



Yleisen suomalaisen ontologian kehitystyö

Tuomas Palonen ja Katri Seppälä

Semanttisen laskennan tutkimusryhmä (SeCo)
Teknillinen korkeakoulu, mediatekniikan laitos

<http://www.seco.tkk.fi/>

Mihin ontologioita tarvitaan?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- mahdollisuus kehittää asiasanastoja ja laajentaa niiden käyttöä ontologiamuodossa
 - kun asiasanastot muokataan koneen tulkittaviksi ontologioiksi niiden
 - » monitulkintaisuutta karsitaan
 - » käsitteiden yksilöimiseksi tarvittavaa ja käsitteiden välisiä suhteita selventävää tietoa lisätään
- mahdollisuus tarjota uusia palveluja nykyisten rinnalla
 - semanttisen webin tekniikat mahdollistavat älykkäiden palvelujen kehittämisen esimerkiksi niin, että sanojen/merkkijonojen sijaan hyödynnetään ontologian kautta määriteltyjä käsitteitä

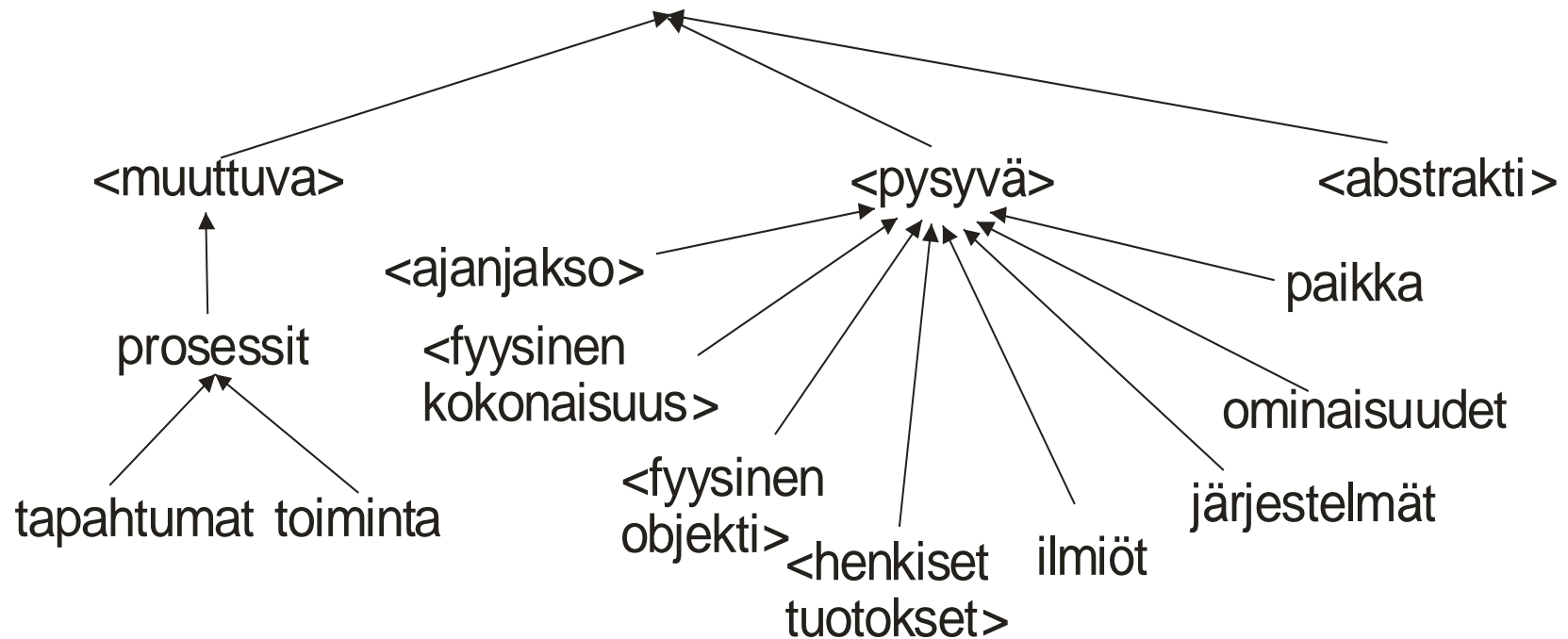
Millainen Yleinen suomalainen ontologia (YSO) on?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- perustuu Yleiseen suomalaiseen asiasanastoon (YSA)
- voidaan käyttää rakenteellisena mallina ja liittymäkohtana erikoisalojen ontologioihin
 - 8 suomalaista erikoisalan ontologiaa on jo liitetty YSOon
- sisältää n. 23 000 ontologisoitua käsitettä
 - YSAN päivitykset toukokuulle 2009 asti sisältyvät YSOon
- kielet: suomi
 - englannin ja ruotsin vastineiden (Allärs) täydentäminen ja tarkistus on käynnissä
- ontologian käsitteiden väliset suhteet
 - yläluokka-alaluokkasuhde
 - osa-kokonaisuussuhde
 - assosiatiivinen suhde

