



Aalto University
School of Science



Kieliteknologian hyödyntäminen Semanttisessa Finlexissa

Aalto-yliopisto

Minna Tamper

Sisällys

1. **Taustaa**
2. **Julkaisussa**
3. **Automaattinen asiasanoitus**
4. **Visualisointi: Avainsanapilvi**

Taustaa

- **Tavoitteena helpottaa säädösten löytämistä parantamalla niiden hakua ja selailua**
- **Tämä saavutetaan rikastamalla metadataa**
- **Tavoitteena myös parempi linkitettävyys**
- **Automaattinen annotointi: kustannustehokkuus, systemaattinen, runsaampi ja harmonisempi sisällön kuvailu**
- **Datan (oikeuden päätökset) avaamisen kannalta merkityksellistä on anonymisointi**

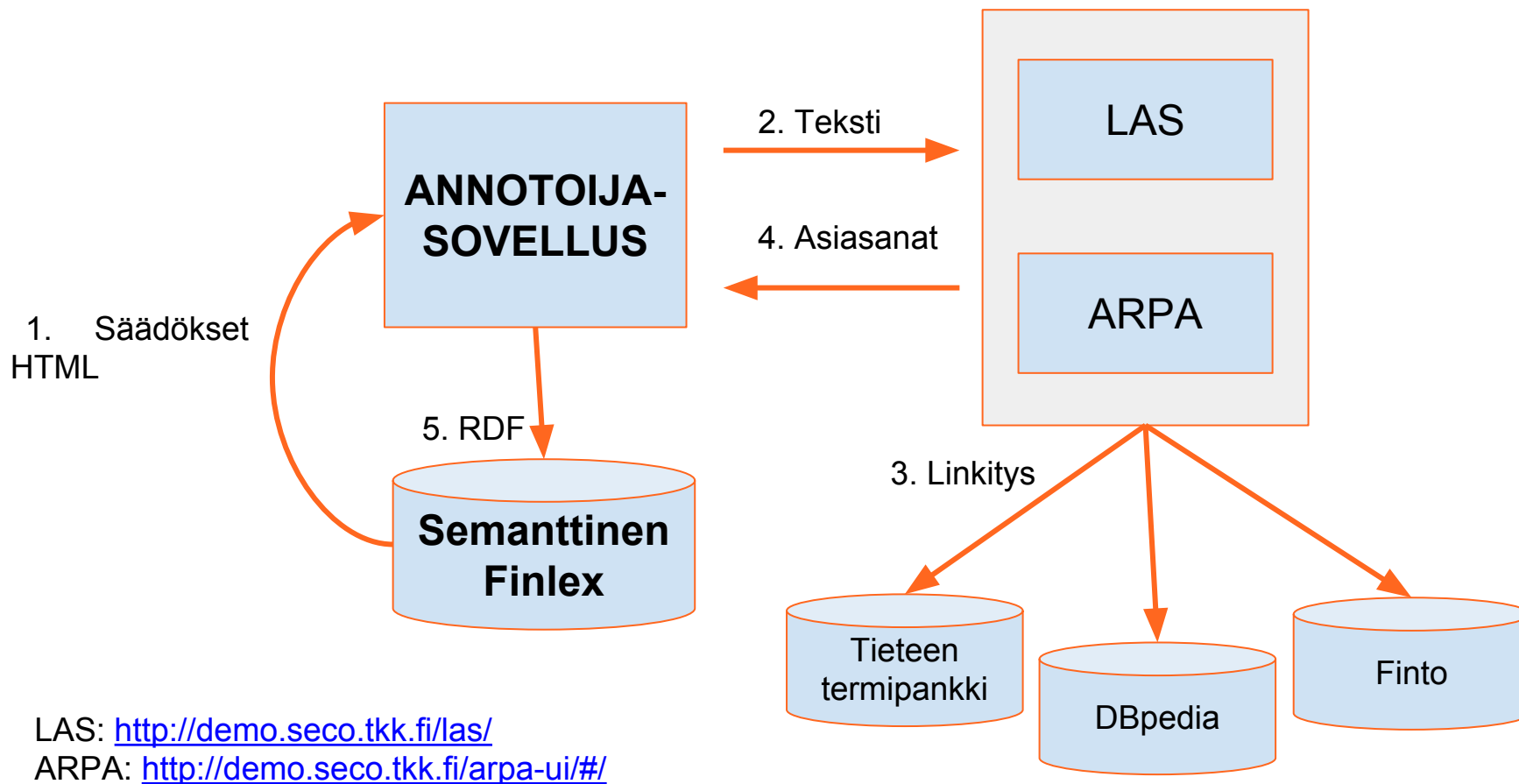
Julkaisussa

- **Asiasanoitus ajantasaiselle säädöksille (~2800)**
 - 103 414 asiasanaa
 - arviolta n. 37 asiasanaa per säädös
- **Avainsanapilvi ajantasaiselle säädöksille**
- **Kaikki suomeksi**

Automaattinen annotointi

- **Tavoitteet:**
 - Helpottaa kiinnostavien säädösten löytämistä
 - Tarjota yhteyksiä säädösten välillä
- **Menetelmät:**
 - Säädosmetadatan rikastus säädöstekstejä louhimalla

Prosessi



Asiasanoituksessa käytetyt sanastot

- **Käytetyt sanastot:**
 - KOKO-ontologia (Finto)
 - Oikeustieteen termitietokanta (Tieteen termipankki)
 - DBpedia
- **Asiasanoitus on eriytetty varsinaisesta datajulkaisusta**
 - oma graafi: <http://ldf.fi/finlex-annotation/sd>
- **Asiasanat on liitetty dataan RDF-ominaisuudella:**
 - <http://data.finlex.fi/schema/common/autRecSubject>

Haasteet

