

Epätäsmällisen tiedon esittäminen semanttisen webin ontologioissa

FinnOnto, 16.11.2005

Markus Holi (markus.holi@tkk.fi)

Semantic Computing Research Group
<http://www.seco.tkk.fi/>



Epätasällisyys

- Maailmaa koskeva tieto on epävarmaa
 - Sataako huomenna?
 - Onko Irakissa joukkotuhoaseita?
- Monet käsitteet ovat epämääräisiä
 - Onko vesi lämmintä?
 - Onko vuosi pitkä aika?
- Käsitteet ovat usein osittain päällekkäisiä ja peittävät toisiaan jossakin määrin
 - Lappi – Suomi
 - Aasia peittää suuremman osan maailman pinta-alasta kuin Eurooppa

Semanttisen Webin Ontologiat

- Semanttisen Webin Ontologiat perustuvat logiikkaan
 - Väitteet ovat tosia tai epätosia
 - Käsitteet ovat täsmällisesti määriteltyjä
 - Käsitteet sisältyvät toisiinsa tai ovat toisistaan erillisiä
- Tarvitaan menetelmiä epätäsmällisen tiedon mallintamiseksi Semanttisen Webin ontologioissa
 - Probabilistiset menetelmät
 - Sumea logiikka
 - Epämonotoninen logiikka

Esimerkkejä (1)

- Ontologioiden väliset käännökset (ontology mapping, ontology alignment)
 - Monissa tapauksissa käännettävälle käsitteelle ei löydy yhtä täsmällistä vastinetta toisessa ontologiassa, vaan mahd. useampi epätäsmällinen vastine
 - Esim. R. Pang & al.: A Bayesian Network Approach to Ontology Mapping
- Semanttisten web palveluiden löytäminen (semantic web service discovery)
 - Verrataan palvelun kuvausta palvelun hakijan tavoitteisiin
 - Mahdollisesti epätäydellinen täsmääminen kuvauksen ja tavoitteen välillä
 - F. Martin-Recuerda and D. Robertson: Discovery and Uncertainty in Semantic Web Services

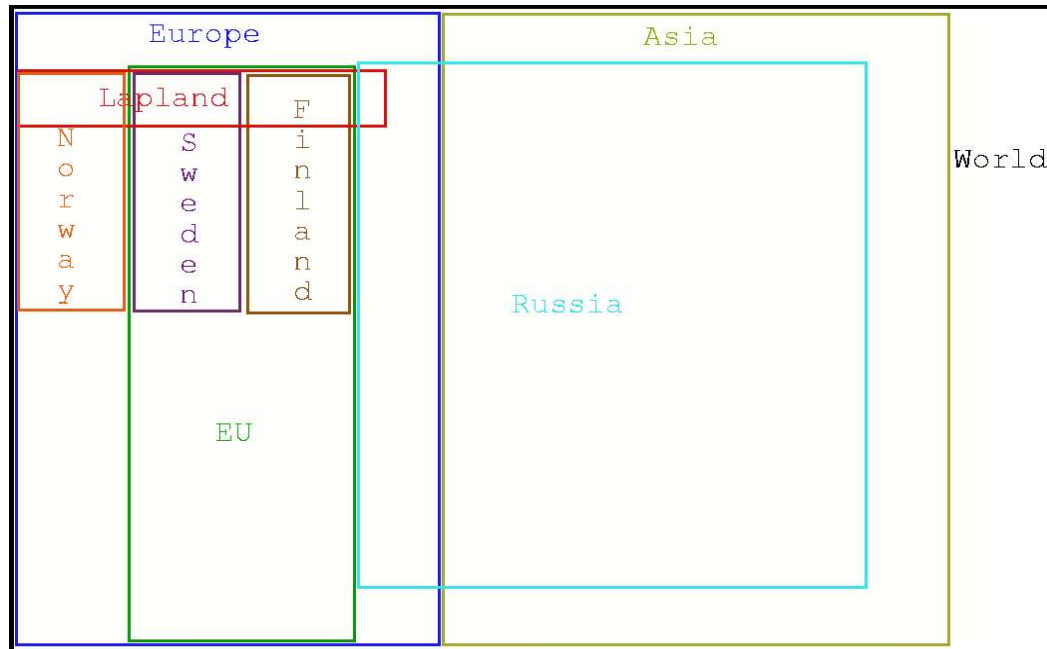
Esimerkkejä (2)

