information access systems - beyond one shot

kth
jussi karlgren
march 2014

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"

back to usefulness for task

happiness, trust, and satisfaction!

so who is our user here? and why?

let's broaden the scope a bit from information retrieval

information access

what is the broader service the information retrieval system is a component of?

what is the primary task of the user?

knowledge sources for understanding information access

text

language

text context

social context

task context

platform

application context

user goals

system provider goals

some example tasks

web search

bibliographic search in a library

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

how do you measure

happiness, trust, and satisfaction

for these tasks?

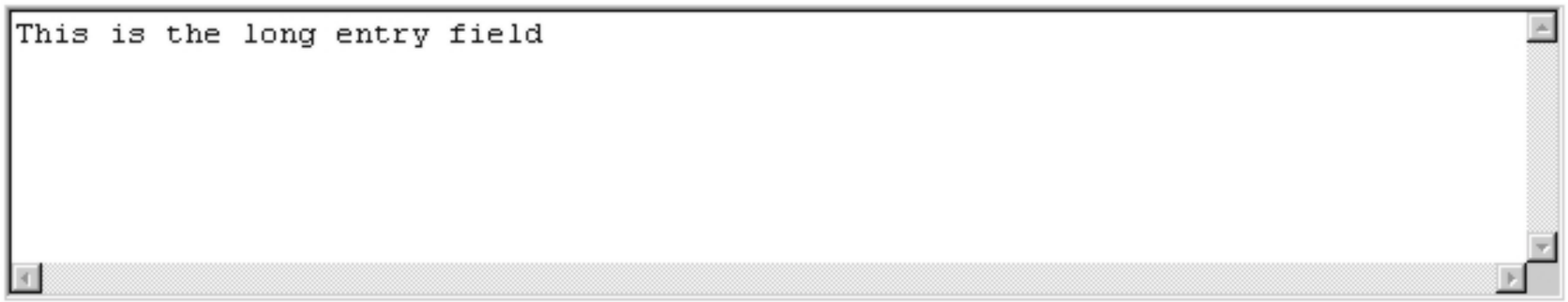
modelling usage in adhoc search is a challenge:

1.87 wds / q

but we have sessions!

and logged-in users' profiles!

