



Airo: Ontologiaperustainen indeksointi- ja hakutyökalu

Matias Frosterus, Olli Alm
Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK),
Laboratory of Media Technology
and
University of Helsinki, Department of Computer Science

<http://www.seco.tkk.fi>

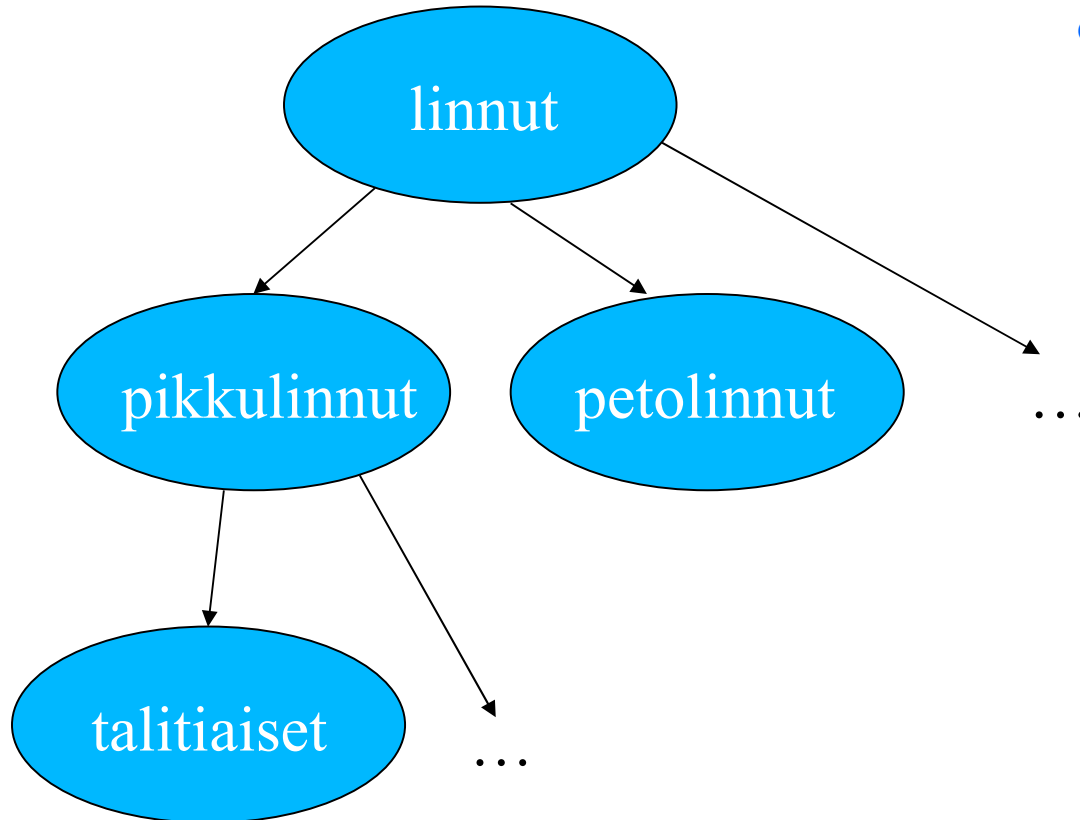


- Hakuominaisuuksien parantamien tekstihauulle ontologioita hyödyntämällä
 - sovelluskohteena esimerkiksi sanomalehtiarkistot
 - » miljoonia artikkeleita
 - » lisää joka päivä
 - vaatimuksena automaattisuus
- Tekstistä löydettyjen ontologisten käsitteiden laajentaminen aineiston ontologista jatkohyödyntämistä silmällä pitäen



