Online and Offline Evaluation of Search Engine Quality

Evangelos Kanoulas

University of Amsterdam

What are the current exhibitions at Prado Museum?

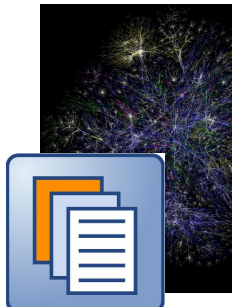
User's Re

Goog

Search Engine

prado
prado **museum**
prado
prado**sham**
paradox
prado**floxacin**
prado**foto**

Collecti



MUSEO NACIONAL DEL **PRADO**

Español

THE COLLECTION
EXHIBITIONS
Current Exhibitions
Upcoming Exhibitions
Past Exhibitions
EDUCATION
RESEARCH
PRADOMEDIA

THE INSTITUTION
CORPORATE AREA
PRESS

Supported by:
Telefónica

Home› Exhibitions› Current Exhibitions

Current Exhibitions

Exhibitions in the Museum

The Plácido Arango donation
July 7
October 04, 2015

The invited work: *The young Saint John the Baptist*, Michelangelo
March 31
September 20, 2015

10 Picassos from the Kunstmuseum Basel
March 18
September 14, 2015

Touching the Prado
January 20
October 18, 2015

Monumental Views of Spanish Cities. The Romantic Painter Genaro Pérez Villaamil
December 17, 2014
September 06, 2015

SEARCH

VISIT THE MUSEUM

Tu visita	La teva visita
A túa visita	Zure bisitaldia
Your visit	La seua visita
Votre visite	La vostra visita
Sua visita	Dein Besuch
Ваш визит	参观
ご来館について	زيارتك للمتحف

BUY YOUR TICKET
and avoid queuing

PRADO SHOP

BECOME A FRIEND

APPS

SOCIAL

Change the search algorithm.

How can we know whether we made the users happier?

HiPPO-driven Entrepreneurship

- Highest Paid Person's Opinion
  - Often wrong

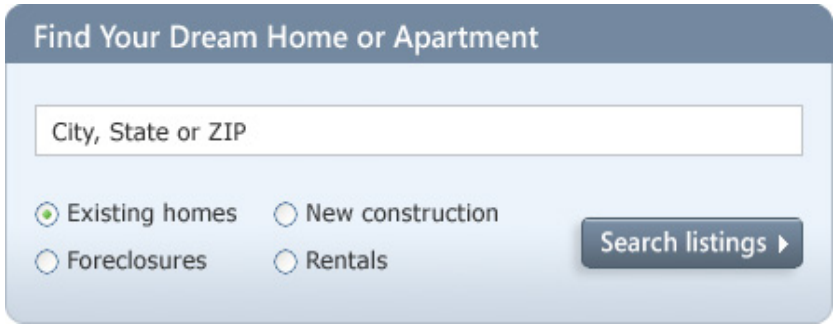
Data-driven Entrepreneurship

- Avinash Kaushik (author and entrepreneur)
  - "80% of the time we were wrong about what a customer wanted"
- QualPro (management consultancy)
  - tested 150,000 ideas over 22 years
  - 75% of important business decisions and business improvement ideas either have no impact on performance or actually hurt performance.
- Google
  - Google ran approximately 12,000 experiments in 2009, with [only] about 10% of these leading to business changes.
- Microsoft
  - At Microsoft only 1/3 of ideas were positive ideas and statistically significant

MSN Real Estate

- "Find a house" widget variations

  A was 8.5% better.



