

information access systems - beyond one shot

jussi karlgren

gavagai & kth

february 12, 2016

take home message

(from last time!)

love evaluation and systematic testing

(the thing to do, whatever you do)

understand precision and recall

know of various measures based on p & r

beware of perils of averages

crucial and central target notion of "relevance"

challenges to "relevance"

what is quality in an information system?

usefulness and effectiveness for task

appealing presentation

authority and trustworthiness and sourceability

relevance and truthfulness

reusability and cost

happiness, trust, and satisfaction!

the target concept of relevance

in everyday language:

a function of task, collection characteristics, user preferences and background, situation, tool, temporal constraints, and untold other factors

in information retrieval research:

a (binary) relation between query and document, disregarding everything contextual

how did your assessment task go?

back to usefulness for task

happiness, trust, and satisfaction!

so who is our user here? and why?

some e how do you measure
we happiness, trust, and satisfaction
bibliographic for these tasks?
patent search

monitoring for copyright violations

question answering

item search

entity search

information extraction

sentiment analysis

topic detection and tracking

plagiarism detection

image search for illustration of text, or for entertainment

video search

music search

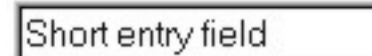
modelling usage in adhoc search is a challenge:

1.87 wds / q

but we have sessions!

and logged-in users' profiles!

design matters for average query length

A large, rectangular text input field with a thin border. It contains the text "This is the long entry field" in a monospaced font. The field has a scrollbar on the right side, indicating it is scrollable.A small, rectangular text input field with a thin border. It contains the text "Short entry field" in a monospaced font.

