

CIRI

Ontologiaperustainen tiedonhakuliittymä

Eija Airio, Kalervo Järvelin, Sari Suomela,
Pirkko Saatsi ja Jaana Kekäläinen
Tampereen yliopisto
Informaatiotutkimuksen laitos

Ontologian kolmitasomalli

- kehitetty tiedonhaun tarkoituksiin
- tasot
 - käsitetaso (conceptual level)
 - hierarkkisesti järjestetty
 - ilmaisutaso (expression level)
 - synonyymeja
 - luonnollista tai synteettistä kieltä
 - jokaista käsitettä vastaa 0, 1 tai useampia ilmaisuja
 - esiintymätaso (occurrence level)
 - indeksiä vastaava täsmäytysmalli
 - jokaista ilmaisua vastaa 1 tai useampi esiintymä



CIRI (Concept-based Information Retrieval Interface)

- perustuu kolmitasomalliin
- Java-servletit, SQL-kanta
- käytetään www-selaimen kautta
- graafinen ontologiaesitys
- käyttäjä voi selata avaamaansa ontologiaa ja valita käsitteitä hakua varten
- käyttäjä valitsee hakukoneen ja tietokannan valikosta

CIRI

- CIRI muodostaa kyselyn käyttäjän valitsemista käsitteistä, käyttäjän valitseman hakukoneen syntaksin mukaisesti
- käyttäjä määrää, kuinka monta tasoa hakua laajennetaan (oletuksena kolme)
 - assosiatiivisuhteita laajennetaan vain yksi taso
- lähettää hakulausekkeen hakukoneelle, joka tekee haun käyttäjän valitsemasta tietokannasta