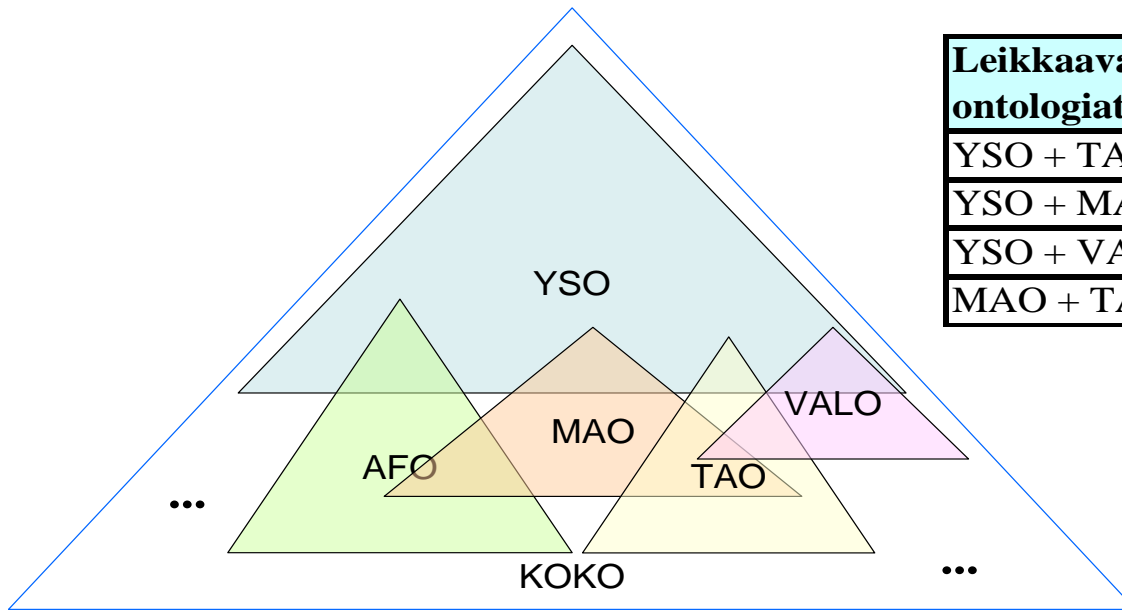
YSO – käsitteistön yhteen liittävä hierarkia



KOKO-ontologia: n. 37 000 käsitettä



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



Leikkaavat ontologiat	Yhteisten ekvivalenttien käsitteiden määrä
YSO + TAO	1990
YSO + MAO	2330
YSO + VALO	950
MAO + TAO	1190

Mitä YSO ei ole?



- kaikenkattava maailmankuva
 - käsitevalikoima on muotoutunut asiasanoituksen tarpeiden mukaan
- hierarkialtaan täydellinen
 - hierarkia on tehty karkealla tasolla ja sitä voidaan korjata ja täydentää yksilöityjen tarpeiden mukaan
- ainoa oikea ratkaisu
 - vaikka asioita voidaan tarkastella monesta näkökulmasta, voidaan silti sopia yhteisesti käytettävästä mallista
 - esimerkiksi yhdistelmäontologioissa esiin tuleva erikoisalojen näkökulma käsitteistä voi olla erilainen kuin YSOssa

Mitä hyötyä YSOsta on?



