



Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?

Prof. Eero Hyvönen
TKK Viestintätekniikka
ja Helsingin yliopisto
Semantic Computing Research Group (SeCo)

Lähde: E. Hyvönen: Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?
TKK Viestintätekniikka, 2005.<http://www.cs.helsinki.fi/group/seco/ontologies/AsiasanastoistaOntologioihin.pdf>



- Miksi asiasanastot eivät riitä semanttisen webin tarpeisiin?
- Miten asiasanasto (tesaurus) voidaan muuttaa ontologiaksi?



- Nykyisten tesaurusten idea (tyyppiä YSA, MUSA, MASA,...)
 - Asiasanojen avulla tiedon etsinnän tarkkuus (precision) ja saanti (recall) paranevat
 - Semanttinen rakenne (aakkosellisen lisänä) helpottaa tiedon indeksoijaa asiasanojen löytämisessä
- Semanttinen rakenne hyödyllinen tietoteknisissä sovelluksissa
 - Esimerkiksi MuseoSuomi ym. semanttisen webin portaalit
 - » Termien laajennus kyselyissä
 - » Sanaston tarjoaminen hierarkioina loppukäyttäjälle
 - » Sisältöjen yhdistäminen ja suosittelulinkit
 - » Hakutulosten ryhmittely
 - » Käsitteiden visualisointi



- Indeksoija pystyy ymmärtämään tietämyksensä avulla käsitteiden väliset merkitykset ja suhteet
- Tietokoneella ei ole maailmankuvaa, vaan kaikki pitää kertoa eksplisiittisesti semanttisina rakenteina
- **Tietokone ei pysty hyödyntämään kunnolla nykyisiä sanastoja, sillä niissä ei semantiikkaa kuvata koneen kannalta riittävän täsmällisesti**



- **Miten asiasanasto muutetaan ontologiaksi?**
 - **Miten asiasanasto pitäisi laatia, jotta myös tietokone voisi sitä käyttää?**

Asiasanatesauruksen rakenne



- Ekvivalenssisuhteet
- Hierarkkiset suhteet
- Assosiatiiiviset suhteet

Suhde englanniksi	Suhde suomeksi	Merkitys
BT Broader term	LT	Laajempi termi, hierarkkinen suhde, käänteissuhde ST
NT Narrower term	ST	Suppeampi termi, hierarkkinen suhde, käänteissuhde LT
RT Related term	RT	Rinnakkais-termi, assosiatiiivinen suhde
USE Use	KAYTTÄ	Ilmaisee samantiltaisen synonyymia, käänteissuhde KORVAA
UF Used for	KORVAA	Ilmaisee ei-suositeltavan synonyymia, käänteissuhde KAYTTÄ
SN Scope note	HUOMAUTUS	Luomaahtaa termin merkitystä ja käyttöä

Taulukko 2.1. Tärkeimmät semanttiset suhteet resauroissa.

Esimerkki: Yleinen suomalainen asiasanasto YSA



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Noin 23.000 termiä
- 179 aluetta
 - Aluesuunnittelu
 - Anatomia
 - Antropologia
 - Arkeologia
 - ...
- 61 aluejoukoista koostuvaa ryhmää
 - Ryhmä 11: Maantiede. Kartografia. Kartoitus. Geodesia. Geologia. Paleontologia
- Perinteiset semanttiset suhteet
 - LT/ST, RT, KÄYTÄ, KORVAA
 - Taksonomiat vain osittaisia

Tesauruksen ontologisoiminen



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- 1. Ekvivalenssisuhteiden esittäminen
- 2. Hierarkkisten suhteiden esittäminen
- 3. Assosiatiivisten suhteiden esittäminen

1. Ekvivalenssisuhteiden kehittäminen



- Kaksi ratkaisutyyppiä
 - Termit luetellaan käsitteen ominaisuuksina
 - » Ongelma:
 - Joskus runsaasti nimiä
 - Esim. eri kielet, translitteroinnin jne.
 - Termit voivat riippua tilanteesta
 - Esim. eri museoiden luettelointitavat, eri indeksointiajat jne.



