

Is Query Translation a Distinct Task from Search?



Daniela Petrelli

George Demetriou

Patrick Herring

Micheline Beaulieu

Mark Sanderson

University of Sheffield - UK

Where the idea comes from



- CLI R literature suggests:
 - Query translation checked by the user
 - Search with the updated-translated query
- Clarity user study with potential users suggests the opposite:
 - Users do not want to control the system
 - Users want to see relevant document

The System Test-bed



- Architecture:
 - User interface + correlated services in UK
 - Translation + search in Finland
 - Communication via SOAP
- Interface:
 - Clarity interface updated for iCLEF: title translation, keywords, proper names added

Hypothesis



Users will perform better and like most the layout with hidden translation

Sub-hypothesis:

- CLI R is good enough to support “delegation”
- Non speakers not necessary prefer control

Netscape: CLARITY - user testing

File Edit View Go Communicator Help

clarity search

Search in: English Finnish

For documents in: English Finnish

Date range: last two weeks Max documents: 10

racism bigotry immigration religion intolerance

Translate Search

Translated terms:

racism: rasismi: (racism)
 racism: rotusorto: (racial discrimination, racism)
 bigotry: fanaattisuus: (fanaticism, zealotry, bigotry)
 bigotry: kiihkoilu: (fanaticism)
 immigration: maahanmuutto: (immigration)
 immigration: siirtolaisuus: (migration, emigration, immigration)
 immigration: passintarkastus: (passport control)
 religion: uskonto: (religion)
 religion: uskontokunta: (religion, denomination)
 religion: oppi: (doctrine, dogma, theory, tenet, learning, erudition, scholarship, e...
 religion: usko: (belief, faith)
 intolerance: intoleranssi: (no translation)
 intolerance: sietokyvyttömyys: (no translation)
 intolerance: suvaitsemattomuus: (intolerance, narrow-mindedness)
 campaign: kampanja: (campaign)
 campaign: tempaus: (wrench, pull, snatch, move, campaign, masterstroke)
 campaign: sotaretki: (campaign, military expedition)
 campaign: sotatoimi: (operation, act of war)
 campaign: kampanjoita: (campaign)

list report overview
 sorted by rank, sort by language, sort by date

Suvaitaan toisiamme
 Suvaitaan (stand, put up, tolerate, condone, condescend) toisiamme (second, another, other, different, one, the others, other people, some, one ... the other, one another)
 #1, #, 2002.08.02, 332 words long, search terms: racism bigotry immigration religion intolerance campaign, terms found: campaign, racism
 proper names: Yksi Jehanesson Jari Nasa Baresi Tilaisuus Matta Quark Desailly Marcel Tanam Tai Franco Amnesty-talossa TUOMI Toimittaja Al Ajaxin Tarzan-jutut Seminaari SYSTEM Kello Litmanen Matta Ajax LIVES RAIMO Eline Rotusortom Peik Markku Teatterimestussa Iita
 keywords: viikko (week) loppu (end, ending, rest) ettei ilta (evening, night) kyyms

Netscape: CLARITY - user testing

File Edit View Go Communicator Help

clarity search

Search in: English Finnish

For documents in: English Finnish

Date range: last two weeks Max documents: 10

genetics

see query translated

list report overview
 sorted by rank, sort by language, sort by date

Eturushassyöä voi myös periytyä
 Eturushassyöä voi (butter)(can, be able to, be capable of, may, be, feel) myös periytyä
 #1, #, 2002.08.02, 179 words long, search terms: genetics, terms found:
 proper names: Eturushassyöän Yksikko Periytyvän Siten Finn-Mediin LIVES Professori Quark Tamperelainen SYSTEM Eturushassyöasta Taysin Syöägenetiikan Jorma Isola
 keywords: voi (butter)(can, be able to, be capable of, may, be, feel) kasvaa (grow, grow up, produce, enlarge, expand, become larger, develop, mature, ripen, increase, grow, grown) esimerkiksi (example, instance, model, pattern, illustration, exemplification) kahden (two) yksikko

Väitökset Kasvatustieteiden lisensiaatti Seija Kojonkoski-Rännin käsityön käsitteen merkityssisällön koskeva väitöskirja tarkastetaan tänään Turun yliopistossa
 Väitökset Kasvatustieteiden (kaat, kietoa, wind, bind, wasp, kiinnittää, fasten, attach, fix, stick, make fast, kravatti, liittää, join, connect, side, bond, sitoa, bind up, tie up, tie down, solmia, knot, solmia, necktie, road, way, path, yhdysside, link)(science)growing, cultivation, breeding, raising, education, upbringing)(pedagogics) lisensiaatti (licentiate) Seija Kojonkoski-Rännin käsityön käsitteen merkityssisällön koskeva (touch, hurt, ache, smart, concern, relate to, have to do with, apply to) väitöskirja tarkastetaan (check, examine, control, revise, audit, make sure, inspect, review) tänään Turun yliopistossa. (college, institute, training centre, training center)(over, more than, past, across)(university)
 #2, #, 2002.08.02, 108 words long, search terms: genetics, terms found: genetics
 proper names: Kari Quark Kojonkoski-Rännin HARKOTE Jorma Tiina Seija Beyreuther Konrad Aaltola Linnea Paunio Palo Niinistö VÄITÖKSET Lindfors Aeyloidoasi SYSTEM
 keywords: tyttö pallo (fire, burning) kuuluu (bome)(moon, month) (have to, be heard, belong to, be the property of, be owned by, be a member of, be associated with, be part of, be) professori (professor) kari (rock, trap, pitfall)

Participants and procedure



- 6 people participated, only data from 4 used
- English speakers, no Finnish
- All positive toward computer and I R
- iCLEF conditions
- Topics described in scenarios
- Parallel use of the system
- Users required to “keep” documents and judge as “highly”, “somewhat”, “non” relevant

Result and discussion

- Official result:
Tran – F = 0.061 NoTran – F = 0.081
- BUT...
 - Only “high relevance” computed officially
 - 16 tasks but 4 valid (3 NoTran, 1 Tran)
 - Only 2 topics covered
 - 1 user did 2, 2 users did 1, 1 did nothing

Result and discussion (cont.)



- Relevant documents non retrieved:
 - technical problems
 - few relevant documents (11 – 17) with low ranking
 - translation with all senses (e.g. green = vegetable)
- User personality (e.g. Bobby Sand for hunger strikes)
- Questionnaires says:
 - None mentioned Tran as preferred or easier
 - Complains about figurative senses of Tran
 - Slowness a general big drawback

Lessons learnt



- Run live is risky but rewarding:
 - Bugs and weakness (slowness) revealed
 - New user requirement discovered (use of "+" and "-", acronyms, proper names)
- Insight on cognitive mechanisms:
 - User expect a result list, not a translation
 - See translation push refining immediately
 - Specific words might be less effective
- Limitation of the current interface design

Can we fix it? Yes, we can!



- Architecture fixed
- Interfaces updated
- Users query log will be used to re-run the experiment and collect reliable data
- Training will be considered