

Elektroniliste teatmeteoste kasutajasõbralikud päringusüsteemid

Jaak Vilo



1997-2005

2000a

2005:

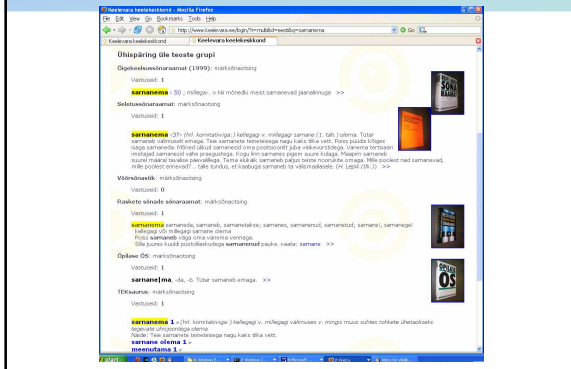
üks miljon nelisada viiskümmend viis tuhat üheksasada kaheksakümmend kaks avalike külastust alates 27. augustist 1998

2005+

Eesmärgid

- I sõnastike ja teiste teatmeteoste struktuurne märgendamine ja ettevalmistamine elektroonilise andmebaasi ning info-otsingute jaoks.
- II lihtsate ja mugavate päringusüsteemide välja töötamine ja arendamine.
- III elektrooniliste sõnastike kasutuse populariseerimine

Struktuur

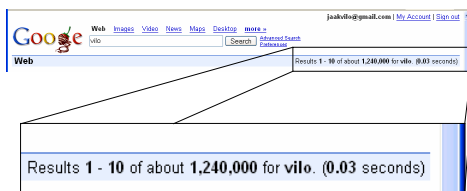


Struktuur

- sõnastikel on väga erinevad ülesehituse ja formaadi printsiibid
- Kasutaja "tarbib HTML-i"
- Otsimootor peaks suutma pakkuda tuge eri formaadis teoste kasutamiseks
- St iga indekseeritud sõna on oma "kontekstiga" ja päringus peaks seda saama otsida

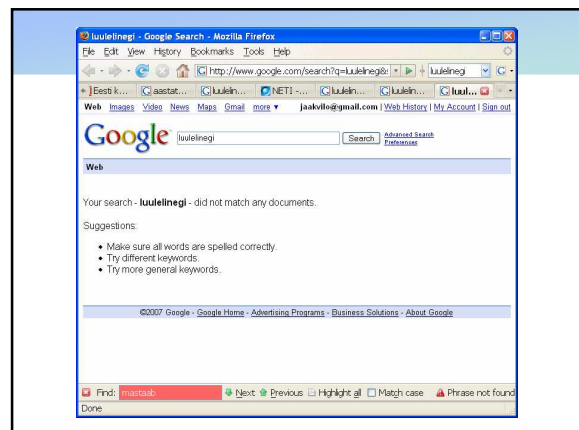
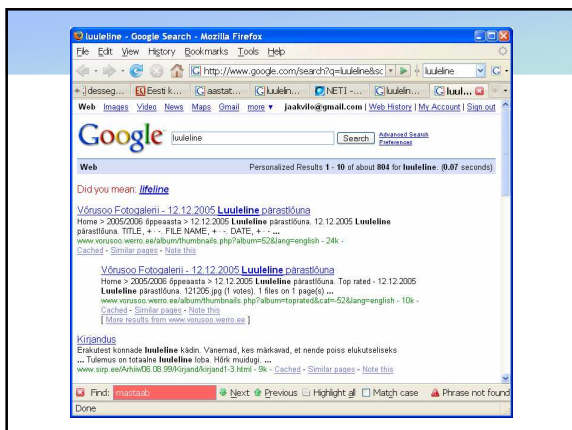
Info otsimine

- Dokumendid milles esineb sõna, sõnad, fraas, ...