Short entry field	2,81
Long entry field	3,43

log analysis

dwel time

click thru

returning visits

conversion rate

user actions

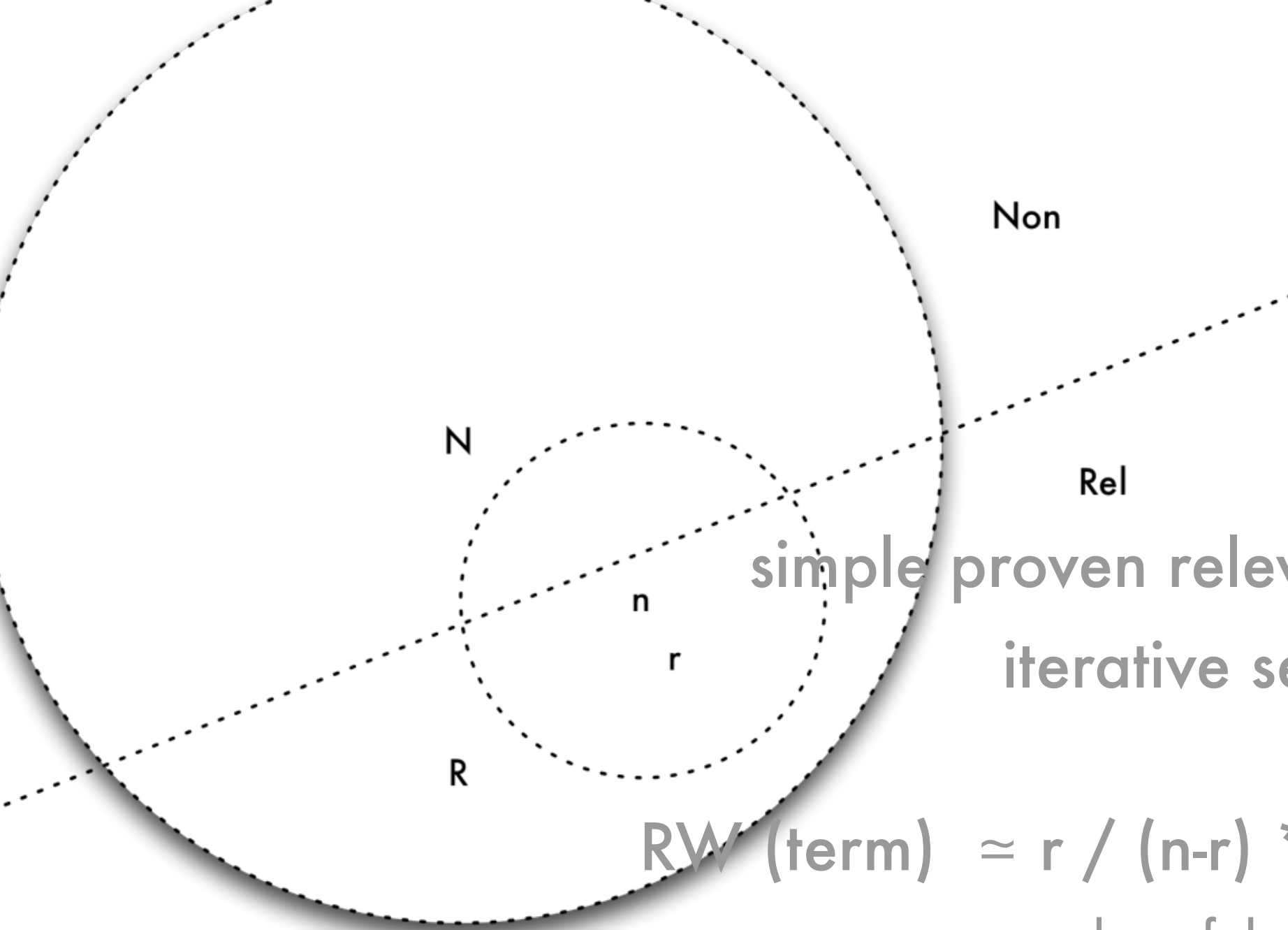
views

what is a successful site? many visitors? long dwell time?

best engagement?

studiotraffic.com	Apr 1, 2006 14:39	1	http://www.studiotraffic.com
studiotraffic admin contact	Apr 2, 2006 9:19		
johndeere.com	Apr 2, 2006 11:12	1	http://www.deere.com
gurneys	Apr 3, 2006 15:31		
studio traffic	Apr 6, 2006 19:51	2	http://www.studiotraffic.org
studio traffic	Apr 6, 2006 19:51	3	http://www.scam.com
studio traffic	Apr 6, 2006 19:51	4	http://www.scam.com
what happened to studio traffic	Apr 6, 2006 19:57	1	http://www.talkgold.com
fort worth tx irving tx area hotels	Apr 26, 2006 11:40		
fort worth tx irving tx hotels	Apr 26, 2006 12:22		
buy cheap cds	Apr 28, 2006 22:36	1	http://www.cheap-cds.com
what is trimpatch 24-7	Apr 29, 2006 18:38	1	http://www.bodyandmindshop.c
what is trimpatch 24-7	Apr 29, 2006 18:38	3	http://www.bodytherm.com
what is trimpatch 24-7	Apr 29, 2006 18:38	4	http://www.bodytherm.com
diet patch reviews	Apr 29, 2006 18:48	1	http://www.ultimatefatburner.co
diet patch reviews	Apr 29, 2006 18:48	4	http://www.reviewcentre.com
diet patch reviews	Apr 29, 2006 18:48	6	http://www.dietfraud.com
diet patch reviews	Apr 29, 2006 18:48	7	http://www.gamer-talk.net
diet patch reviews	Apr 29, 2006 18:48	9	http://www.nextag.com
treatment for double vision	Apr 30, 2006 9:40	4	http://www.medical-library.org
once a day diet pill	Apr 30, 2006 9:47	5	http://buyinnovations.com

relevance feedback - beyond the one shot



simple proven relevance feedback
iterative sessions

$$RW(\text{term}) \cong r / (n-r) * ((N-n)-(R-r))/(R-r))$$

n = number of documents with term

N = number of documents

r = known relevant documents for term

R = known relevant documents for a query

positive feedback generally more valuable
implicit \approx negative if not clicked through