CIRI:n käyttöliittymä

Haku

Ontologia **Elintarvikealan sanasto** **Ominaisuudet**

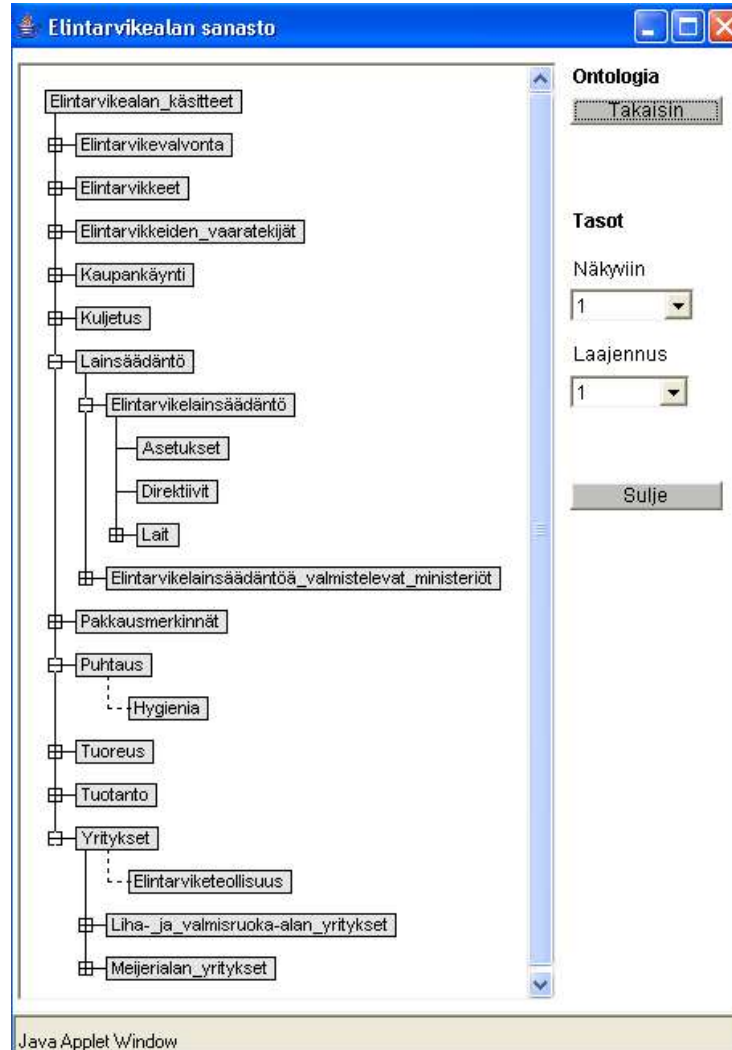
 

Ciri-haku

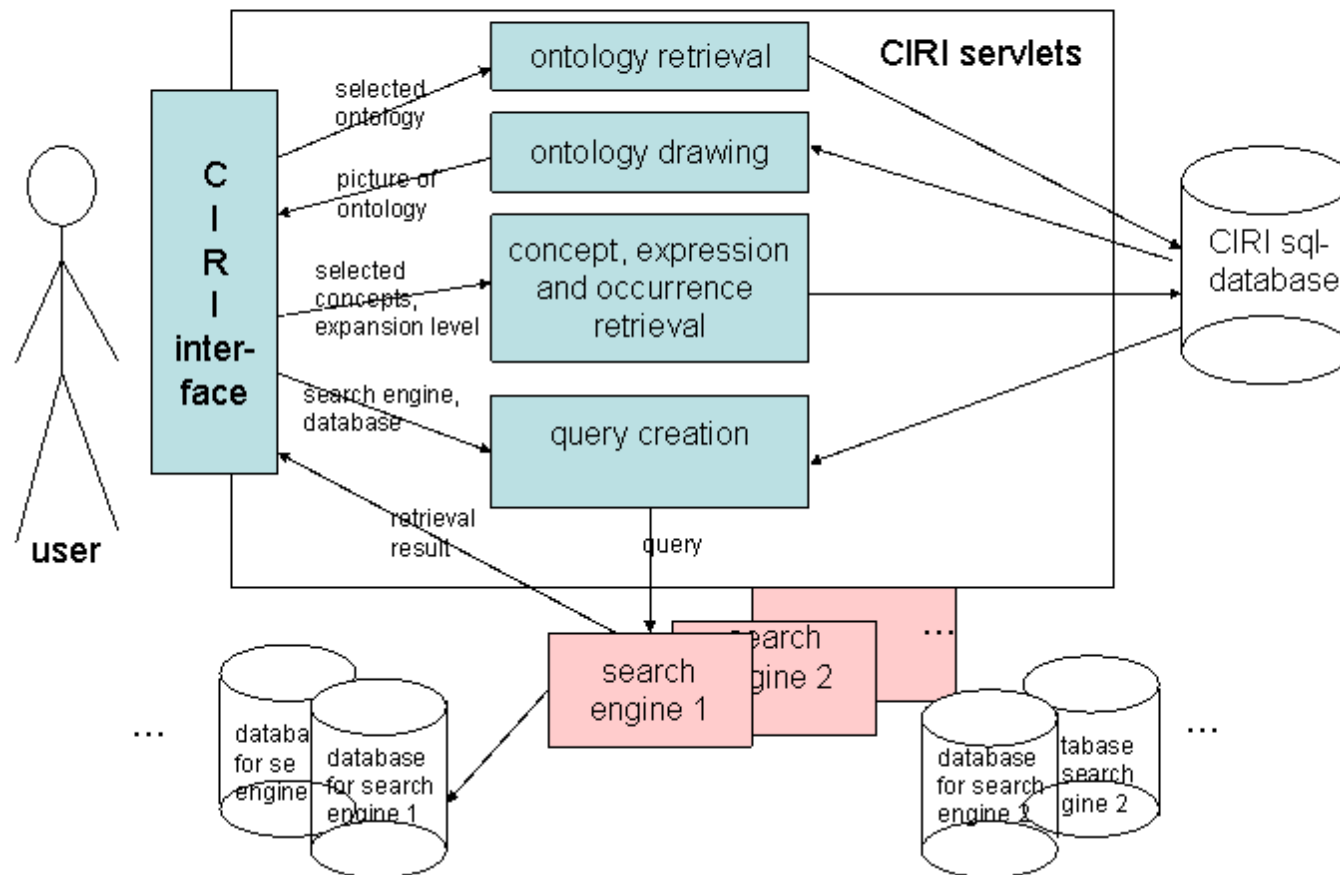
Käyttöohjeet

1. Paina **Ominaisuudet** -nappia ja valitse haluamasi tietokanta.
2. Valitse sanasto kohdan *Ontologia* alavetovalikosta ja paina **Avaa**. Sanaston navigointipuu aukeaa omaan ikkunaan.
3. Valitse sanastosta haluamasi käsitteet, sekä montako tasoa haluat käsitteitä laajentaa. Sitä mukaa, kun valitset käsitteitä, hakulause muodostuu kohtaan *Haku*.
4. Kun olet valinnut kaikki haluamasi käsitteet, älä sulje *Ontologia*-ikkunaa, vaan jätä se auki. Paina *Muodosta kysely*, jolloin Ciri laajentaa valitut käsitteet haluttuun tasoon asti, ja muuttaa laajennuksen tulokset niitä vastaaviksi täsmäytysmalleiksi. Tämä laajennettu kysely ilmestyy haku-appletin alapuolella sijaitsevaan tekstikenttään. Voit vielä halutessasi muokata syntynyttä hakuja tai lisätä sen perään omia hakusanajasi. Suorita haku painamalla *Lähetä haku*. *Poista valinnat* -nappi tyhjentää kaikki valinnat.

