Internetin valtavasta, huonosti jäsenetystä tietomassasta. Assosiativinen hypertekstin käyttö (Agosti, Smeaton, 1996) ja sisältöjen metakuvaukset tarjoavat uusia, sisältöperustaisia mahdollisuuksia tiedonhakuun ja laajentavat näin perinteisten hakusanoihin perustuvien menetelmien (Korfage, 1997) mahdollisuuksia. Tuloksena syntyy esimerkiksi erilaisia semanttisia portaaleja, kuten Karlsruhen yliopiston AIFB-instituutin intranet (Maedhe et al., 2001).

Tietämyksen hallinnan (Smith, Farquhar, 2000) merkitys liittyy yritysten ja muiden organisaatioiden alati lisääntyvään tarpeeseen hankkia, ylläpitää, löytää ja hyödyntää omaa tietämystään kilpailuedun saamiseksi ja toimintojen tehostamiseksi. Haasteena ovat mm. organisaatioiden tietovarastoissa olevien dokumenttien vapaamuotoisuus ja hajautus maapalloistumisen myötä. Semantic Web -teknologioilla saavutettavia etuja ovat esimerkiksi: tietoa voidaan hakea niiden sisällön (eikä vain hakusanojen) perusteella; pelkän tiedon haun sijaan voidaan tuottaa vastauksia kysymyksiin; dokumentteja voidaan vaihtaa (esimerkiksi XSLT-muunnosten avulla) eri standardeja käyttävien systeemien välillä; dokumentteihin voidaan luoda erilaisia näkökulmia. Standardien avulla voidaan saada alakohtaiset tietojärjestelmät (esimerkiksi yritys- ja tuotekatalogit, museoiden kokoelmätietokannat, eri sairaaloiden potilastietojärjestelmät jne.) yhteismitallisiksi ja voidaan luoda yhteisiä kieliä systeemien väliseen viestintään.

Verkkokaupan innovaatioita ovat mm. on-line markkinapaikat, ostoagentit ja huutokaupat (Turban et al., 2000, Deitel et al, 2001b). Alan lupauksena on uusi sähköinen jakelu- ja markkinointikanava sekä tämän mahdollistamat uudentyyppiset liiketoimintamallit. Näiden sisältöjen, palveluiden ja prosessien kuvauksessa Semantic Web –teknologioilla tulee olemaan yhä tärkeämpi rooli.

Sähköisessä liiketoiminnassa (Turban et al., 2000, Deitel et al, 2001b) keskeinen kehityskohde on mm. liiketoimintaan liittyvien transaktioiden hallinta. Toinen tärkeä kehityskohde on tuote- ja palvelukuvaukset ja luettelot sekä näihin liittyvät hakemistopalvelut. XML-perustaisten teknologioiden ja ontologioiden avulla voidaan eri liiketoiminta-alueille luoda sisällöllisiä standardeja, jotka integoituvat luontevasti yritysten normaaliin dokumenttien hallintaan ja Internet-viestintään. Yhteisten terminologia- ja kommunikointikielten kautta käy mahdolliseksi esimerkiksi luoda eri toimialoille yhteisiä tuoteportaaleja.

Voimakkaan kansainvälisen standardointityön kohteena ovat mm.:

- EDI:n (Electronic Data Interchange) XML-versio XML-EDI (<http://www.oasis-open.org/cover/xml.html#xml-edi>). EDIFACT ei ole vastannut siihen alunperin asetettuja toiveita, vaan on osoittautunut kömpelöksi ja eristyneeksi ratkaisuksi. Apua haetaan XML-maailmasta.
- Sähköisen liiketoiminnan kehysstandardit. Esimerkiksi ebXML (Electronic Business XML) (www.ebxml.org) on laaja modulaarinen standardointihanke, jonka kunnianhimoisena tavoitteena on mahdollistaa globaalien sähköisten markkinapaikkojen luominen kaikenkokoisten ja -laisten yritysten globaalia kaupankäyntiä varten. Hankkeen vetureina ovat OASIS ja UN/CEFACT. UDDI (Universal Description, Discovery and Integration) on yritysvetoinen, Ariban, IBM:n ja Microsoftin standardointihanke. Sen spesifikaatioilla yritykset voivat julkaista tarjoamiaan palveluita erityisen rekisteröintikeskuksen kautta ja vastaavasti löytää ja käyttää itse tarvitsemaansa palveluita. WSDL (Web Services Description Language) on kieli, jolla voidaan kuvata verkon eri päätepisteiden (port) palveluita (esimerkiksi annetun yhtiön viimeisimmän pörssikurssin lukeminen tietystä pörssipalvelusta). (http://www.xml.org/xml/resources_cover.shtml)
- Rakenteinen ja tietämysperustainen kommunikointi. SOAP (Simple Object Access Protocol) (http://www.xml.org/xml/resources_focus_soap.shtml) on HTTP-protokollaan perustuva standardi, jonka avulla järjestelmät voivat vaihtaa XML-muotoisia viestejä. KQML (Knowledge Query and Manipulation Language) (<http://www.cs.umbc.edu/kqml/>) on kehys älykkäiden agenttien kommunikoinnille. Se perustuu vastaavanlaisiin pragmaattisiin rakenteisiin (puheaktit, speech acts) kuin ihmisten välinen viestintä. FIPA (Foundation for Intelligent Physical Agents) (<http://www.fipa.org>) ja OMG (Object Management Group) (<http://www.omg.org>) kehittävät tahoillaan kieliä agenttien kommunikointiin.
- Toimialakohtaiset standardit. Edellä kuvatuilla standardeilla voidaan kuvata sähköiselle kaupankäynnille ja viestinnälle lähinnä yleiset kehukset ja toimintamekanismit. Niissä ei kuitenkaan otata kantaa siihen, millaisista tuotteista tai palveluista on kysymys tai millaisia sisältöjä viesteissa kulkee. Alakohtaisten ontologioiden avulla järjestelmiin voidaan antaa ymmärrys liiketoiminnan ja viestinnän kohteena olevista tuotteista ja palveluista. Ilman yhteistä toimialakohtaista terminologiaa ja kieltä eivät järjestelmät voi toisiaan syvällisemmin ymmärtää.

Tunnetuin esimerkki toimialakohtaisesta standardoityyöstä lienee elektroniikkateollisuuden piiristä syntynyt RosettaNet (<http://www.rosettanet.org>).

8. Semantic Web ja Suomi

Tulevaisuuden mobiili- ja muiden verkkojen rakentamisen perimmäisenä tavoitteena on palveluiden tarjonta, joiden kehittämiseksi Semantic Web näyttäisi tarjoavan uuden lähtökohdan. Aihepiiri onkin ollut viime aikoina sekä vilkkaan tutkimuksen että käytännön standardoityyön kohteena, sillä yhteiset pelisäännöt ovat edellytys teknologioiden luonnostaan globaalille käytölle WWW-ympäristössä.

Internetin kehitystä koordinoiva W3C-järjestö käynnisti helmikuussa 2001 erityisen Semantic Web Activity -ohjelman (<http://www.w3c.org/2001/sw>) edesauttamaan ja yhtenäistämään alan kehitystä. Tähän päänavaukseen liittyen ilmestyi mm. Scientific American -lehdessä keväällä näytävä cover story -artikkeli Semantic Webistä kirjoittajana WWW:n isänä tunnettu Tim Berners-Lee ja kumppanit (2001). Euroopassa käynnistyi kesällä 2001 erityinen OntoWeb -tutkimusverkosto (<http://www.ontoweb.org>) alan tutkimuslaitosten, yritysten ja tutkijoiden yhteistyötä lujittamaan. USA:ssa puolustusministeriön DARPA rahoittaa laajaa DAML-ohjelmaa (DARPA Agent Markup Language) (<http://www.daml.org>). Maailmalle on lyhyessä ajassa syntynyt suuri määrä erilaisia yhteenliittymiä Semantic Web -konseptin eri kielten, teknologioiden ja standardien luomiseksi (<http://www.semanticweb.org>). Vaarana jopa on, että runsas standardoiminen kääntyy itseään vastaan: jo kahden päällekkäisen "standardin" olemassaolo merkitsee, ettei olemassa on ensimmäistäkään standardia.