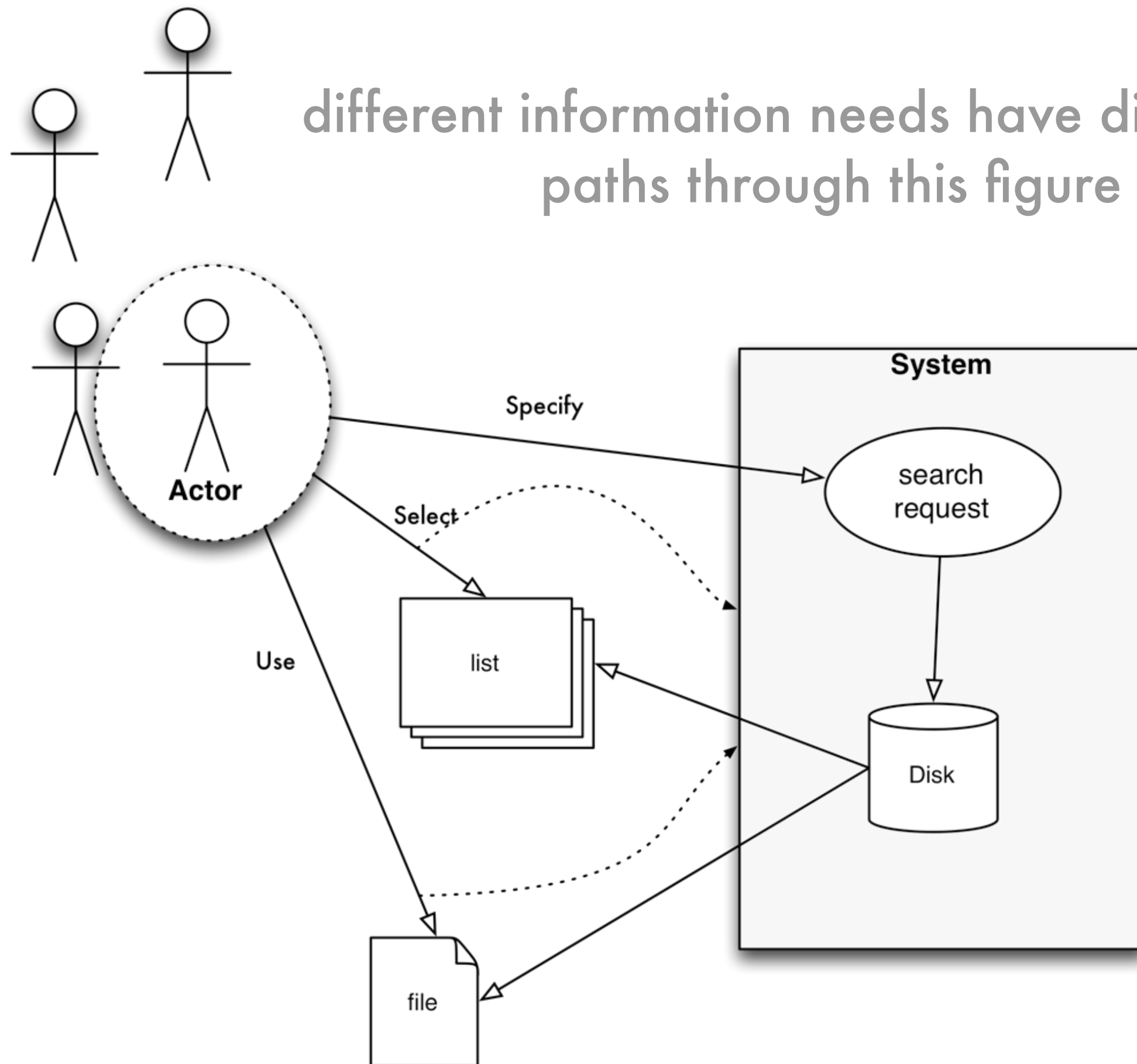
given relevance feedback, how to use it?

query expansion

interactive

automatic

different information needs have different
paths through this figure



use cases

differential needs

and thus differential eval!

evaluation = benchmarking (what we talk about here) + validation (matching to use case)

best practice: satisficing vs optimisation

(in practical system design, optimising the search component by benchmark may not be the best use of budget and time - appropriate validation steps in evaluation will provide the balance to make this decision)

take home message

information retrieval is a component
sessions are more informative than one-shot
relevance feedback is potentially useful
log analysis can yield relevance feedback
different use cases have different needs
... and should be evaluated differently
top line is not always best practice

<http://lexicon.gavagai.se>

Gavagai



- how to evaluate?
- benchmarking?
- validation?