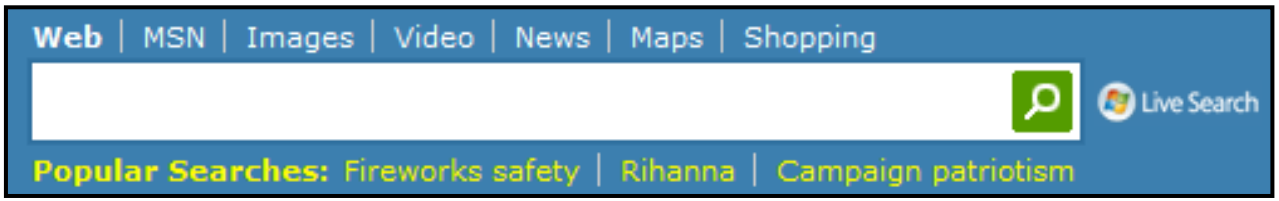
A

B

Ronny Kohavi, Microsoft

MSN Home Page Search Box

A



B

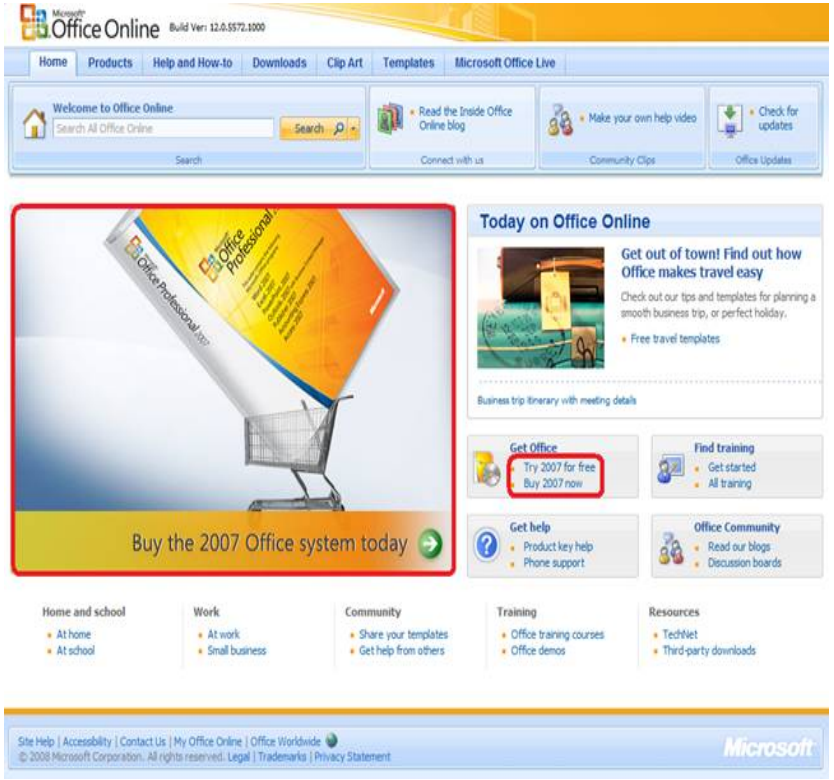
No significant difference

Ronny Kohavi, Microsoft

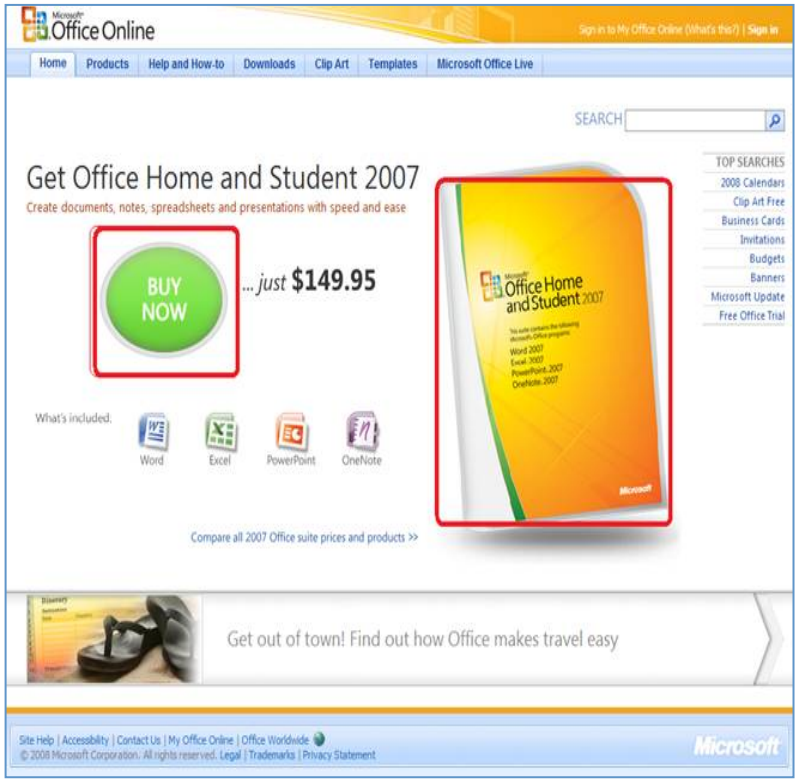
Office Online

A was 64% better

A



B



Ronny Kohavi, Microsoft

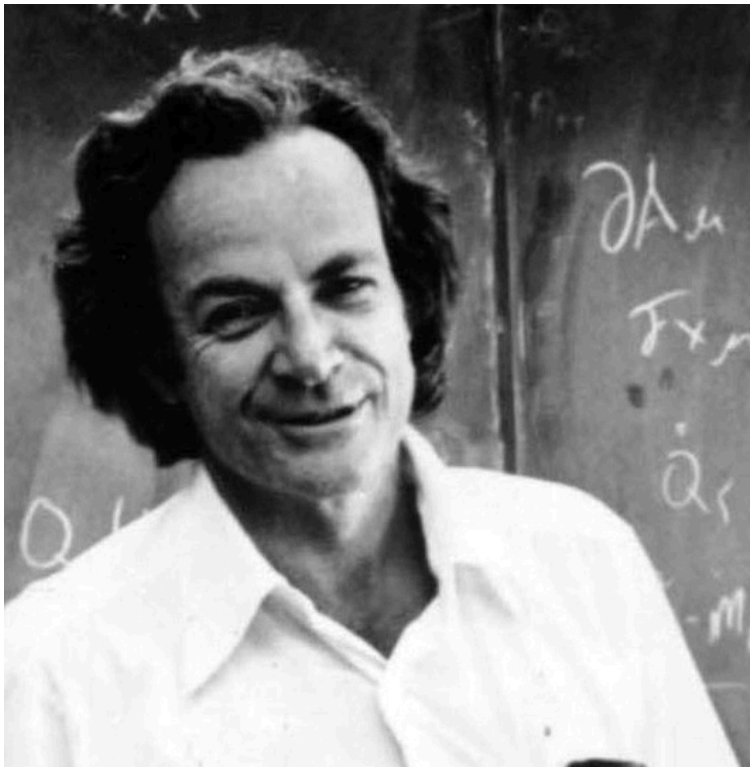
Key Lessons from Industry

- Don't trust the HiPPO
- Experiment often
  - *To have a great idea, have a lot of them* -- Thomas Edison
- Avoid building optimal features without early testing
- Try radical ideas
  - *If you're not prepared to be wrong, you'll never come up with anything original* -- Sir Ken Robinson