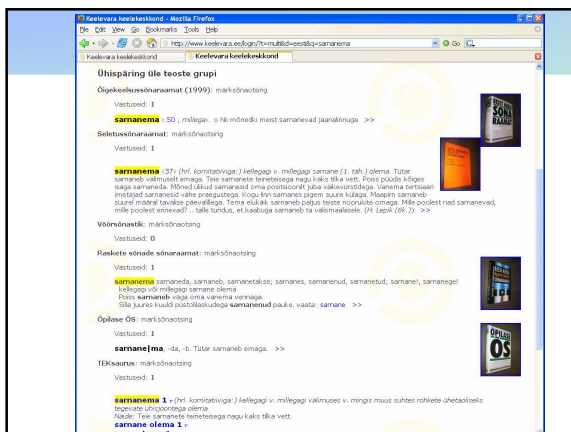
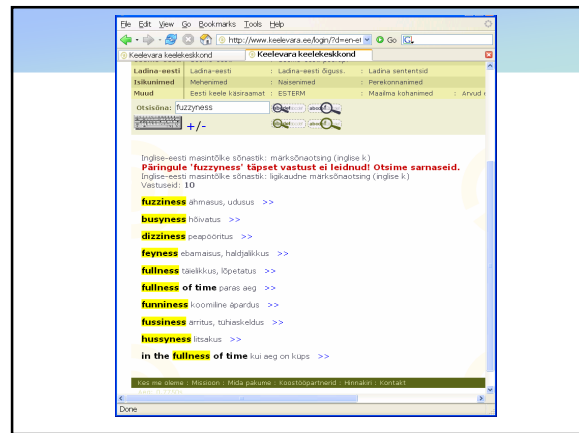
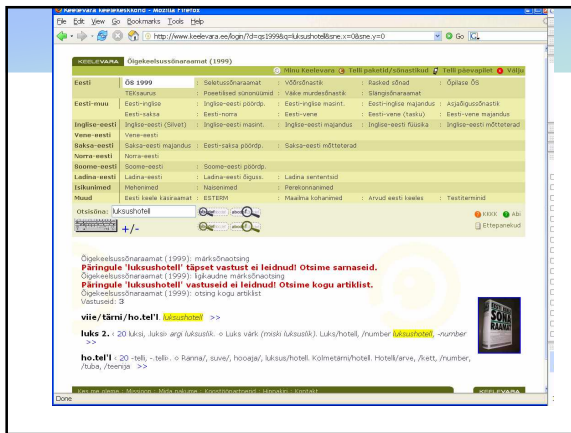
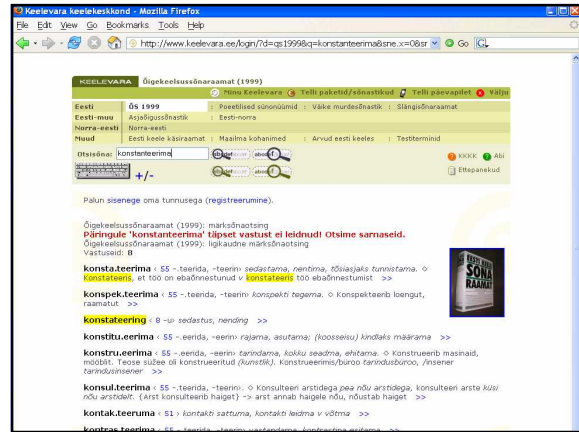
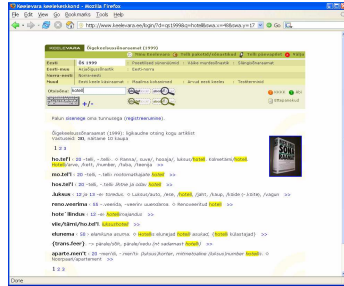


Kasutaja jaoks

- Kuidas lihtsalt ja täpselt esitada oma päring
- Kuidas saada relevantseid tulemusi?
- Kõrgem loogika:
 - Kus on homme lõunal odavam tume õlu?
 - Millised maksureeglid kehtisid 5 aastat tagasi?



Kogu artikkel täpselt/ligikaudselt



OpenOffice.org writer

Format Table Tools Window Help

Times New Roman 12 B U

http://www.posti.ee
http://www.presid.ee

Mu kallid kaasmaekestsellentsid, austatavad daamid ja härrad.

Morph search
Find: sünd Next

1. märtsi 89. sünnipäeva Tartusse, kus tänavu pühitseme ka kui rüürik kaib käsikäes vaimlise ajakausisega; ulikooli olemas ei. Esimene oli eelduseks teise sünnile. Ilma eesti demokraatlikust rahvusriigist jaama nimetuks tungiks.

Variandid:

- Morfoloogiline analüüs
 - Analüüs igale tekstile kui ka päringu sõnale
 - Sõnade sobitamine ainult põhivormis
- Morfoloogiline süntees
 - Sünteesida kõik võimalikud sõnavormid
 - otsida kõiki neid korraga

Kristo Tammeoja:

OpenOffice.org writer

Format Table Tools Window Help

Times New Roman 12 B U

http://www.posti.ee
http://www.presid.ee

Mu kallid kaasmaekestsellentsid, austatavad daamid ja härrad.