Liggande logotyp

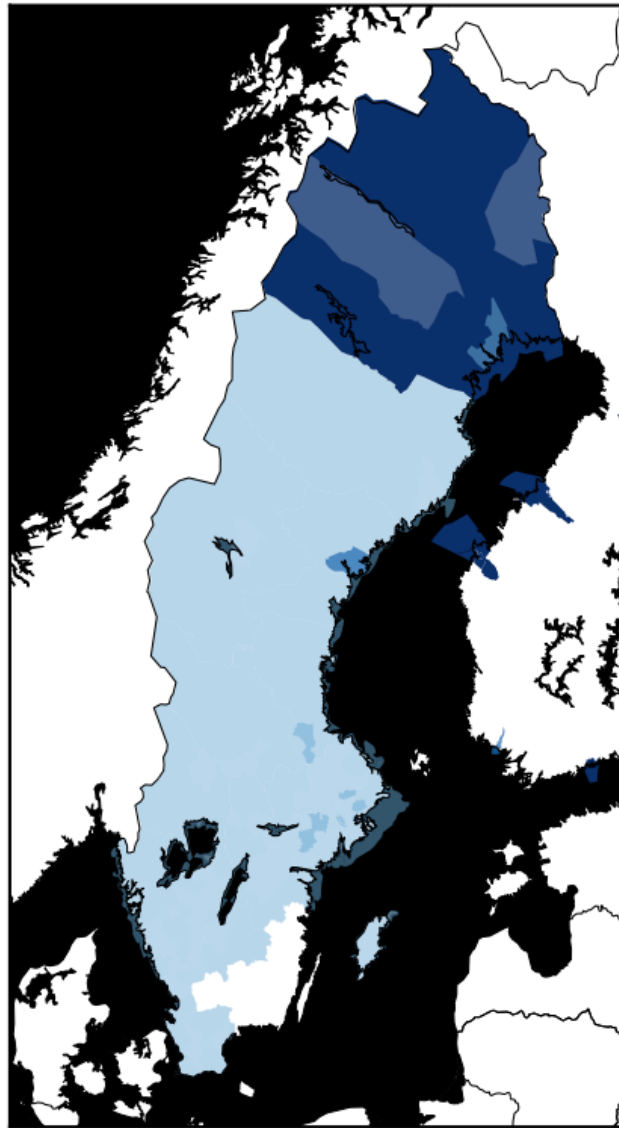
Gavagai



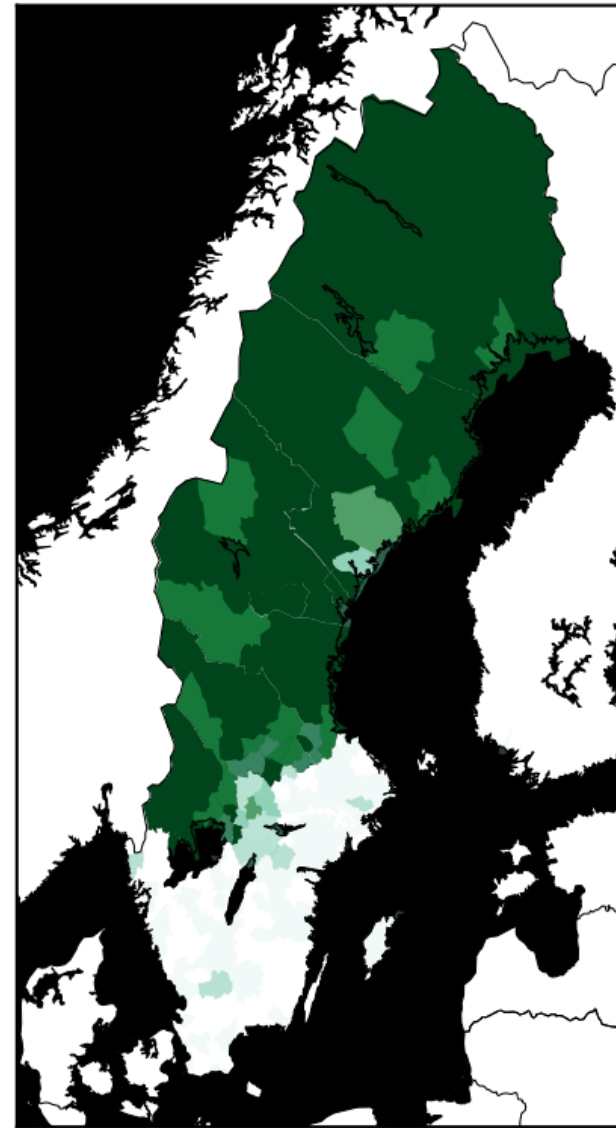
nästkusin - hits: 959



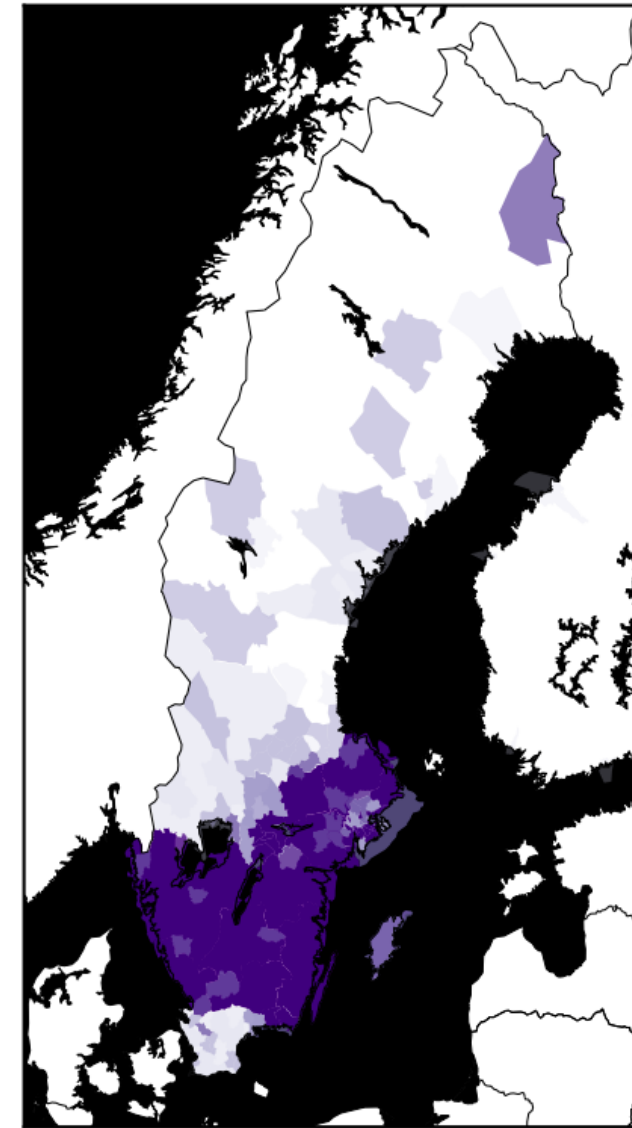
småkusin - hits: 678