Scientific Method

Finding a new law:

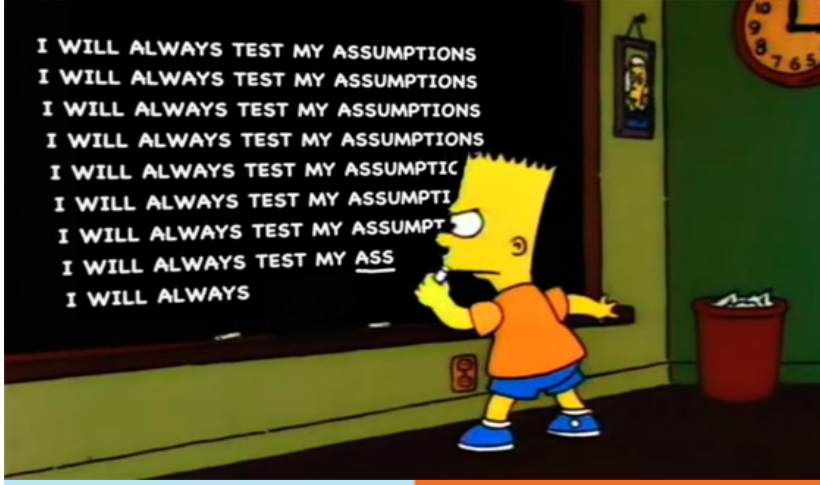
1. We guess it.
2. We compute the consequences of the guess
3. We compare the result of the computation to observation
4. If it **disagrees** with **experiment** it is **wrong**.

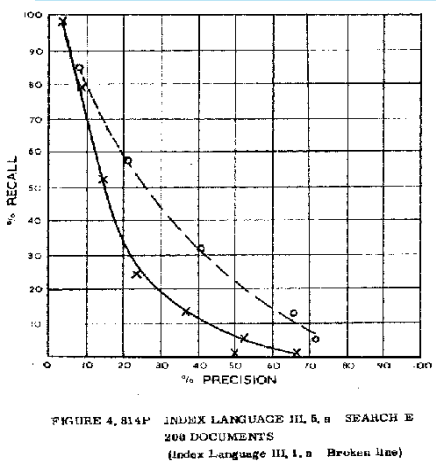


"What you can't measure you can't improve"
Lord Kelvin

Different approaches to evaluation

- User-studies
- Collection-based evaluation
- In-situ evaluation
  - A/B Testing
  - Interleaving





Outline

PART I

1. Collection-based Evaluation
2. Comparative Evaluation

PART II

3. Observable User Behavior
4. A/B Testing
5. Interleaving
6. Comparative Studies

1. Collection-based Evaluation

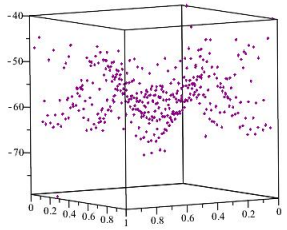
Happiness: elusive to measure

- Most common proxy: relevance of search results
  - But how do we test relevance?

