

Hylkeet syövät lohia

DAML-ontologian hyödyntäminen Itämeriportaalissa



Aleksi Grym

Ch5 Finland Oy

Itämeriportaali

- Merentutkimuslaitoksen ylläpitämä verkkosivusto
- 2300 kpl puolitieteellisiä Itämeri-aiheisia dokumentteja
- Dokumentit tallennettu XML-muotoisina SQL-tietokantaan
- Dokumentteja haetaan käsitteiden perusteella
- www.itameriportaali.fi

Käsitehaku

- DAML-kielellä koodattu ontologia
- Hakusanana toimii jokin ontologian käsitteistä
- Ontologia muodostaa *käsiteavaruuden*
- Dokumentit sijaitsevat käsiteavaruudessa
- Hakutuloksina valitaan ne dokumentit, jotka ovat *lähellä* haettua käsitettä
- Etäisyys (läheisyys) määritellään predikaattien (rdf:property) avulla

Ontologia

- Käsitteet jaettu aihepiireihin:
 - Maantieteelliset
 - Biologiset
 - Fysikaaliset
 - Yhteiskunnalliset
 - Tutkimukselliset

Ontologia

- Jokaisen aihepiirin sisällä käsitteet muodostavat useita hierarkioita
- Käsitteitä yhteensä joitakin satoja
- Predikaatteja yhteensä joitakin kymmeniä
- Ontologialauseita (rdf-triplettejä) yhteensä toista tuhatta

Ontologia

Esimerkki (maantiede):

```
<Country rdf:ID="Denmark">
  <hasCoastOn rdf:resource="ArkonanAllas" />
  <hasCoastOn rdf:resource="DanishStraits" />
  <hasCoastOn rdf:resource="NorthSea" />
</Country>

<SeaRegion rdf:ID="DanishStraits">
  <rdf:type rdf:resource="SampleSiteObservation" />
  <isPartOf rdf:resource="BalticSea" />
</SeaRegion>

<Sea rdf:ID="BalticSea" />

<rdfs:Class rdf:ID="Sea">
  <rdfs:subClassOf rdf:resource="#SeaRegion" />
</rdfs:Class>
```

Ontologia

Esimerkki (yhteiskunta):

```
<Government rdf:ID="GovOfDenmark">  
  <rules rdf:resource="Denmark" />  
  <hasHQIn rdf:resource="Koebenhavn" />  
  <isMemberIn rdf:resource="EuropeanUnion" />  
</Government>
```

```
<Capital rdf:ID="Koebenhavn">  
  <isLocatedIn rdf:resource="Denmark" />  
  <isCapitalOf rdf:resource="Denmark" />  
</Capital>
```

Ontologia

Esimerkki (yhteiskunta):

```
<rdfs:Class rdf:ID="Capital">
  <rdfs:subClassOf rdf:resource="#City" />
  <rdfs:subClassOf>
    <daml:Restriction daml:mincardinality="1" daml:maxcardinality="1">
      <daml:onProperty rdf:resource="#isCapitalOf" />
      <daml:toClass rdf:resource="#Country" />
    </daml:Restriction>
  </rdfs:subClassOf>
</rdfs:Class>

<Union rdf:ID="EuropeanUnion" />

<rdfs:Class rdf:ID="Union">
  <rdfs:subClassOf rdf:resource="#Organization" />
</rdfs:Class>
```


Ontologia

Esimerkki (yhteiskunta):

```
<Hylje rdf:ID="Harmaahylje">
  <rdf:type rdf:resource="Sekundaarituottaja" />
  <rdf:type rdf:resource="Saalistaja" />
  <eats rdf:resource="Kilohaili" />
  <eats rdf:resource="Silakka" />
  <eats rdf:resource="Turska" />
  <eats rdf:resource="Lohi" />
</Hylje>

<Kala rdf:ID="Lohi">
  <rdf:type rdf:resource="Sekundaarituottaja" />
  <rdf:type rdf:resource="Saalistaja" />
  <eats rdf:resource="Kilohaili" />
  <eats rdf:resource="Silakka" />
</Kala>
```

Ontologia

- Predikaateille on annettu painokertoimet
- Painokertoimet määrittävät kahden käsitteen välisen etäisyyden
- Dokumentista etsitään kaikki siellä esiintyvät käsitteet
- Predikaattien avulla löydetään myös läheisiä käsitteitä
- Dokumentti on sitä lähempänä käsitettä, mitä enemmän dokumentissa esiintyy käsitettä tai sitä lähellä olevia käsitteitä

Käsitteiden esiintyminen

- Käsitteille on määritelty, minkä näköisenä ne esiintyvät dokumenteissa
- Haku ymmärtää, että käsitteestä puhutaan, vaikka sitä ei suoraan esiintyisi dokumentissa
- Tämä tekee käsitehausta riippumattoman
 - Kieliversioista
 - Taivutusmuodoista
 - Synonyymeistä

Mitä opimme?

- Se toimii!
- Minkälainen ontologia tuottaa intuitiivisesti järkeviä tuloksia?
 - Pienelläkin määrällä käsitteitä ja predikaatteja saadaan aikaan suuri määrä lauseita
 - Painokertoimien (etäisyysmittojen) kalibrointi tärkeää ja vaikeaa

Yhteystiedot



Aleksi Grym

Ch5 Finland Oy

aleksi.grym@ch5finland.com