tremänning - hits: 1717



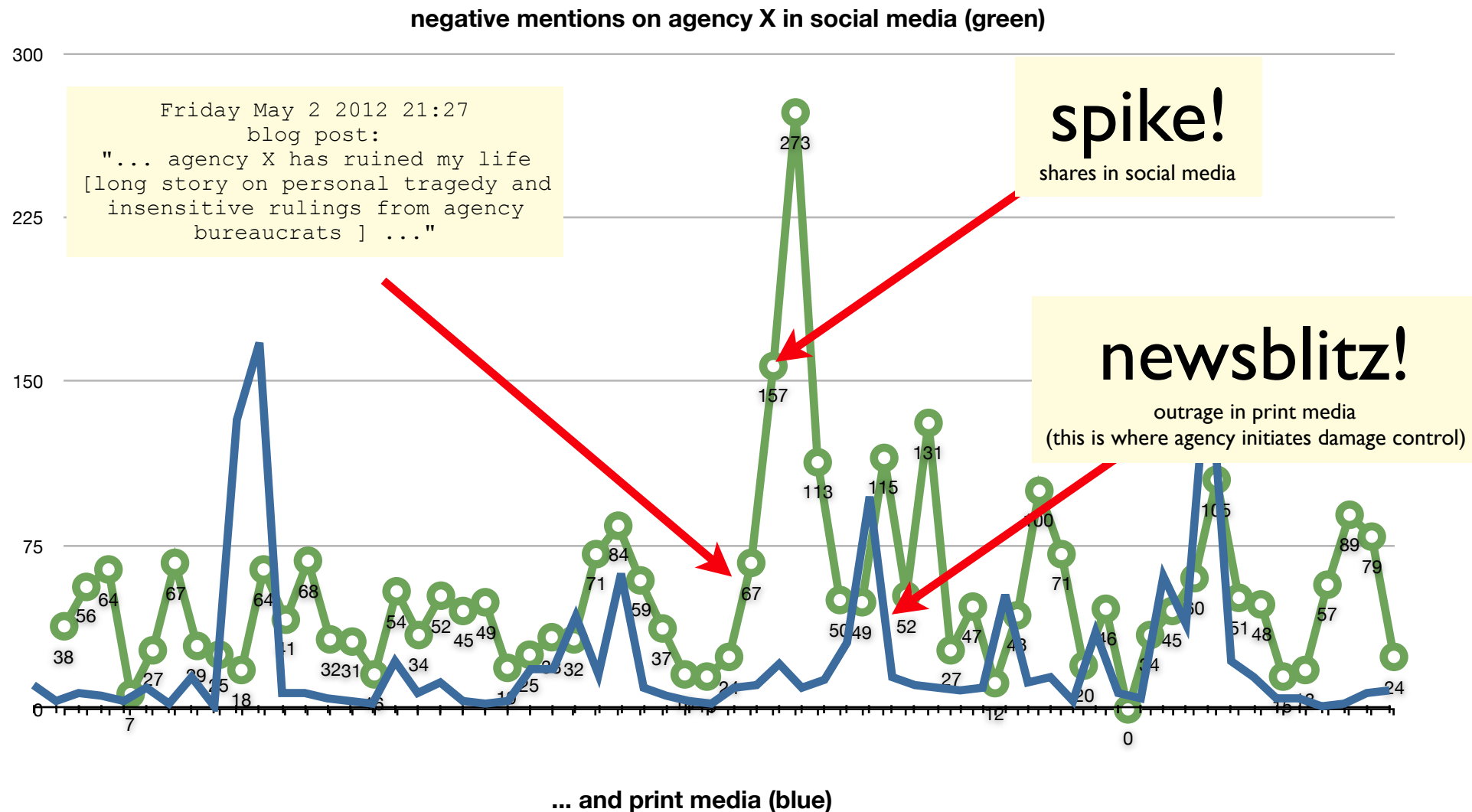
syssling - hits: 7204



Gavagai



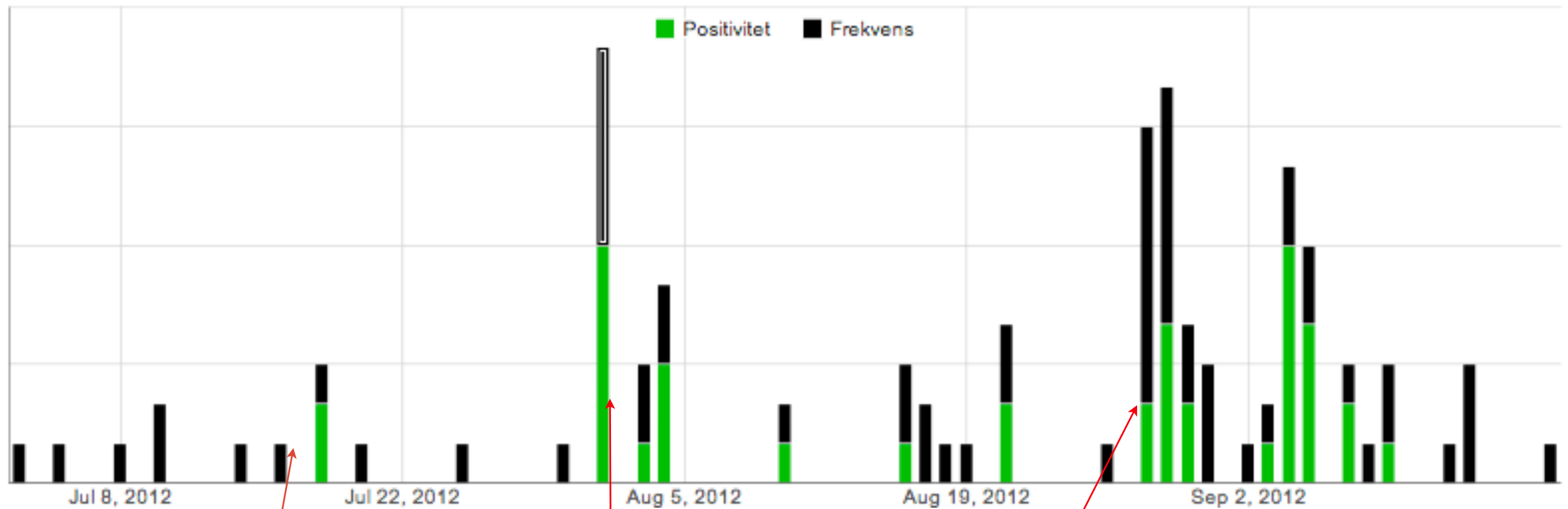
early warning from social media



Gavagai



evaluating a marketing campaign



old school marketing campaign

freebie with a periodical

gift in a goodie bag on a cruise

Gavagai



Evaluating Learning Models

evaluate adaptivity?