Static Test Collections

Machine Learning

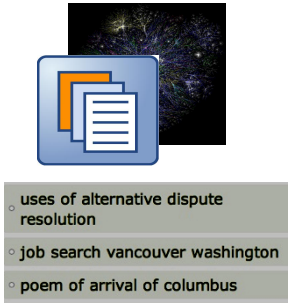
- Feature vectors



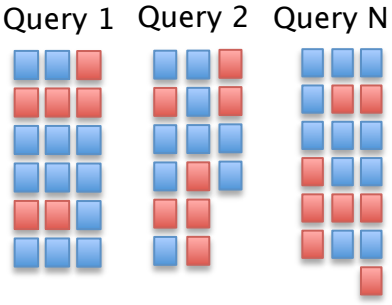
- Labels

Information Retrieval

- Documents
- Queries



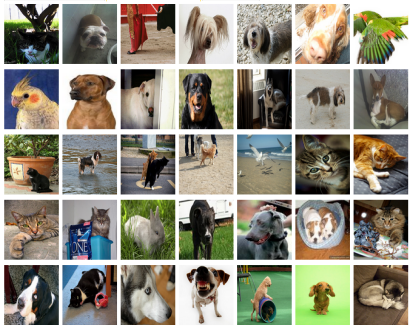
- Labels
  - relevance judgments



Static Test Collections

Computer Vision

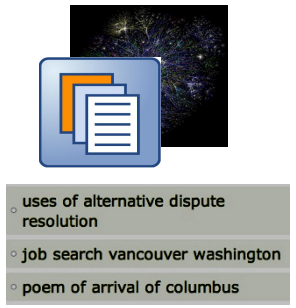
- Images



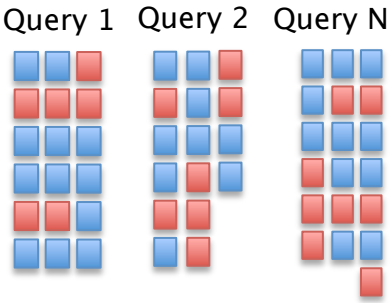
- Labels

Information Retrieval

- Documents
- Queries



- Labels
  - relevance judgments



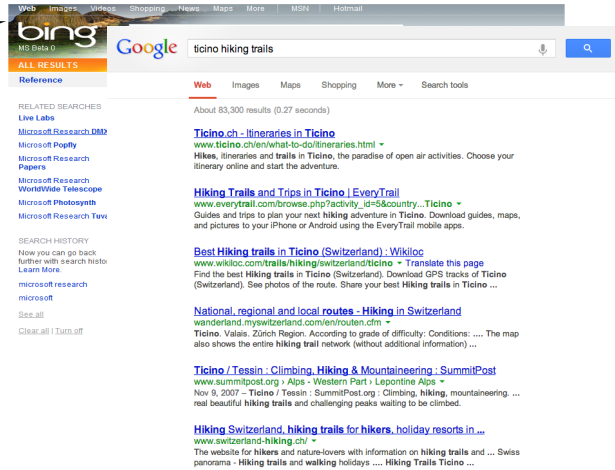
SE Evaluation/Optimization

Id	Query	Relevant Web Pages	Not Relevant Web Pages
1	prado	• https://www.museodelprado.es	
2	real madrid	• http://www.realmadrid.com/en • http://www.goal.com/en/teams/spain/real-madrid/2016 • http://bleacherreport.com/real-madrid	
3	uned	• http://portal.uned.es/ • https://twitter.com/uned	
4	malasaña restaurants	• http://www.lonelyplanet.com/spain/madrid/malasana-chueca/restaurants • http://www.tripadvisor.com/Travel-g187514-d246519/Madrid:Spain:Malasana.html • http://www.labarraca.es/	
5	flamenco madrid or sevilla	• http://www.fodors.com/is-it-better-to-see-flamenco-in-seville-or-madrid.cfm • http://www.tripadvisor.com/Best_Flamenco_Shows_Madrid_or_Seville-Spain.html	
	…	…	

Test Collection Construction

- uses of alternative dispute resolution
- job search vancouver washington
- poem of arrival of columbus

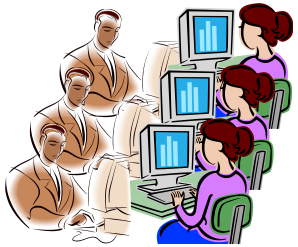
Results



Search Algorithms

Skewed Data

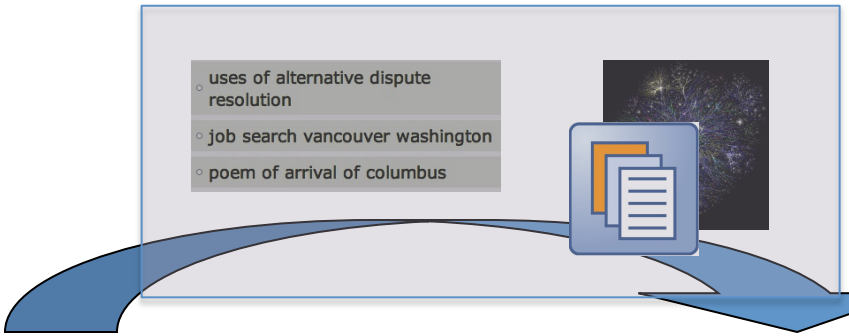
Judges



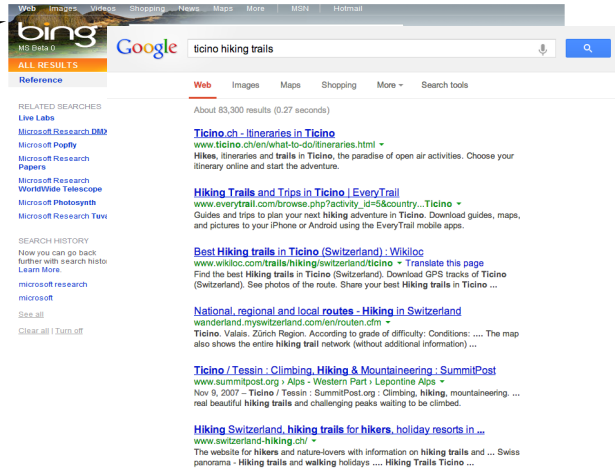
QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

Queries and Doc. Collection

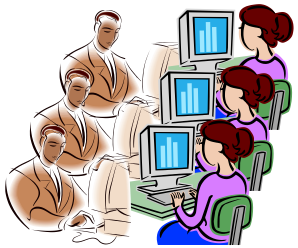


Results



Search Algorithms

Judges



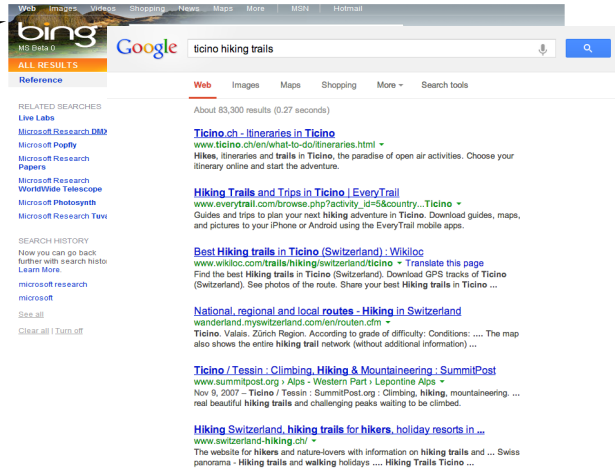
QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

