

Tampereen yliopisto
Informaatiotutkimuksen laitos

**Suomenkielen ontologinen
tiedonhaku**

Käsiteperustainen tiedonhaku tietokannasta

- perustuu kolmeen abstraktiotasoon: käsite-, ilmaisu- ja esiintymätasot
- ontologian käyttö parantaa hakutulosta: parempi tarkkuus ja saanti
- voidaan huomioida suomen taivutusmuotoihin liittyviä ongelmia
- voidaan ottaa haussa huomioon hakuavaimiin semanttisesti liittyvät toiset avaimet, kuten synonyymit, hyponyymit, meronyymit

Mitä tehty tähän mennessä

- käsiteperustaisen tiedonhakuliittymän prototyyppi (CIRI)
 - **Concept-based Information Retrieval Interface**
 - perustuu kolmitasomalliin
 - verkkosovellus, jonka avulla voidaan käsitetasolla tehtyjen käsitevalintojen perusteella automaattisesti rakentaa kyselyjä tekstitietokantoihin
- ilmaisu- ja esiintymätasot ovat hakuontologioille spesifejä
 - johdetaan toisaalta käsitetason perusteella
 - toisaalta muiden käytettävissä olevien resurssien perusteella
- ontologioiden vaihtaminen ja yhteiskäyttö muiden ontologiasovellusten kanssa tapahtuu käsitetasolla
- elintarvikealan ontologia
 - käyttäjät testit CIRI:llä

Toisen tutkimus- ja kehitysvuoden tavoitteet I

■ **A) Hakuontologiaeditori**

- 1. Editorin tukivälineet
 - a. Monikielisyyden tuki
 - b. Sanastuksen tuki
 - c. Ilmaisujen hallinnan tuki (NLP)
 - d. Tarkistimet (suhdeverkko)
- 2. Käyttäjättestaus

Toisen tutkimus- ja kehitysvuoden tavoitteet II

■ (B) CIRI

- 1. Hakuympäristöjen tuki
 - a. Perinteiset ympäristöt (TRIP, InQuery)
 - b. Web-ympäristöt, ontologia –metahaku
- 2. Monikielisyyden tuki, indeksityyppien tuki
- 3. Hakutulosten visualisointi ja organisointi
- 4. Käyttäjättestaus

Miksi kolmitasomalli?

- esiintymä voi olla melkein mitä vaan kohdeindekseistä riippuen
 - tekstiä, kuvaa, ääntä
- eri kohdeindeksejä vastaavat esiintymät ovat samassa ontologiassa
 - valitaan hakukoneen mukaan
- ei manuaalista annotointia