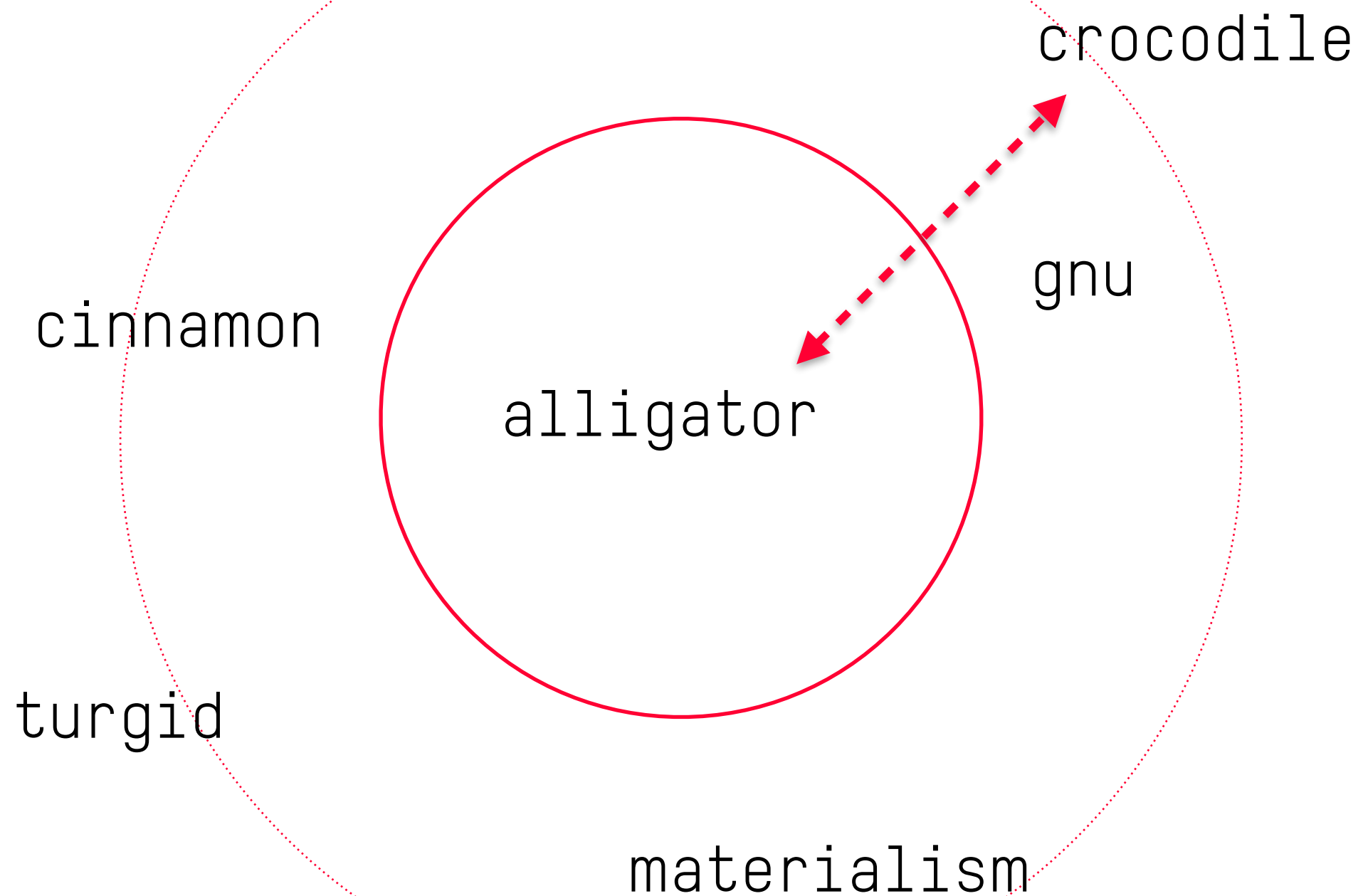
test process, not outcome.