- Ontologioiden automaattinen oppiminen datasta
 - Luonnollisen kielen prosessointia ja koneoppimista hyödyntäen opitut ontologiat sisältävät epätasällisyyttä
 - P. Haase and J. Völker: *Ontology Learning and Reasoning – Dealing with Uncertainty and Inconsistency*
- Ontologiaperustainen tiedonhaku
 - Tuloksia voidaan järjestää käsitteiden välisen päällekkäisyyden perusteella
 - M. Holi and E. Hyvönen: *Probabilistic Information Retrieval Based on Conceptual Overlap in Semantic Web Ontologies*
- Käsitteiden merkitysten muutoksien mallintaminen ajassa
 - Käsiteversioiden osittainen päällekkäisyys
 - T. Kauppinen and E. Hyvönen: *Modeling and Reasoning about Changes in Ontology Time Series*

Käsitteiden välisen päällekkäisyyden mallintaminen

Ongelma

- Ontologiat perustuvat täsmälliseen logiikkaan eivätkä tarjoa keinoja käsitteiden välisen päällekkäisyyden määrän mallintamisen
- Täsmällisen partOf -hierarkian avulla ei voida ilmaista alla olevan Venn diagrammin tilannetta
 - Käsitteiden välinen osittainen päällekkäisyys
 - Peittämisen aste

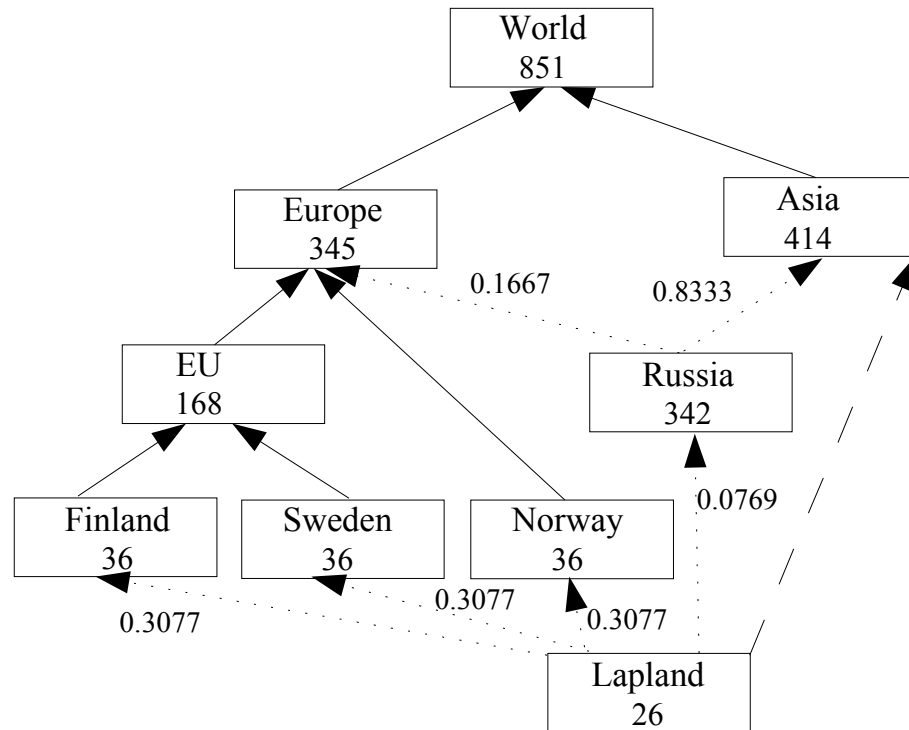


Tavoite

- Osumien relevanssin määrittäminen hakukoneessa käsitteiden välisen päällekkäisyyden perusteella
 - $o = |\text{Hakuk.} \cap \text{Annot.}| / |\text{Annot.}|$
- Tavoitteeseen päästään kahdessa vaiheessa:
 1. Luodaan notaatio, jonka avulla voidaan esittää sisältyminen asteet
 2. Lasketaan päällekkäisyysarvot kääntämällä ontologia Bayesverkoksi

| Haku-käsite | Annotaatio | Relevanssi (o) |
|-------------|------------|----------------|
| Lapland | Finland | 0.2222 |
| Lapland | Norway | 0.2222 |
| Lapland | Sweden | 0.2222 |
| Lapland | EU | 0.0953 |
| Lapland | Europe | 0.0754 |
| Lapland | World | 0.0306 |
| Lapland | Russia | 0.0059 |
| Lapland | Asia | 0 |