Relevance Judgments

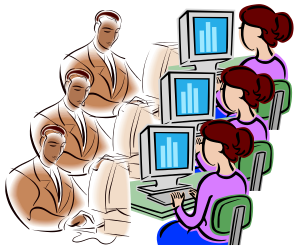
- uses of alternative dispute resolution
- job search vancouver washington
- poem of arrival of columbus

Results



Search Algorithms

Judges



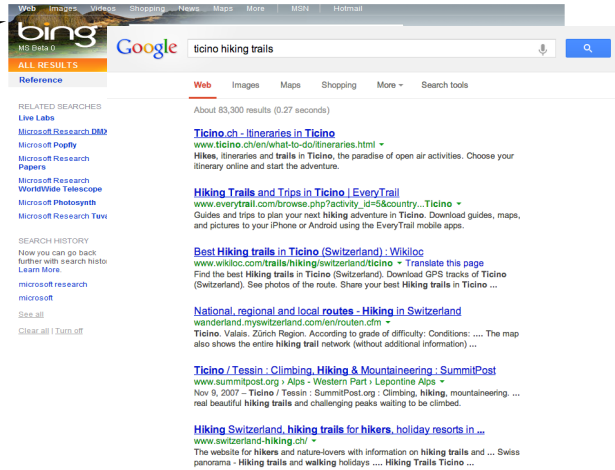
QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

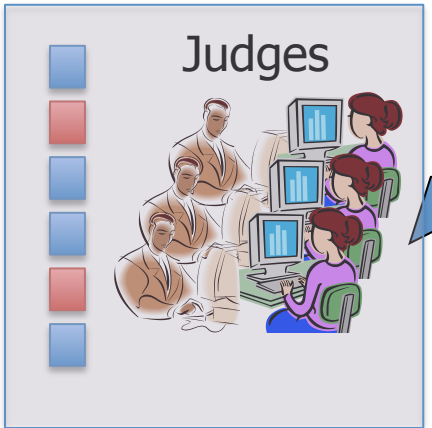
Relevance

- uses of alternative dispute resolution
- job search vancouver washington
- poem of arrival of columbus

Results



Search Algorithms



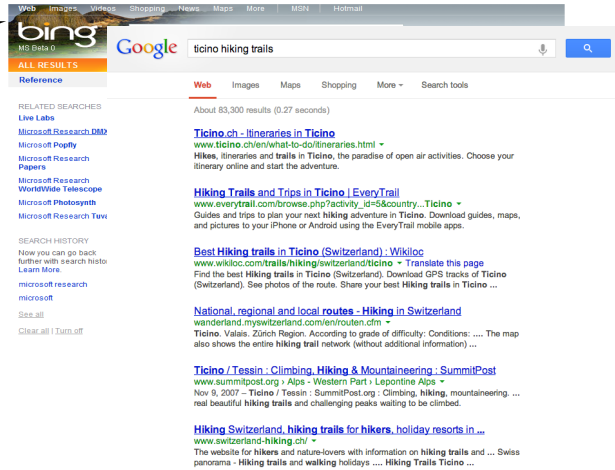
QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

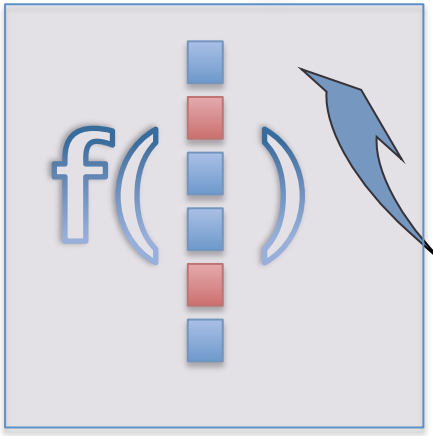
Evaluation Measures

- uses of alternative dispute resolution
- job search vancouver washington
- poem of arrival of columbus

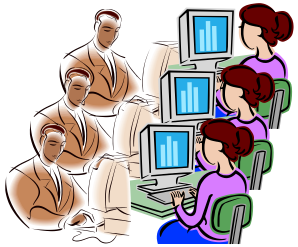
Results



Search Algorithms



Judges



QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

Part 1: Evaluation Measures

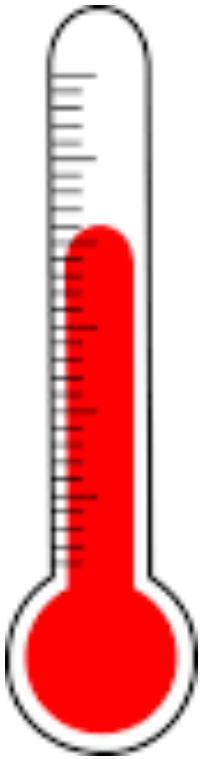
- Evaluation measures
  1. Traditional evaluation measures
  2. Model-based measures
  3. Novelty and Diversity-based measures
  4. Session-based measures

- Evaluation of evaluation measures

What is temperature?