- mahdollistaa uudenlaisten palveluiden tarjoamisen
 - mahdollisuus semanttiseen tiedonhakuun
 - käsitelähtöisyys ja ontologian monikielisyys mahdollistavat kielten joustavan käytön hakutilanteessa riippumatta siitä, millä kielellä kuvailu on tehty
- helpottaa käsitteiden valintaa ja yhtenäistä käyttöä aineistoja kuvailtaessa, koska käsitteet on eritelty hierarkian avulla
 - ontologian avulla kuvailtaessa riittää, että kuvailuun valitaan samasta hierarkiasta vain yksi, sopivan tarkka käsite
- vähentää eri aloilla tehtävän työn päällekkäisyyttä, kun erikoisalojen ontologioita kehitetään entistä tiiviimmin suhteessa YSOon

YSOn käännös englantiin



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- YSOssa käännettävänä alunperin n. 8000 asiasanaa; tällä hetkellä kääntämättä n. 2000 asiasanaa
- kääntämätön osuus muodostui seuraavista osista: YSAn 6000 HKLJ:ssa kääntämätöntä asiasanaa + 2000 myöhemmin YSAan lisättyä asiasanaa
- käännöstyö alkoi huhtikuussa 2009 ja päättyi joulukuussa 2009; työtä on tehnyt pääosin Tuomas Palonen apunaan Katri Seppälä ja Anu Ylissalmi
- käännöstyö edennyt arviolta 70 käsitteen päivävauhtia
- jotkin asiasanat helposti käännettäviä, toiset huomattavasti vaikeampia

Käännöksen apuvälineet



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- painetut sanakirjat: lääketieteen, biologian, lakikielen ja talouselämän sanakirjat jne.
- nettisanastot: MOT-sanakirjat, IATE-termipankki, opetushallinnon, Eduskunnan kirjaston ja Sanastokeskuksen sanastot, terveydenhuollon Metatesaurus Rex jne.
- muut nettilähteet: Fennica, Wikipedia, Tilastokeskus, International Standard Classification of Occupations, EU-asiakirjat, lakikäännökset, Google (termien yleisyys) jne.
- asiantuntija-apu tarvittaessa

Käännösongelmia



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- käsitteet, joille ei ole vastinetta englantia puhuvassa maailmassa: viileä vyöhyke, kestävyyslajit, hirvikoirat jne; sosiaalialan termit: erikoisnuorisotyö, lähikasvattajat jne.
 - termit eivät aina hahmotu samalla tavalla eri kielissä: verolautakunta = tax committee, metsälautakunta = forest board, kunnallislautakunta = local council
- suomalaisen kulttuurin spesifit käsitteet: kalterijääkärit, korpelalaisuus, heimoaate
- erisnimet: Olhavan taistelu, Pyhän Elian kirkko, Sanoja & Säveliä
- ammatit ja oppiarvot: hallintopäälliköt, kunnanlääkärit, kandidaatit, leipurimestarit

Käännösratkaisuja



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- selvitetään, mikä on käsitteen asiasisältö VESAssa ja käyttökonteksti Fennicassa, monitulkintaisten sanojen kohdalla tämä on erityisen tärkeää
- etsitään mahdollisimman luotettava käännöslähde; harrastelijakäännökset eivät kelpaa
- tarkistetaan, että suomen kielen asiasana ja englannin vastine viittaavat samaan käsitteeseen
- jos sopivia vaihtoehtoja on useita, tarkistetaan yhtäältä, mikä englannin käsitteistä vastaa tarkimmin suomen käsitettä ja toisaalta mikä englannin käsitteistä on yleisimmin käytössä

Kehitettävää yhä



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- HKLJ-käännösten tarkistus: käännöksissä ei ole otettu huomioon ontologiahierarkiaa eikä monimerkityksisyyttä
- vaikeiden, kääntämättä jääneiden asiasanojen kääntäminen
- mahdollinen WordNetin ja YSO:n automaattisen siltaamisen kartoitus ja aiempien YSO-käännösten tarkistus





- YSOlla on keskeinen asema eri ontologioiden yhdistäjänä
 - englannin kieli on monilla aloilla tärkeä linkki kotimaisen kieliyhteisön ulkopuolelle
- kuten asiasanastot, ontologiat edellyttävät jatkuvaa kehitys- ja ylläpitotyötä ollakseen toimivia
- ontologioiden ylläpito on mahdollista toteuttaa hyvin organisoidulla yhteistyöllä
- lisätietoja ja sovellusten esittelyitä SeCo-ryhmän kotisivulta
 - <http://www.seco.tkk.fi/>
 - ONKI-palvelu: <http://www.yso.fi/>