

Keeleveeb

Heiki-Jaan Kaalep

Filosoft

Kus, mis, kes?

- <http://www.keeleveeb.ee> ,
<http://keeleveeb.edu.ee>
- Täitjad:
 - Andrei Dementjev, Heiki-Jaan Kaalep, Rene Prillop, Katrin Tsepelina, Tarmo Vaino
 - TÜ tudengid

Ideoloogia

- Mõeldud eesti keele jaoks
- Üks koht, kust alustada (portaal)
- Erialsõnastikud (=ainesõnastikud, oskussõnastikud) on kah vajalikud
- Ühispäring

Ühispäring

- erinevaid sõnastikke ei panda kokku
 - võimaldab ühendada rohkem sõnastikke (ei pea midagi endale tõmbama)
 - autoriõigused
 - versioonide uuendamine
 - võimaldab valida valdkonda ja kvaliteeti
 - nt. “aken” ehituses ja IT-s
 - eri kasutajad, eri nõudmised (õpilane vs spetsialist)
 - vastuste hulk on piiratud (vähem kerimist ja klikkimist)

Ühispäring (järg)

- kergendab arendustööd
 - uute sõnastike lisamise lihtsus
 - sisu esitatakse autori esitusele lähedaselt
(eri sõnastikes on eri liiki infot)

aga...

- sõltub sõnastikutegija suvast

Uued sõnastikud 2006-2007

- 25 tk
- teisenduskaik
 - el. versioon:
 - algfail -> TBX -> HTML
 - paberversioon:
 - digitaliseerimine -> parandamine -> (TBX) -> HTML
- esitus nagu keeleveebi-välistel sõnastikel
- päring leiab ka käändevormis olevad sõnad (nt. väljendites)

Keeleveeb ja TÜ

- EuroTermBank (2005-2006)
 - ülevaade olemasolevatest terminisõnastikest (ETERi andmebaas)
 - läbirääkimised autoritega
 - digitaliseerimine (Tilde)
 - TBX-sse (osa)
- koolisõnastikud (PHARE) (2004-2005)

TBX

- Term Base eXchange
 - www.lisa.org standard
 - terminisõnastike vahetuse formaat
- sobib, kui sõnastikus on terminid mõistete kaupa rühmitatud
- ei sobi, kui tähendused on sõnade järgi rühmitatud (nt. Tehnikasõnastik)

Teisendamine

(e. mis see ka maksab?)

- sõnastiku sisuline struktuur ja vormiline esitus on põimunud ning mõjutatud otsimisprobleemist
- TBX on struktuuri esitamiseks (nii sisuline kui formaalne kontroll)
- autori omapärad vs vead
- sõnastiku (paber)esitusest tulenev liiasus
- esitus ekraanil \neq esitus paberil, aga see võib olla sarnane

Korpusepäring

- soov: tasakaal kasutusmugavuse ja võimaluste vahel
- paremad võimalused kui tavaline otsing, kuid kehvemad kui UNIXi käskudega saaks teha
- korpus indekseerida
- indeksiks sõnavorm, lemma, grammatiline info
- päringus võimalik neid omavahel kombineerida ja seda ka mitme sõna puhul
- tulemus esitatakse KWIC-indeksina

Aitäh!