The measure defines the problem:

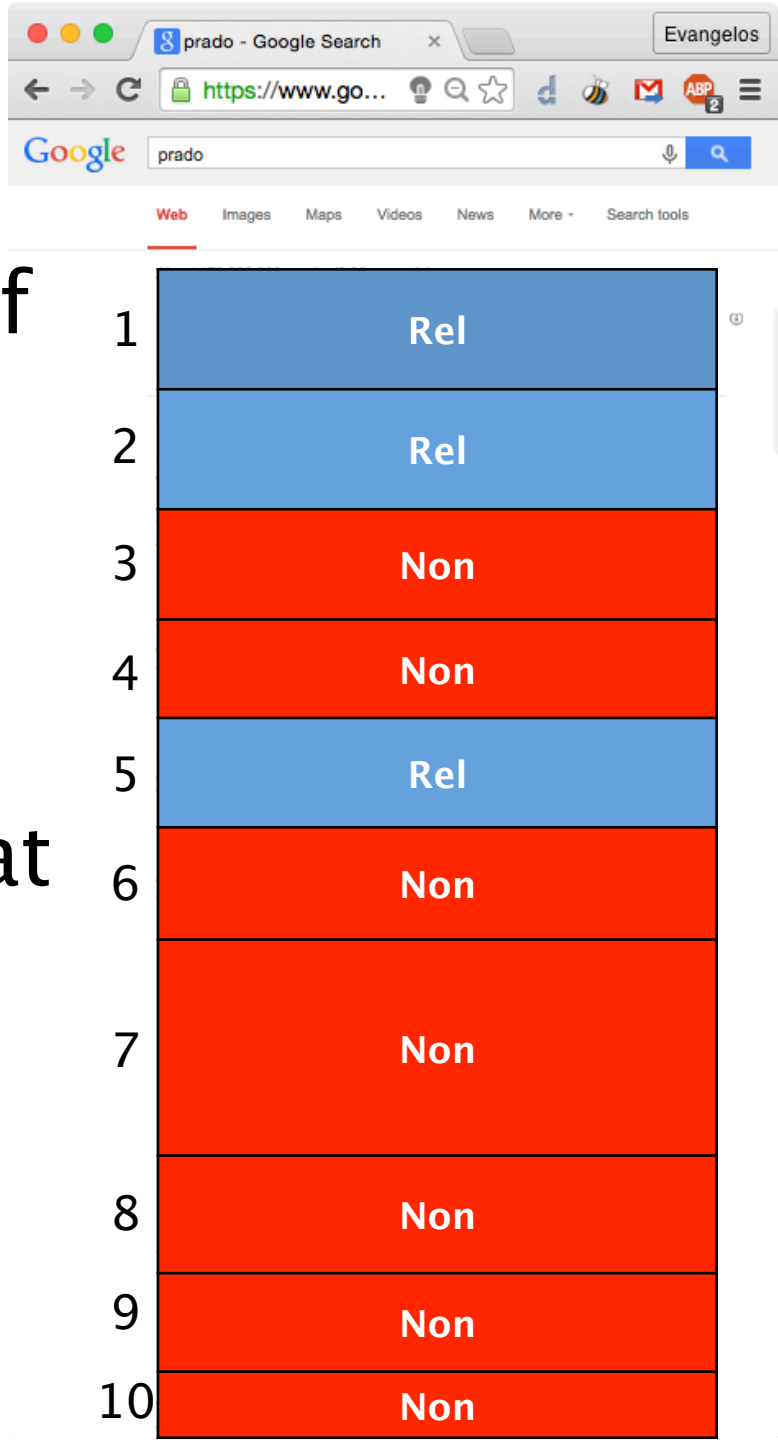
- An adequate measure is fundamental towards solving the problem
- An inadequate measure implies focusing on the wrong problem



Courtesy: Julio Gonzalo

Precision and Recall

- Precision (P): fraction of retrieved documents that are relevant
- Recall (R): fraction of relevant documents that are retrieved



High precision vs. high recall retrieval

- There is usually a tradeoff between precision and recall
  - by returning more documents, recall increases (monotonically)
  - by returning fewer documents, often (not always) precision increases

- High precision search
    - (web search)
  - usually user only looks at very few top results : e.g. precision@3
  - repository is large, easy to find some relevant pages
  - imperfect recall is often tolerable

- High (total) recall search
    - (vanity, e-discovery, health, patent, legal)
  - total recall challenge: find documents that are not "obvious" matches
    - http://plg.uwaterloo.ca/~gvcormac/total-recall/

Precision and Recall for rankings

- Precision/recall assume that:
  - We retrieve a fixed set of documents; and
  - The user examines all of them.

