



Kielitieteelliset aineistot

Antti Leino <antti.leino@cs.helsinki.fi>

Tietojenkäsittelytieteen laitos

13.9.2010



Siis mitä?

- Tietojenkäsittelytieteen syventävien opintojen seminaari
- Tiedon säilytystä, kun sovellusalana kielitiede
 - yleensä rakenteista
 - eri tavoin esikäsiteltyä
 - tekstimuotoista tietoa
 - sanakirjamaista tietoa
 - paikkatietoa



Taustaa

- Antti Leino
 - suomen kielen ja sovelletun tietojenkäsittelytieteen dosentti
 - päätyö Kotimaisten kielten tutkimuskeskuksessa
 - aiemmin mm. TKTL, HY:n atk-keskus
- Muut
 - kuka?
 - miksi täällä?
 - mahdollinen aiempi kosketus aihepiiriin?



Työmuodot

- Tapaamiset maanantaisin 16–18, C220
 - Alkupuolella esityksien hahmottelua
 - Loppupuolella esitykset
- Itsenäistä työskentelyä
 - Kirjallinen esitys n. 10 s
 - Suullinen esitys n. 25 min
 - Keskustelua esityksistä



Aikataulu

- 20.9. aiheen valintaa ja hahmottelua
 - Tavoite: karkea aihe sovittuna
- 27.9.–11.10. Aiheen tarkentamista
 - Joka kerralla lyhyt katsaus, missä kukin on menossa
 - Tavoite: periodin loppuun mennessä kirjoitustyö hyvällä alulla
- 1.–29.11. Esitykset
 - Kullakin kerralla kaksi esitystä
 - Itse esitykset á n. 25 min, loppu keskustelua



Mahdollisia aiheita

- Valmiita otsikoita ei tarjolla
- Muutamia karkeita aihepiirejä
 - Aineistojen kokoaminen
 - Mistä, miten, mihin käyttöön?
 - Rakenteen suunnittelu
 - XML – mutta miksi ja miten?
 - Aineistojen hallinta ja käyttö
 - säilytys, tarjollepano, hakutyökalut
- Tarkka aihe täsmentyy seuraavien viikkojen aikana



Mahdollisia aiheita

Aineistojen kokoaminen

- Painopisteenä tekstimuotoiset aineistot
- Mistä tekstiä löytyy?
 - verkosta? muualta?
 - millaista kieltä mistäkin löytynyt teksti edustaa?
- Miten teksteistä koostetaan tutkimusaineisto?
 - verkosta louhitun tekstin esikäsittely
- Käyttö- ja tekijänoikeudet?



Mahdollisia aiheita

Rakenteen suunnittelu

- Itse aineisto vai metatiedot
- Aineiston rakenne
 - teksti
 - sanakirja
 - maantieteellinen vaihtelu
- Olemassaolevat standardit
 - TEI, Dublin Core



Mahdollisia aiheita

Aineistojen hallinta ja käyttö

- Tekstiaineistojen käsittely
 - jäsenitys, haku, ...
- Paikkatiedon käsittely
 - kielen ilmiöiden maantieteellinen vaihtelu
- Aineistojen pitkäaikaistalletus
- Aineistojen tarjoaminen tutkijoille
 - verkon kautta koko aineisto / vain metatiedot
 - käyttöoikeuksien hallinta
- Aineistopohjainen kielentutkimus



Ja töihin!

- Kurssisivulla artikkeleita
 - paperilla selattavissa täällä
- Lisää löytyy
 - Tiki-kurssin opit tarpeen
- Kysyä saa sekä nyt että myöhemmin