- Termeille kehitetään erilliset ontologiat, joista viitataan käsitteisiin
 - » Ongelma:
 - Useamman ontologian ylläpito
 - » Esim. MuseoSuomen eri museoiden termikortit
 - Termikorttiluokan yksilö kertoo
 - sanaan liittyvän käsitteen ontologiassa
 - onko sana suositeltu
 - monikon ja yksikön (esim. ooppera vs. oopperat)

Merkitysten erottelu



- Esimerkki: YSA:n termi johtaminen
- Termillä on kuitenkin monia merkityksiä
 - johtaminen (musiikki)
 - johtaminen (liiketalous)
 - johtaminen (sähkötekniikka)
 - johtaminen (matematiikka)
 - ...
- Johtaminen asiasanalla hakeminen ”johtaa” suureen sekaannukseen
- Ratkaisumalli:
 - eri merkityksille oman URI-tunnisteet
 - » esim. <http://yso.fi/johtaminen#>
 - näihin voidaan viitata sanoilla, esim. ”johtaminen”
 - **indeksointi on tehtävä URI-tunnisteilla eikä monimerkityksisillä sanoilla**

2. Hierarkkisten suhteiden esittäminen



- Yksilöiden erottaminen luokista
- Partitiivisten ja geneeristen suhteiden erottelu
- Hierarkioiden täydellistäminen

Yksilöiden erottaminen luokista



- Olio-perustainen (object oriented) mallinnus ja ohjelmointi keskeistä tietotekniikassa
 - Luokka määrittelee sen yksilöiden ominaisuudet
 - » Luokalla voi olla myös omia ominaisuuksia
 - Esim. luokan yksilöiden määrä
 - Yksilö perii ominaisuudet luokaltaan ja sen yläluokilta
- Esimerkki YSA:sta
 - Halley'n komeetta LT komeetat
 - Ongelma: ” Halley'n komeetta” on yksilö luokasta ”komeetat”, mutta kone ei voi tietää sitä
 - » Voi olla useita eri komeettoja, joilla on ”komeetat”-luokan ja sen yläluokkien ominaisuudet
 - » Ei voi olla useita ”Halley'n komeettoja”
 - » Yksilö ei ehkä voi olla kuin yhden luokan (ja sen yläluokkien) yksilö
 - Tämä oletus tehdään esim. Protege-editorissa

Partitiivisten ja geneeristen suhteiden erottelu



- LT/ST suhteella on monia merkityksiä
 - Ihmiset niiden ymmärtävät eron mutta koneet eivät
- Esimerkki YSA:sta
 - sairaalat LT terveydenhuoltolaitokset
komeetat LT aurinkokunnat
 - Kone hämmentyy säikähtäen
 - » sairaalat subClassOf terveydenhuoltolaitokset
 - OK: sairaalat perivät terv.huoltolaistosten ominaisuudet
 - » komeetat subClassOf aurinkokunnat ?
 - Väärin: komeetta ei ole aurinkokunta, sillä ei esim. ole keskustähteä, planeettoja jne.
- Ratkaisumalli:
 - sairaalat subClassOf terveydenhuoltolaitokset
 - komeetat isPartOf aurinkokunnat

Partitiivisen suhteen alatyyppejä



Meronyminen suhde	Esimerkki
part / whole	branch / tree
member / collection	tree / forest
piece / whole	piece-of-cake / cake
material / object	aluminium / airplane
feature / activity	paying / shopping
phase / process	childhood / growing-up
place / region	Helsinki / Finland

Taulukko 3.1. Hierarkkisia osa-kokonaisuussuhdetyyppejä.