Ajatus kotimaisen Semantic Web Kick-Off in Finland -tilaisuuden järjestämisestä syntyi huhtikuussa 2001 kansainvälisen kehityksen innoittamana. Tilaisuuden tavoitteena on herättää Suomessa laajempaa kiinnostusta aihepiiriä kohtaan niin yliopistojen, tutkimuslaitosten, yritysten kuin julkishallinnonkin piirissä. Yhteistyöverkoston luominen heti alussa on tärkeää, sillä Semantic Web alueella yhdistyvät hedelmällisellä tavalla:

- **Teollinen intressi.** Yritysmaailmalla on suuri tarve tuottaa mielekkäitä palveluja ja liiketoimintaa tulevaisuuden Internet- ja mobiiliverkkoihin.
- **Tekninen mahdollisuus.** WWW- ja Internet-alusta sekä XML-perustaiset standardit, tekniikat ja työkalut mahdollistavat vision toteuttamisen globaalisti ja käytännöllisinä sovelluksina.
- **Tieteellinen haaste.** Ala tarjoaa mahdollisuuden hyödyntää älykkäiden järjestelmien ja tietämystekniikan menetelmien tuloksia koko tietojenkäsittelyalan kannalta keskeisessä uudessa kontekstissa, Internetissä.
- **Kansallinen intressi.** Älykäs WWW edellyttää mm. suomenkielen kieliteknologian kehittämistä ja soveltamista.

Ala on uuden WWW-tutkimuksen ytimessä, kun verkkoa tarkestellaan sen sovelluskäytön ja palveluiden näkökulmasta. Meneillään olevan teknologisen murroksen tuloksena tarjoutuu laadullisesti uudenlainen mahdollisuus kehittää tietosisältöihin perustuvia, aiempaa olennaisesti älykkäämpiä järjestelmiä käytettäväksi Internet- ja intranet-ympäristöissä.

WWW:n infoähky ja hyödyllisten palveluiden suuri kysyntä on nostattanut oireenaan Semantic Web -kuumeen. Lääketieteessä kuumeen tarkoituksena on taudin haasteeseen vastaaminen: kuume merkitsee tervettä puolustusmekanismien toimimista. Syytä on tietysti muistaa, että liian korkeana kuume voi johtaa hallusinaatioihin, kuten IT-alalla niin usein on nähty.

Nähtäväksi jää, miten nopeasti visio Semantic Webistä, merkitysten Internetistä alkaa toteutua käytännön sovelluksina. Lähtöasetelma alalla on haasteiden ja visioiden suuruudesta huolimatta lupaava ja alan kehittäjien asenne nöyrä. Selvää parannusta nykytilaan on luvassa asteittain semanttisia kuvauksia tietoverkkoihin lisäten.

Kirjallisuutta

- P. Angeles, Dictionary of Philosophy. Harper Reference, New York, 1981.
- M. Agosti, A. Smeaton (eds.) Information retrieval and hypertext. Kluwer, New York, 1996.
- K. Ahmed, D. Ayers, et al., Professional XML Meta Data. Wrox Press, 2001.
- M. Agosti, A. Smeaton (eds.) Information retrieval and hypertext. Kluwer Academic Publishers, 1996.
- T. Berners-Lee, M. Fischetti, M. Dertouzos, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. Harper Business, 2000.
- T. Berners-Lee, J. Hendler, O. Lassila, The semantic web. Scientific American, May, 2001.