- **Automaattisen asiasanoituksen suurimmat haasteet ovat olleet väärät tai epäolennaiset asiasanat**
 - Yksikäsitteisyys
 - Merkitsevien ja merkityksettömien asiasanojen erottelu kokonaisuuden kannalta
- **Asiasanojen löytäminen lyhyistä teksteistä**
- **Puuttuvat asiasanat, kuten asiasanojen yläkäsitteet**

Visualisointi: Avainsanapilvi

- **Tavoitteet**
 - Hahmottaa lukijalle säädösten keskeinen sisältö
 - Tiedon etsintä (tulevaisuudessa)
- **Menetelmät:**
 - Säädosmetadatan rikastus säädöstekstejä louhimalla ja ontologioita hyväksikäyttämällä

Avainsanapilvi

Maa- ja metsätalousministeriön asetus eräitä elintarvikkeita koskevista vaatimuksista



Avainsanapilvi

- **Perustuu avoimen lähdekoodin pyTagCloud-sovellukseen**
 - sovellus pohjautuu Wordleen
 - lisätty: suomen kielen tuki, mukautettu palvelemaan Semanttisen Finlexin tarkoitusperiä
- **Semanttisen Finlexin säädöstekstit on perusmuotoistettu LAS-työkalulla ennen avainsanapilven muodostusta**
- **Sovelluksen tuloksia on rikastettu KOKO-ontologian yläkäsitteillä**

Avainsanapilvi: Haasteet

- Epäoleellisten ja merkittävien termien rajanveto kokonaisuuden kannalta ja säädöskohtaisesti
- Tekniset kielisidonnaiset haasteet
 - Sanojen sovittaminen kuviin
 - Kuvien skaalaus



Laki metsätuhojen torjunnasta

Tulevaisuus

- **Hakukäyttöliittymän kehittäminen asiasanojen ja avainsanapilven avulla**
- **Asiasanoituksen yksikäsitteistäminen, tarkentaminen ja rikastaminen**
- **Avainsanapilven kehittäminen**
- **Mahdollisesti: Anonymisointi (oikeuden päätökset)**