Hierarkioiden transitiivisuuden esittäminen

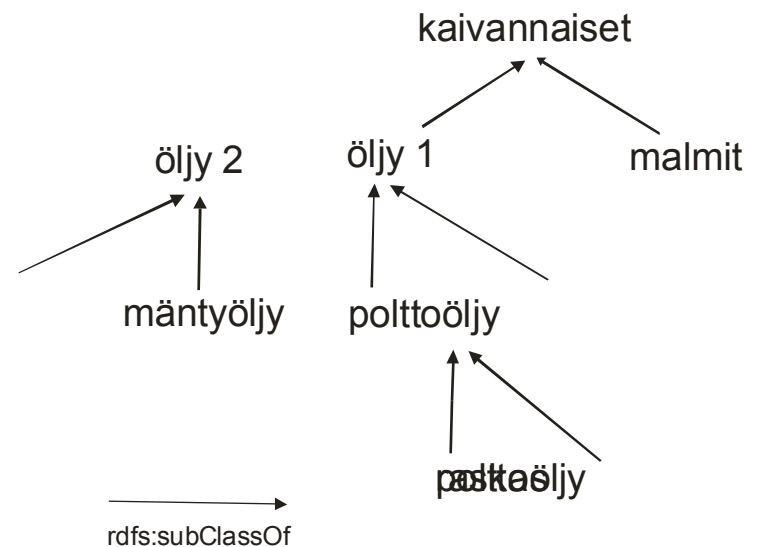


- Esimerkki YSA:sta

- mäntyöljy LT öljy
- polttoöljy LT öljy
- öljy LT kaivannaiset
- Ongelma: mäntyöljy EI OLE kaivannainen, mutta polttoöljy on

» Kone ei voi tajuta miksi tämän esitysmuodon perusteella, vaikka ihminen voikin

- Ratkaisu oikealla



Hierarkioiden täydellistäminen



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Kaikilla käsitteillä on oltava yläkäsite
 - Poikkeuksena kaiken juuri ”käsite”
- Esimerkki Museoalan asiasanaston MASA muunnoksesta Museoalan ontologiaksi MAO
 - MASA:ssa n. 2600 käsitettä ilman yläluokkaa!
- Käsitteiden jako ortogonaalisiin pääfasetteihin

3. Assosiatiivisten suhteiden esittäminen



Assosiatiivinen suhte	Esimerkki
osa / kokonaisuus (ei lineaariskinen)	kirurgia saksit RT leikkaussalit
ala / tutkimuskohde	seismologia RT maanjäristykset
prosessi / instrumentti	karjapaiminen RT kammat
ammattiryhmä / henkilö	karjapaimo RT karjapaimijat
toiminta / tuote	julkaiseminen RT kirjat
toiminta / kohde	opettaminen RT oppilaat
käsite / ominaisuus	naiset RT naisellisuus
käsite / alkuperä	vesi RT kaivot
syvä / seurauus	liika RT kuluminen
objekti / vastavoima	tuholaiset RT tuholaismyrkyt
raaka-aine / tuote	nahka RT nahkavyöt
toiminta / ominaisuus	ajamittaus RT tarkkuus
käsite / vastakohta	suvaitsevaisuus RT suvaitsemattomuus

Taulukko 3.2. Assosiatiivisia suhdetyyppejä (Aitchison, 2000). Suomeksi esitetyt esimerkit ovat luovutettuja eikä perustu YSA:an.



- Semanttinen erottelutarkkuus riippuu sovelluksesta
- Suhteiden esittämismalleja
 - Kuten taulukossa
 - » Ongelmana suhteiden suuri määrä ja niiden merkityksen ”epämääräisyys”
 - Kehyspohjainen esitystapa, jossa suhteiden merkitys yritetään palauttaa joukkoon primitiivejä

Kehysperustainen ratkaisumalli



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Assosiaatiot muodostuvat verbityyppisten kehystapahtumien kautta (frame)
 - Käsitteet toimivat eri temaattisissa rooleissa (thematic role)
 - » Esim. myymis/ostamistapahtumassa
 - agentti (agent)
 - kohde (patient)
 - hyötyjä (benefactor)
 - » myyminen ja ostaminen määritellään saman kehyksen avulla
 - Tekoälyssä ym. käytetty mallinnustapa
 - » Esimerkiksi. FrameNet USA:ssa
- Esimerkki YSA:sta
 - aurinkotuuli RT revontulet
 - Aiheuttaminen_1 agentti aurinkotuuli
tulos revontulet



- Perinteisen tesauruksen kevyt ontologisointi
 - Yksilöiden erottelu
 - Päämerkitysten erottelu
 - Partitiivinen suhde erotettu geneerisestä
 - Täydelliset hierarkiat
 - Transitiivisuus toimii
 - Semantiikka perustuu W3C:n RDFS-standardiin
- Tämän pohjalta voi laatia pidemmälle meneviä ontologioita
 - Ks. esimerkiksi W3C:n Web Ontology Language OWL