CIRI:n ontologianäkymä



CIRI:n toiminta



Esiintymätason lisääminen ontologiaan

- ei tarvitse annotoida kunkin uuden tietokannan jokaista dokumenttia, vaan ontologian esiintymätason päivitys riittää
 - samantyyppisille tietokannoille tarvitsee lisätä esiintymät vain kerran
- erilaisia tietokantatyyppejä - kutakin varten luodaan esiintymä
 - tekstitietokannat
 - perusmuotoindeksi
 - stemmattu indeksi
 - taivutusmuotoindeksi
 - kuvatietokannat
 - äänitietokannat

Esiintymätason lisääminen ontologiaan tekstitietokantaa varten

- NLP (natural language processing) –työkalut
 - parasta käyttää samaa välinettä kuin tietokannan indeksoinnissa
 - NLP-työkalujen liittäminen ontologiaeditoriin
- perusmuotoindeksi
 - normalisointityökalut
 - yhdyssanan pilkkominen
- stemmattu indeksi
 - sama stemmeri kuin indeksoinnissa
- taivutusmuotoindeksi, katkaisu mahdollista
 - sanavartalogeneraattori
- taivutusmuotoindeksi, ei katkaisua
 - sanavartalogeneraattori
 - grepataan indeksiä sanavartaloilla
 - mahdollisesti tarkistetaan normalisointityökaluilla tulos, koska indeksin greppaus tuottaa yleensä paljon roskaa

Käyttäjätetit elokuussa 2004

- elintarvikealan ontologia
 - 479 käsitettä
 - 689 ilmaisia
 - 2087 esiintymää
 - 603 käsitteiden välistä suhdetta
- Tähän mennessä 12 testihenkilöä, jotka edustivat
 - elintarviketeollisuuden yrityksiä
 - ravintoloita
 - elintarvikealaan liittyvää tietojärjestelmäyritystä
 - elintarvikealan opiskelijoita

Käyttäjätestien taustaa

- osittaistäsmäyttävä INQUERY
- Aamulehti ja Kauppalehti artikkeleita v. 2000-2004
- jokainen henkilö sai neljä testiaihetta
 - kahdessa sai käyttää ontologiaa
 - kaksi ilman ontologiaa
- aihe-järjestelmä-yhdistelmiä kierrätettiin
 - jokainen aihe tuli tehtyä yhtä monta kertaa ontologian kanssa ja ilman
 - tehtävien suoritusjärjestystä vaihdettiin

Käyttäjätestit – tuloksia I

- käyttäjiltä kysyttiin, kumpaa järjestelmää oli helpompi käyttää
 - 8 käyttäjää oli sitä mieltä, että järjestelmää oli helpompi käyttää ilman ontologiaa
 - kahden mielestä ontologiahaku oli helpompi
- käyttäjiltä tiedusteltiin, kummasta järjestelmästä he pitivät enemmän
 - 7 piti enemmän järjestelmästä ilman ontologiaa
 - 3 piti enemmän ontologiahausta

Käyttäjätестit – tuloksia II

- ”Sanastosta löytyi helposti tehtävissä tarvittavat hakusanat”
 - täysin tai melko samaa mieltä oli 8 käyttäjää
- ”Sanasto auttoi hakusanojen ideoinnissa”
 - täysin tai melko samaa mieltä oli 8 käyttäjää
- ”Sanasto auttoi jäsentämään aihetta”
 - 9 käyttäjää oli täysin tai melko samaa mieltä
- ”Sanaston käyttäminen oli työlästä”
 - 4 käyttäjää oli melko samaa mieltä
 - 6 oli vähän tai täysin eri mieltä

Käyttäjätestien yhteenvetoa

- kokeessa käytetty ontologia vastasi käyttäjien sanastoa
 - useimmat käyttäjät löysivät ontologiasta samat termit, joita itse käyttävät samoista käsitteistä
 - ontologian rakenne oli käyttäjien mielestä looginen
- haku ilman ontologiaa tuotti tulosjoukon kärjessä parempia tuloksia kuin ontologiahaku
- käyttäjät pitivät käyttöliittymää ilman ontologiaa helpompana käyttää...
 - käyttäjien kommentteista kävi ilmi, että uuteen järjestelmään täytyisi tottua
- ...mutta hakujärjestelmän käyttö ontologian kautta oli myös helppoa
- useimpien käyttäjien mielestä ontologia auttoi jäsentämään hakuaihetta