  - However, more typically we return a ranked list (open-ended); and
  - The user examines the top with higher probability than the bottom

Average Precision

- Average Precision (AP) = Average of precisions at relevant documents

Rank	Rel.	Precision@k	Recall@k
1	R	1/1	1/10
2	N	1/2	1/10
3	R	2/3	2/10
4	R	3/4	3/10
5	N	…	…
6	R	4/6	4/10
7	N		
8	N	…	…
9	N		
10	R	5/10	5/10
…	…	…	…
∞	R	0	10/10

$$AP = \frac{\frac{1}{1} + \frac{2}{3} + \frac{3}{4} + \frac{4}{6} + \frac{5}{10} + \ldots}{10}$$

Beyond Binary Relevance

- Discounted Cumulative Gain (DCG)
  - Used by many web search companies
    - Graded relevance judgments as opposed to binary
      - Emphasis on retrieving highly relevant documents
  - Two assumptions:
    - Graded relevance as a measure of the usefulness or gain from examining a document; hence the higher the better
    - The lower the ranked position of a document, the less useful it is for the user, since it is less likely to be examined

Discounted Cumulative Gain (DCG)

- DCG is the total gain accumulated at a particular rank *k*:

$$\mathrm{DCG@k} = \sum_{\mathrm{rank}\ r=1}^{k} \frac{2^{rel_r} - 1}{log_2(1+r)}$$

non-linear gain

discount

Discounted Cumulative Gain

1	Nav= 5
2	Nav= 5
3	Non = 0
4	Non = 0
5	Rel = 1
6	Non = 0
7	Non = 0
8	Non = 0
9	Non = 0
10	Non = 0
…	…

Gain by relevance

$2^{rel}-1$

Gain

31
31
0
0
1
0
0
0
0
0

Discount by rank

$1/log_2(r+1)$

Discounted Gain

31
19.55
0
0
0.38
0
0
0
0
0

DCG@10 = 50.94

Normalized DCG

- Normalizes DCG against the best possible DCG result (the perfect ranking) for the query
  - 0<=NDCG<=1
  - makes averaging easier for queries with different numbers of relevant documents

Part 1: Evaluation Measures

- Evaluation measures
  1. Traditional evaluation measures
  2. Model-based measures
  3. Novelty and Diversity-based measures
  4. Session-based measures

- Evaluation of evaluation measures

Measures of Effectiveness

- Determine how good the system is at finding and ranking relevant documents
- A good effectiveness measure should correlate with the user's experience
- Thus interest in effectiveness measures based on explicit models of user interaction

