



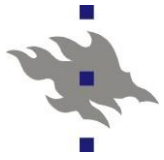
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

OMorFi

Avoimen lähdekoodin sanakirjat ja sovellukset

Krister Lindén / Kieliteknologia

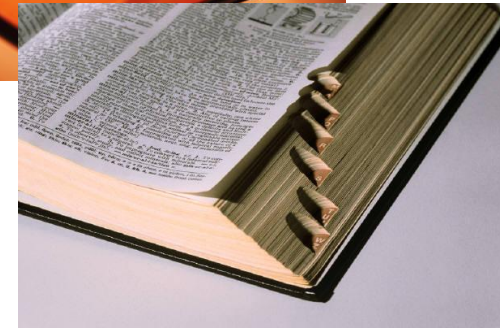
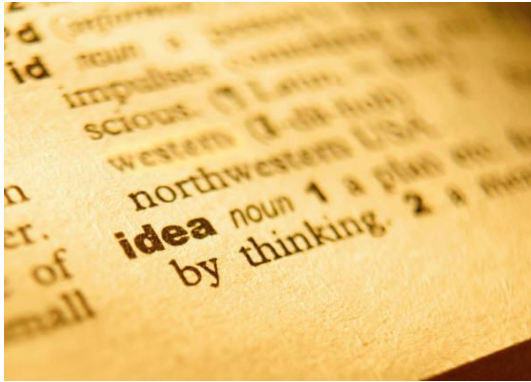
5.12.2007



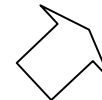
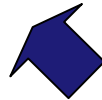
Toimintaympäristö

Muut toimittajat ja käyttäjät

OMorFi



Muu maailma





Sanakirjojen toimittajat ja käyttötarkoitukset

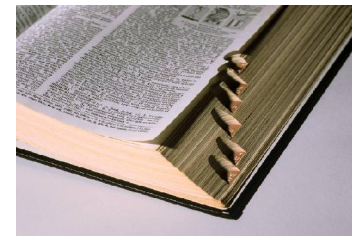
■ Toimitus

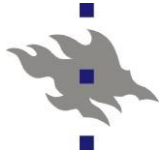
- Kotus
- Voikko / COSS, ISpell / Gnu
- kustantajat (WSOY, ...)
- käännöstoimistot
- yritysterminologit (Semantic Web)
- kieliteollisuus (Lingsoft, Connexor, Kielikone, BitLips, ...)



■ Käyttö

- oikoluku
- tiedonhaku
- kääntäminen
- puhesynteesi ja -tunnistus
- syntaktinen analyysi, ...

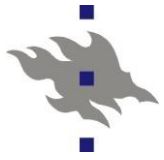




Muu maailma tarjoaa havaintoja sanojen käytöstä

- Kielipankki / CSC
- CLARIN / EU
- internet / WWW
- YLEn puhearkisto
- ...



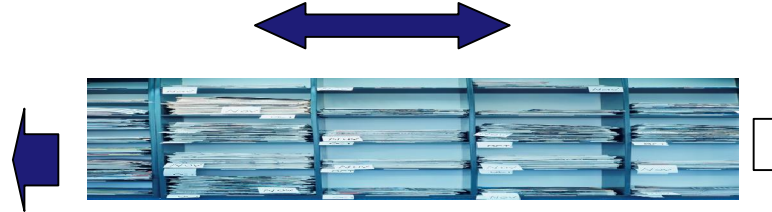


Perinteisesti myös kieliteknologian sanakirjat on rakennettu suljetussa ympäristössä

Mitä



Miten



Miksi



morfologinen analyysi:
sanalista + taivutus

leksikografi

oikeinkirjoitus / Lingsoft

syntaktinen analyysi:
sanaluokat + säännöt

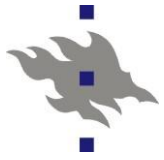
lingvisti

syntaksi / Connexor

käännössanakirja:
käännös + konteksti

kääntäjä

lukemisapu / Kielikone

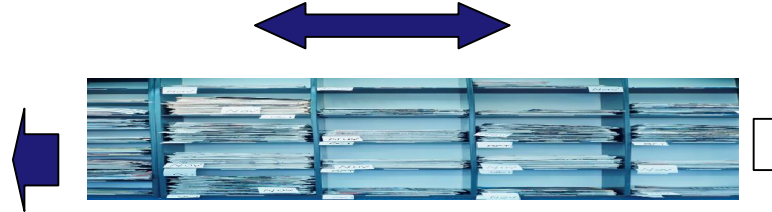


Nykytavoitteet ovat avoimien lähteiden yhteensovittamisessa ja laajentamisessa

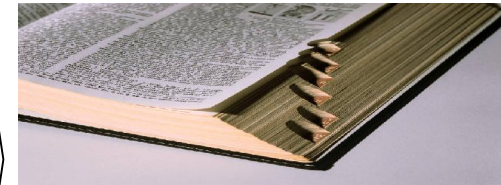
Mitä



Miten



Miksi



1 morfologia

automaattisesti

(uusi) sanalista / Kotus

2 morfologia

automaattisesti

oikoluku / Voikko

3 morfosyntaksi

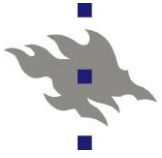
puoliautomaattisesti

morfosyntaktinen
analyysi

...

5.12.2007

Krister Lindén / Kieliteknologia



Myös sanakirjatyökaluista tehdään avoimia