1. Päällekkäisyyden esittäminen



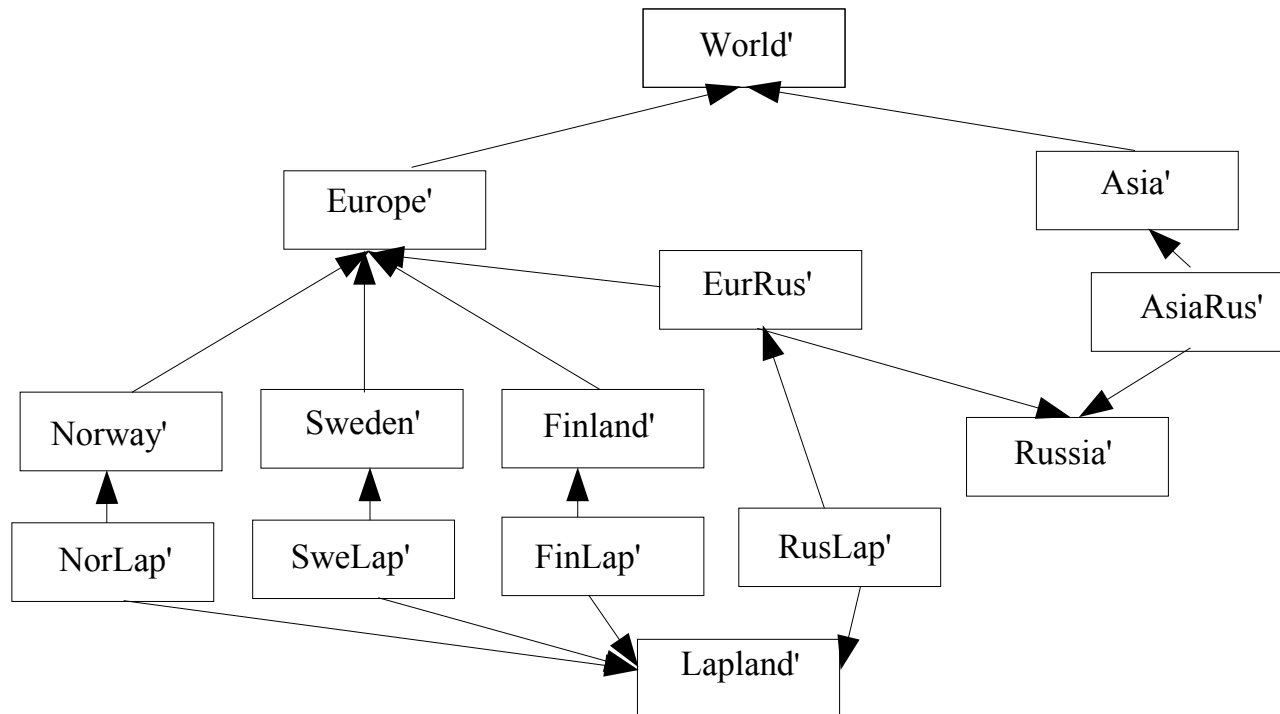
sisältyminen täysin —————▶

osittainen sisältyminen▶

erillisyys - - - - -▶

2. Pällekkäisyysarvojen laskeminen

- Käsitteiden A (valittu) ja B (annotaatio) välinen päällekkäisyysarvo voidaan esittää ehdollisena todennäköisyytenä $P(B'=true|A'=true)$
- A' on binäärimuuttuja, s.e., A'=true tarkoittaa, että käyttäjä on kiinnostunut käsitteestä A
- Pällekkäisyysarvojen laskemiseksi päällekkäisyysgraafi muutetaan Bayesverkoksi



Yhteenveto

- Yksinkertainen notaatio käsitteiden välisen sisällymisen määrän esittämiseen
- Laskutoimitukset palautuvat Bayesverkkopäättelyksi
- Joukko-oppi ja Bayesverkot valittiin niiden tarjoaman vakaan matemaattisen pohjan vuoksi