design matters for average query length



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

log analysis

dwell time

click thru

returning visits

conversion rate

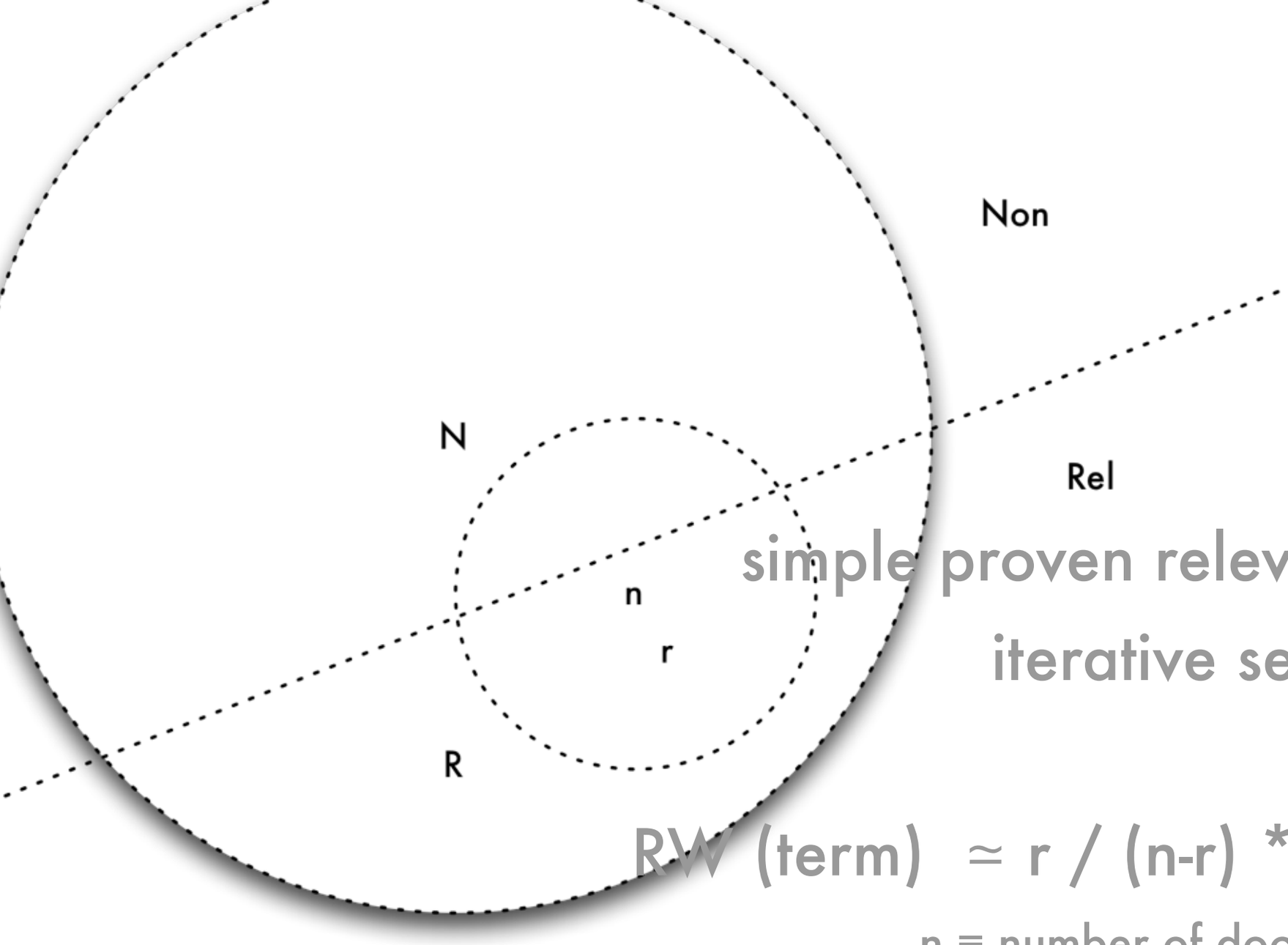
user actions

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
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.o
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

$$RW(term) \simeq r / (n-r) * ((N-n)-(R-r))/(R-r))$$

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

implicitly: negative ≈ if not clicked through

(but this depends on what positive and negative are assumed to mean: if the negative feedback is about picking out one specific negative topic it will work quite well)

given relevance feedback, how to use it?

interactive query expansion

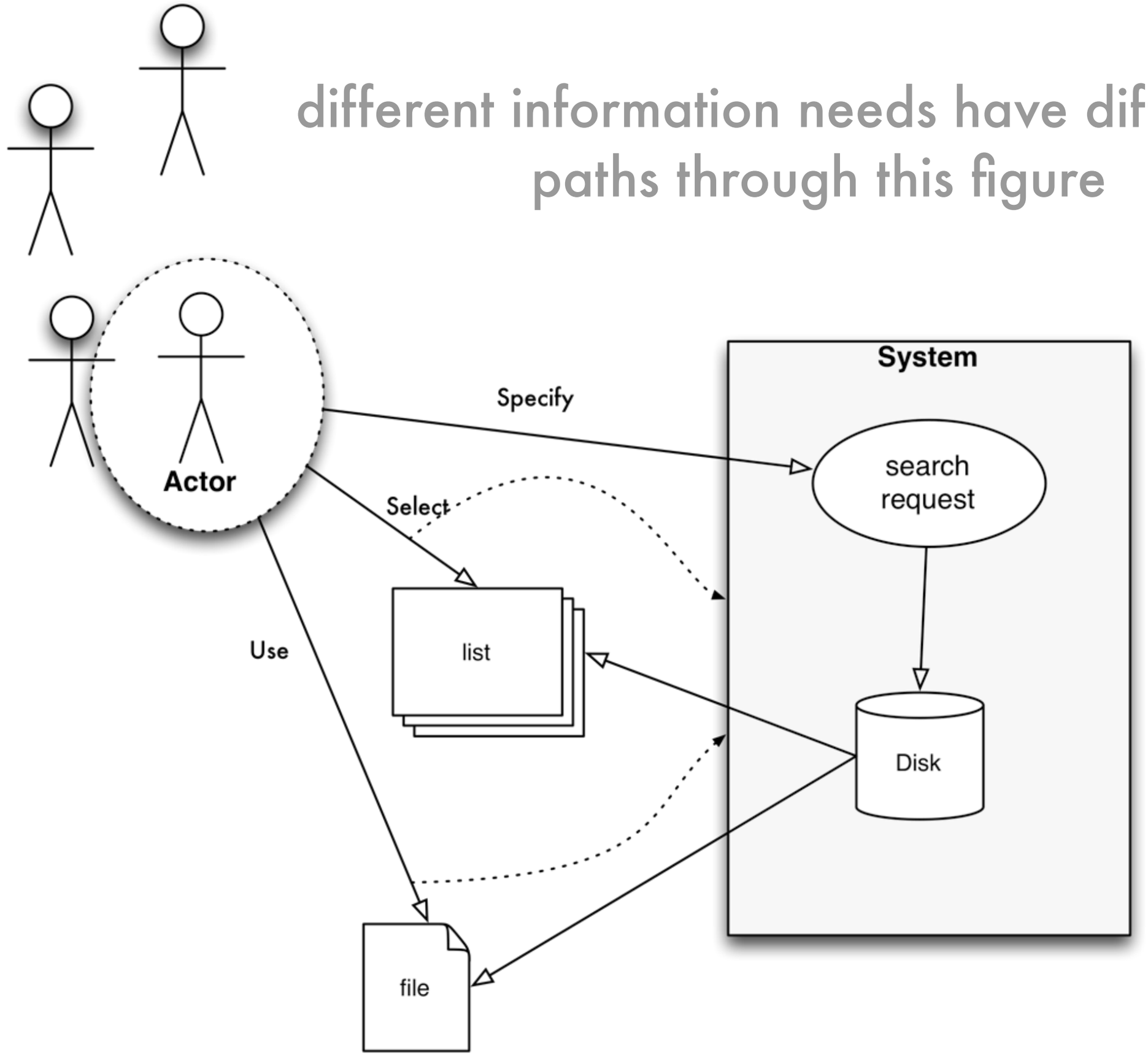
(more work for the user)

