Automating the Analysis of Open Question Responses: Faster, More Accurate, More Consistent and in Every Language

Jussi Karlgren
An efficient way to analyze answers to open-ended survey questions using language technology
gavagai

- text analysis at large scale since 2008
- media monitoring, questionnaire processing
- customers: gov't agencies, finance, market analysis, security
- based in Stockholm
- builds on decades of academic language technology research

jussi karlgren

- cofounder, adjoint professor of language technology
Open-ended survey answers?

- expressive replies
- deep insights
- finegrained results
- unconstrained by questionnaire

Tell us about what you like to drink and when:

"Well, in the mornings I usually like to have a cup of coffee because it doesn’t give me heartburn but after a little something just to clean my palate. My wife...

- expensive to analyse and code
- does not scale

we have technology for this!
gavagai's semantic base technology
we learn from usage to build a lexicon

the weather is great in barcelona
the weather is gray in stockholm
the climate is sunny in nice
the weather is chilly in helsinki
the weather is cold in moscow
the weather is nice in hong kong
the weather in barcelona is pleasant

... the climate is passable in nice
the weather is nippy in moscow
the weather in barcelona is balmy
... the climate is chilly at the office
the weather is great in barcelona
the weather is gray in stockholm
the **climate** is sunny in nice
the weather is chilly in helsinki
the weather is cold in moscow
the weather is nice in hong kong
the weather in barcelona is **pleasant**
...
gavagai explorer

highly interactive
rapid analysis
language magic in the background
(no need to be a wizard!)
1. upload texts
2. inspect clustered utterances
3. define theme folders
4. join clusters
5. accept synonym suggestions and merge
6. discard irrelevant terms
7. repeat until happy
8. deliver report

csv or pdf
Heen had a kewl day morph?

Greeting

i came home 2 am ..and went up at 9

sum bastard woke me up @ 2pm

=P

heh naah.

Irritation i h8 dat guy

lol

phoner on the sms?

:)

whutta dowk

hewwo?

no worries

many markets, many demographics, many languages
main points

human intelligence remains in the loop
results debuggable and inspectable
highly interactive
consistent coding
invisible background technology
immediate refinement of responses
convenient customer reports

An efficient way to analyze answers to open-ended survey questions using language technology
advantages of gavagai's text analysis

picks out salient multiword expressions
understands and suggests synonyms and related terms
automatically learns any language
faster and more consistent results

An efficient way to analyze answers to open-ended survey questions using language technology