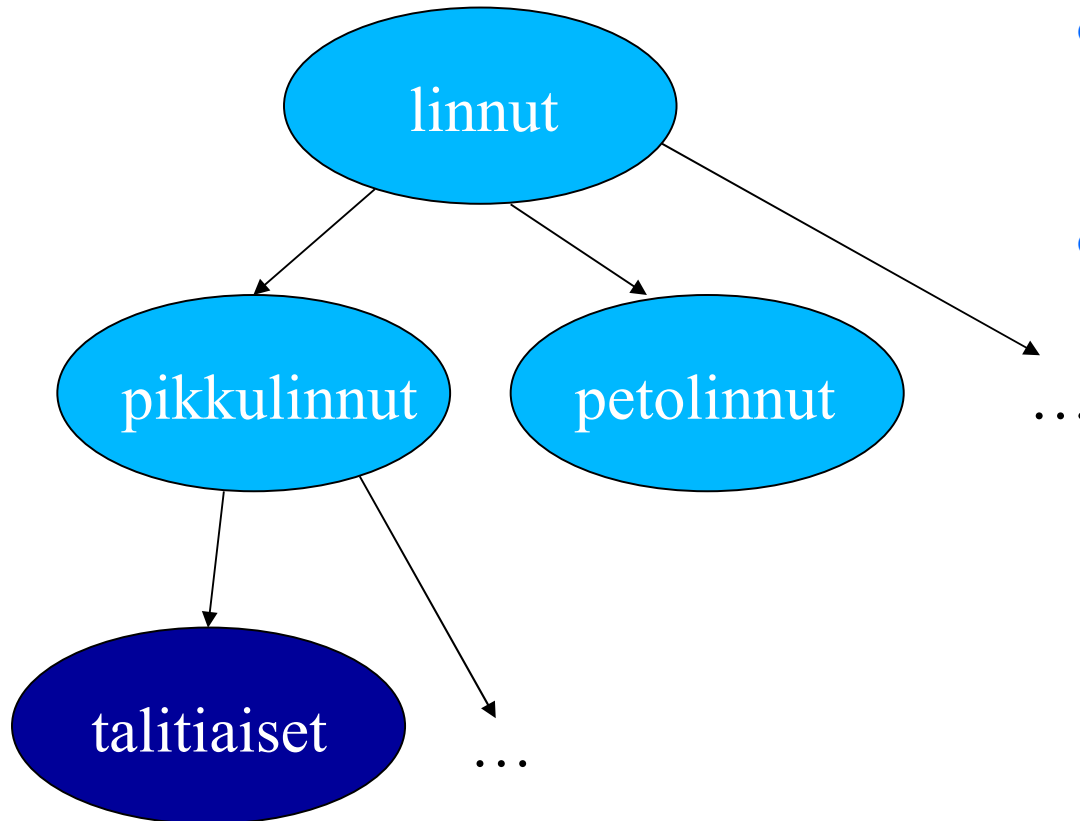
- Automaattinen käsitteiden tunnistaminen tekstistä
 - tapahtuu Pokalla
 - eri ontologioiden käsitteet talletetaan omiin indekseihinsä, jolloin haun aihepiiriä voi rajata valitsemalla ontologia, jonka käsitteitä haussa hyödynnetään
- Käsiteklusterointi