- D. Brickley, R. Guha, RDF Schema specification 1.0. <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>, 2000.
- N. Bradley, *The XML Companion*. Addison-Wesley, 2000.
- J. Broekstra, A. Kampman, F. van Harmelen, *Sesame: An architecture for storing and querying RDF data and scheme information*. Free University of Amsterdam, 2001.
- R. Cooley, M. Bamshad, S. Jaideep, *Web Mining: Information and Pattern Discovery on the World Wide Web*", <http://www-users.cs.umn.edu/~mobasher/webminer/survey/survey.html>, 1997.
- H. Deitel, P. Deitel, H. Nieto, *XML. How to program*. Prentice Hall, New Jersey, 2001a.
- H. Deitel, P. Deitel, H. Nieto, *e-Business & e-Commerce. How to program*. Prentice Hall, New Jersey, 2001b.
- M. Erdmann, R. Struder, *Ontologies as conceptual models for XML documents*. Article, University of Karlsruhe, 2000.
- D. Fensel (ed.), *The semantic web and its languages*. IEEE Intelligence Systems, Nov/Dec 2000.
- D. Fensel, *Ontologies: Silver bullet for knowledge management and electronic commerce*. Springer-Verlag, 2001.
- D. Fensel, J. Hendler, H. Lieberman and W. Wahlster (eds.), *Semantics for the WWW*. The MIT Press, 2001 (forth-coming).
- T. R. Gruber, *A translation approach to portable ontology specifications*. Knowledge Acquisition, Vol. 5, 199-220, 1993.
- J. Han, M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.
- A. Halevy. *IEEE Data Engineering*, special issue on XML data management, June, 2001.
- F. van Harmelen, I. Horrocks. *FAQs on OIL: The ontology inference layer*. IEEE Intelligence Systems, Nov/Dec 2000.
- J. Hjelm, *Creating the Semantic Web with RDF*. John Wiley & Sons, New York, 2001.
- E. Hyvönen, I. Karanta, M. Syrjänen (toim.) *Tekoälyn ensyklopedia*. Gaudeamus, 1993.
- R. Korfhage: *Information storage and retrieval*. John Wiley & Sons, New York, 1997.
- O. Lassila, R. Swick, *Resource Description Framework (RDF): Model and syntax specification*. W3C Recommendation, syntax, 1999.
- A. Levy, D. Weld (eds.): *Intelligent Internet systems*. Artificial Intelligence, Vol 118, 2000.
- S. Luke, L. Spector, D. Rager, J. Hendler: *Ontology-based web agents*. Proceedings of First International Conference on Autonomous Agents, 1997.
- A. Maedche, S. Staab, N. Stojanovic, R. Struder, Y. Sure, *Semantic portal – the SEAL approach*. Report, Institute AIFB, Univ. of Karlsruhe, Germany, 2001.
- N. Noy, D. McGuinness, *Ontology Development 101: A guide to creating your first ontology*. Stanford University, Medical Informatics, 2000.
- N. Noy, C. Hafner: *The state of the art in ontology design*. AI Magazine, No. 3, 1997.
- B. Omelayenko, D. Fensel, *A Two-Layered Integration Approach for Product Information in B2B E-commerce*, In: K. Bauknecht, S.-K. Madria, G. Pernul (eds.), *Electronic Commerce and Web Technologies*, Proceedings of the Second International Conference on Electronic Commerce and Web Technologies (EC WEB-2001), LNCS 2115, 226-239, Springer-Verlag, 2001
- R. Smith, A. Farquhar, *The road ahead for knowledge management. An AI perspective*. AI Magazine, Winter 2000, 17-40.
- H. Stuckenschmidt (ed.), *Ontologies and information sharing*. Working notes, IJCAI-2001 workshop, AAAI, 2001.
- T. Paziienza (ed.) *Information extraction: a multidisciplinary approach to an emerging information technology*. Springer-Verlag, Berlin, 1997.
- S. Pepper, *The TAO of Topic Maps*. Proceedings of XML Europe 2000, 2000a.
- S. Pepper, *Topic Maps and RDF: A first cut*, <http://www.ontopia.net/topicmaps/materials/rdf.html>, 2000b.
- S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*. Prentice-Hall, 1995.
- J. Sowa, *Knowledge representation. Logical, philosophical, and computational foundations*. Brooks/Cole, 2000.
- E. Turban, J. Lee, D. King, H. M. Chung, *Electronic commerce. A managerial perspective*. Prentice Hall, New Jersey, 2000.