Morph search
Find: sünd Next

1 viga

1

2

C ↔ C[ieua]
d → t, j, 0
nd → nn

1. märtsi 89. sünnipäeva Tartusse, kus tänavu pühitseme ka kui rüürik kaib käsikäes vaimlise ajakausisega; ulikooli olemas ei. Esimene oli eelduseks teise sünnile. Ilma eesti demokraatlikust rahvusriigist jaama nimetuks tungiks.

Morfoloogiline otsing...

- Käändelõpud
- Lõpuvaheldus (9)
 - [lrs][aeiou] ↔ [aeiou][lrs]
 - manner : mandri
- Astmevaheldus (20)
 - nn → nd manner : mandri
- Vokaaliteisendus sõna "tüves" (3)
 - i → [ie] ri+da : re+a

Morfoloogiline otsing...

sõna lõpp → MANNER

lõpuvaheldus → <ma>NNER

astmevaheldus → <ma>(NNER | NN[lrs][aeiou] | NNER[ieua])

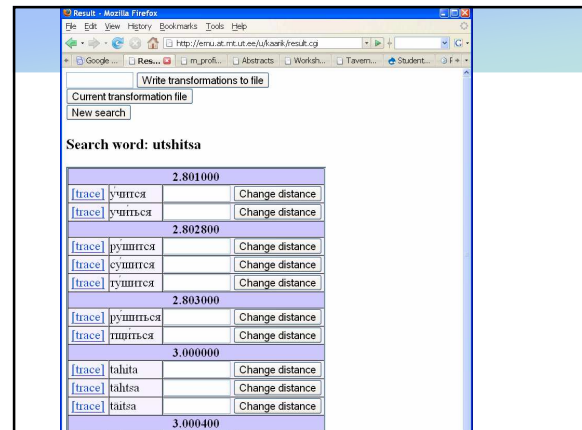
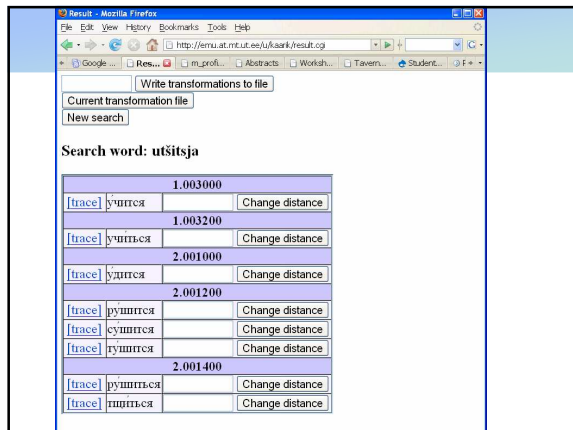
<ma>((nn|nd)er | (nn|nd)[lrs][aeiou] | (nn|nd)[ieua])

Morfoloogiline otsing...

<ma>((nn|nd)er | (nn|nd)[lrs][aeiou] | (nn|nd)[ieua])

(t|d|da|sid|id | (te|de|e|i)?(sse|s|st|...)?)?

Regulaaravaldise ligikaudne sobitamine!



- pishu - tulemusteks пишу, aga esimeste sõnade seas on ka nt. писать ja писали
- pisat - tulemuseks писать, aga leiab ka писали
- problemu - проблему, leiab ka sõna проблем lähedaseks.
- možno - leiab можно, aga ka можем ja может
- znakomilsja - знакомился aga ka знакомить ja знакомый
- svoei - esimesteks tulemuse sõnadeks - своей, свой, звою, своих

- ## Täpne ja ligikaudne otsimine
- Täpited (vs tapited vs ta"pita"hed)
 - Aga kas meie oskame teistes keeltes?
 - ~ ~ ~ ~ ~ \ _ c ~ ~ ~ ~ ~
 - Hélène
 - Susisevad – seniajani ei saa alati nende sisestamise või näitamisega hakkama
 - “kavalam” otsimine
 - Dush, dušš, dushsh ?
 - Gorbatošov, Gorbatošov, Горбачов, Gorbachev
 - režim, rezhiim, riim

- Arendusversioon:
 - <http://emu.at.mt.ut.ee/u/kaarik/>
 - praegu parooli taga
 - oleme valmis partnerlusteks

- ## Inimesed
- Maia-Liisa Natus; Leive Turi
 - Reina Käärik – üldistatud teisenduskaugus
 - Kristo Tammeoja – ligikaudne reg. avaldis
 - otsime tegijaid ☺
-