Käsiteklusterointi 1/4



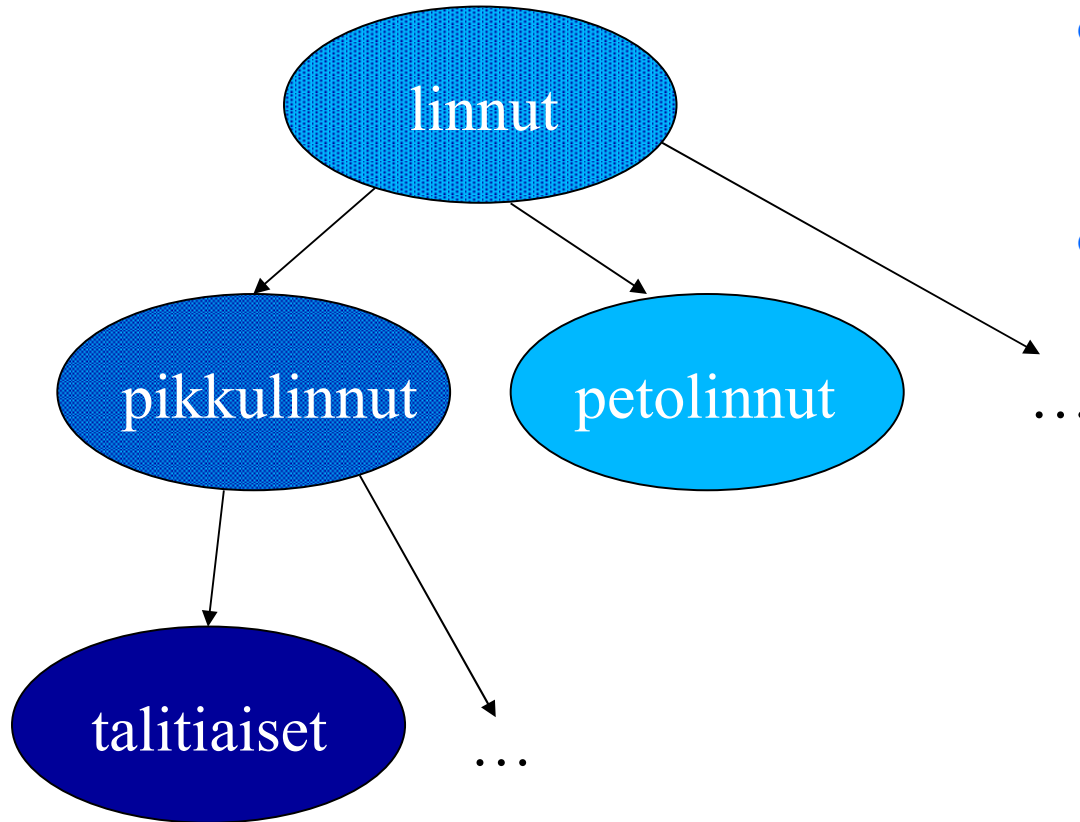
- Ontologinen hierarkia
 - nuolet kuvaavat aliluokkasuhteita

Käsiteklusterointi 2/4



- Ontologinen hierarkia
 - nuolet kuvaavat aliluokkasuhteita
- Tekstistä löytyy termi "talitiainen"
 - täsmää ontologian käsitteeseen talitiaiset

Käsiteklusterointi 3/4



- Ontologinen hierarkia
 - nuolet kuvaavat aliluokkasuhteita
- Tekstistä löytyy termi ”talitiainen”
 - täsmää ontologian käsitteeseen talitiais
 - Nostaa myös lähikäsitteiden painoa
 - laajuus ja voimakkuus vapaasti konfiguroitavissa

Käsiteklusterointi 4/4 - Esimerkki



HELSINKI UNIVERSITY OF TECHNOLOGY

Media Technology

- ”Maija Meikäläinen on missivuotensa jälkeen luonut uraa mallina.”
 - Tekstistä löydetään mm. käsitteet missi ja malli
 - Termi ’malli’ vastaa kolmeen käsitteeseen
 - » ”pienoismalli”-malli
 - » ”valokuvamalli”-malli
 - » ”roolimalli”-malli
 - YSO:ssa missit-käsitteellä ja ”valokuvamalli”-mallit-käsitteellä on assosiatiivinen suhde
 - » Missi-termi voimistaa ”valokuvamalli”-tulkintaa eli kyseisen lauseen ”valokuvamalli”-malli-käsite saa lisää painoa
 - Hakukone tulkitsee lauseen relevantimmaksi ”valokuvamalli”-malli haulle kuin lauseen ”Uudet Volvo-mallit ovat saapuneet kauppoihin.”