test robustness in face of variation

Gavagai



relative position in semantic space



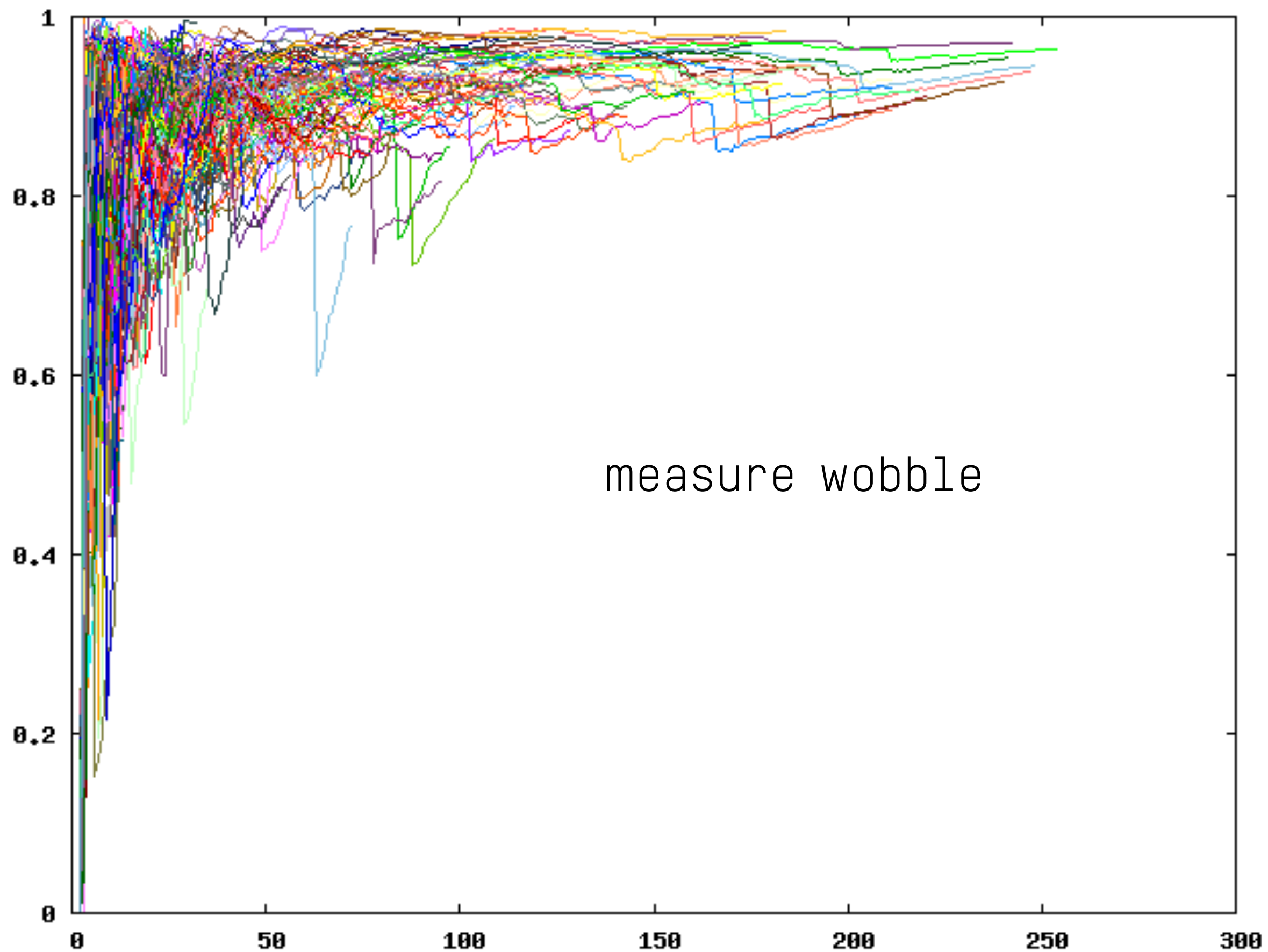
Gavagai



Example contexts, some found in naturally occurring human-generated text, some generated by term substitution. (Can you pick out which are which?)

1. health benefits of coconut oil include hair care, skin care, and proper digestion and metabolism
2. the coconut tree is a member of the palm family
3. in the rendang beef stew from sumatra, chunks of beef are cooked in coconut milk along with other spices
4. thanks to a promotion from the airline you can now book a coconut to frankfurt for 100 off
5. he looks dapper in a coconut as he arrives for the emporio armani show during milan fashion week
6. it's on the 28th floor of the coconut and it's got all the charms of a corporate headquarters





take home message

information retrieval is a component
sessions are more informative than one-shot
relevance feedback is potentially useful
log analysis can yield relevance feedback
different use cases have different needs
... and should be evaluated differently
top line is not always best practice