automatic (hidden) query expansion

(risky, for obvious reasons)

now let's look a little at the session.

different information needs have different paths through this figure

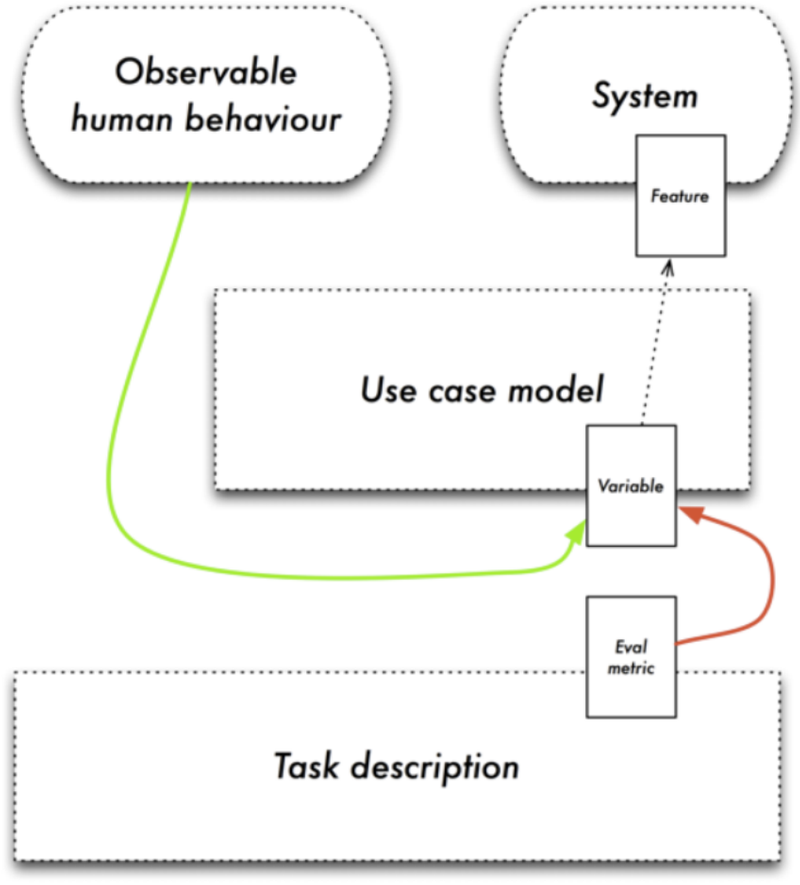


examples of high recall needs?

examples of high precision needs?

examples of other target notions than relevance?

use cases as a modelling framework



use cases have different needs and require differentl eval!

(this needs to be made explicit and should be part of user modelling step in system development)

best practice in an application context?

satisficing vs optimisation: optimal solution may be something else than the optimal algorithm

10 min exercise

formulate use case for entertainment field

with quantitative evaluation criteria

"big data"

game changer

evaluation not by lab experiment on canned data but by methods similar to meteorology: based on similar data, results should be *sufficiently* similar

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

big data changes use cases and thus evaluation

top line is not always best practice