Järjestelmän hyödyt



HELSINKI UNIVERSITY OF TECHNOLOGY

Media Technology

- Ontologiaa voidaan hyödyntää hakujen saannin parantamiseen
 - esim. Iranista ja öljystä kertova artikkeli vastaa hakuun 'öljyvaltiot' vaikka sana ei edes esiintyisi
- Suosittele
 - "Nämä artikkelit vastasivat sanahakuun. Katso myös nämä, jotka sisältävät samoja käsitteitä."
- Käsitteistys antaa pohjan aineiston jatkohyödyntämiselle ontologisia menetelmiä hyödyntäen



- Mahdollisuus sanahakuun ja käsitehakuun, sekä näiden yhdistelmään
- Voidaan esimerkiksi toteuttaa perinteinen sanahaku ja lisätä sen rinnalle suosittelu käsitteillä
- Yksinkertainen hahmokieli klusterointiohjeiden luomiselle



- Toteutuksessa käytettiin Yleistä suomalaista ontologiaa
- Haku toteutettiin Lucenella
 - käytössä mm. Wikipediassa
- Testattiin European Language Resources Associationin CLEF Test Suitella
 - Tulokset positiivisia
- Projektin aikana valmistui diplomityö otsikolla Tekstiaineiston ontologiaperustainen indeksointi ja haku



Haku:

Ajalta -

Tulokset näytetään aikajärjestyksessä

Näytetään:

Koodi Paivamaara Otsikko Score

Näytetään myös:

pelkan sanahaun tulokset pelkan kasitehaun tulokset yhdistehaun tulokset suosittelun tulokset

Sanahaun tulokset:

[Bush: Irakin tiedustelutiedot virheellisiä \(2005-12-15\)](#)
[Bushin suosio nousi hieman Yhdysvalloissa \(2005-12-09\)](#)
[Rumsfeld: USA vähentää taistelu- joukkojaan Irakissa \(2005-12-24\)](#)
[Varapresidentti Cheney Irakissa yllätysvierailulla \(2005-12-19\)](#)
[Bushille voitto on ainoa vaihtoehto \(2005-12-01\)](#)
[Paine USA:n vetäytymiseksi Irakista kasvaa kotirintamalla \(2005-12-23\)](#)
[Pahin tappio olisi sisällissota \(2005-12-27\)](#)
[Bush taipui puheisiin Iranin kanssa \(2005-12-02\)](#)
[Morales toivoi poliittista sopua Boliviiaan \(2005-12-22\)](#)
[Bush: Irakissa yhä tavoitteena voitto \(2005-12-20\)](#)
[Bagdadista löytyi lisää pahoinpideltyjä vankeja \(2005-12-13\)](#)
[Nopeasti vahvistuva Kiina haastaa Yhdysvallat ainoana supervaltana \(2005-12-18\)](#)
[USA:n kannattaisi lähteä Irakista \(2005-12-31\)](#)
[Yhdysvalloilla olisi kiire voittaa \(2005-11-26\)](#)
[Bush pakotettiin nielemään kielto vankien kiduttamiselle \(2005-12-17\)](#)
[Dokumenttielokuva kertoo aina jostakin \(2005-12-03\)](#)
[Kidutus on kidutusta \(2005-12-20\)](#)
[Laillisia vai laittomia ratkaisuja? \(2005-12-17\) \(2005-11-26\)](#)
[USA:n johto antaa varovaisia vihjeitä Irakin-joukkojen vetämisestä \(2005-12-01\)](#)
[Kapinalliset ottivat kaupungin hetkeksi haltuunsa Irakissa \(2005-12-02\)](#)

Suosittelun tulokset:

[USA ja Irak tarvitsevat toisiaan \(2005-12-30\)](#)
[Sunnit voivat pelastaa Irakin vaalit \(2005-12-12\)](#)
[USA:n Irak-liittouma ohenee hiljalleen \(2005-12-31\)](#)
[USA valmistautuu pitkään läsnäoloon Irakissa \(2005-12-27\)](#)
[Kansalaiset, medborgare! \(2005-12-31\)](#)
[USA on hylännyt perusaatteenensa sekä kotimaassa että maailmalla \(2005-12-04\)](#)
[Francon aika on yhä arka aihe Espanjassa \(2005-12-28\)](#)
[Bush myönsi hyväksyneensä salakuuntelun Yhdysvalloissa \(2005-12-18\)](#)
[HJK:n perheen jäsenenä \(2005-12-17\)](#)
[Miksi presidentti on ylipääällikkö? \(2005-12-04\)](#)
[Miehiä romahduksen partaalla \(2005-12-18\)](#)