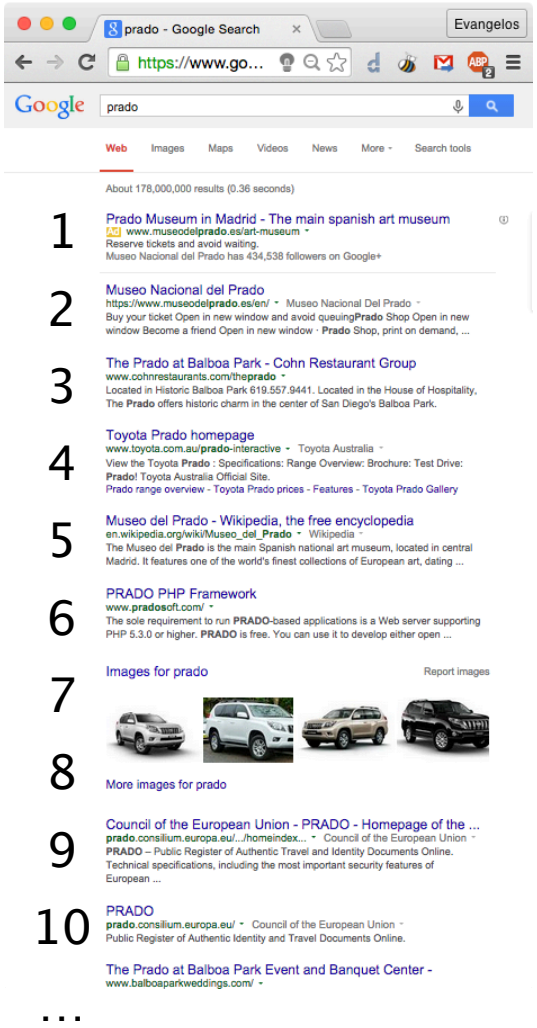
Model-based Measures

1. A browsing model that describes how a user interacts with results;

2. A model of document utility, describing how a user derives utility from individual relevant documents;

3. A utility accumulation model that describes how a user accumulates utility in the course of browsing

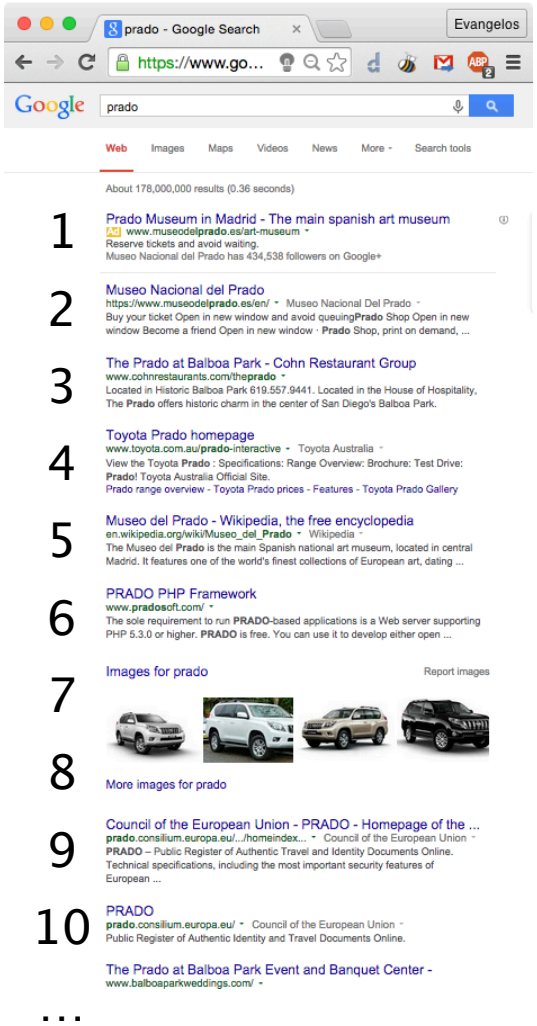
Model Browsing Behavior



1 Position-based models

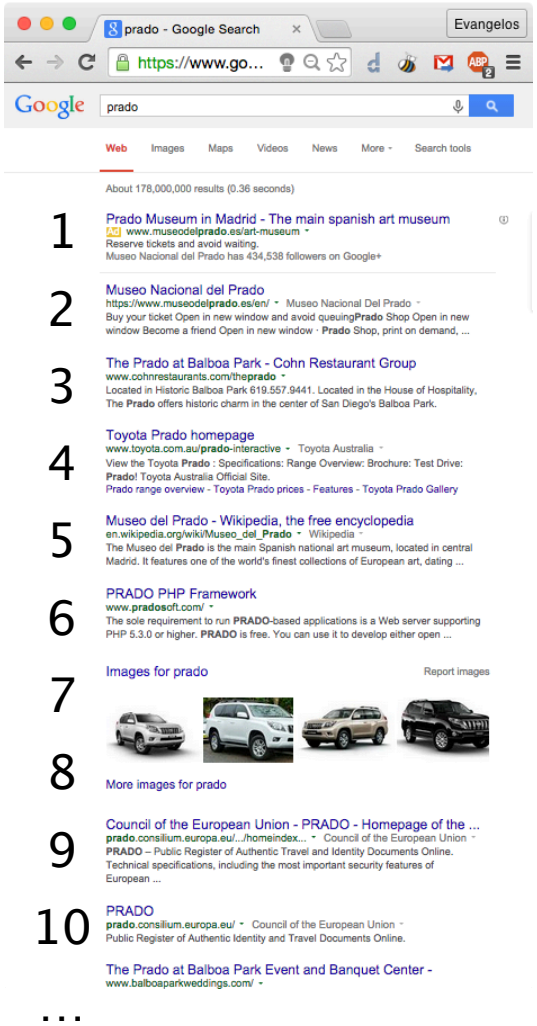
The chance of observing a document depends on the position of the document in the ranked list.

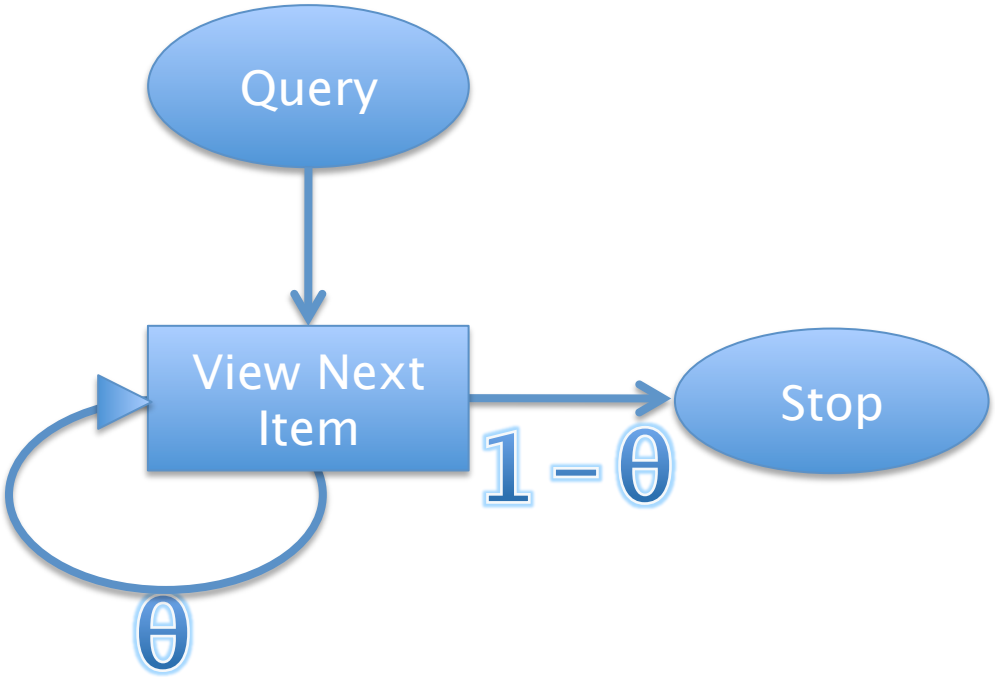
Ranked Biased Precision Model



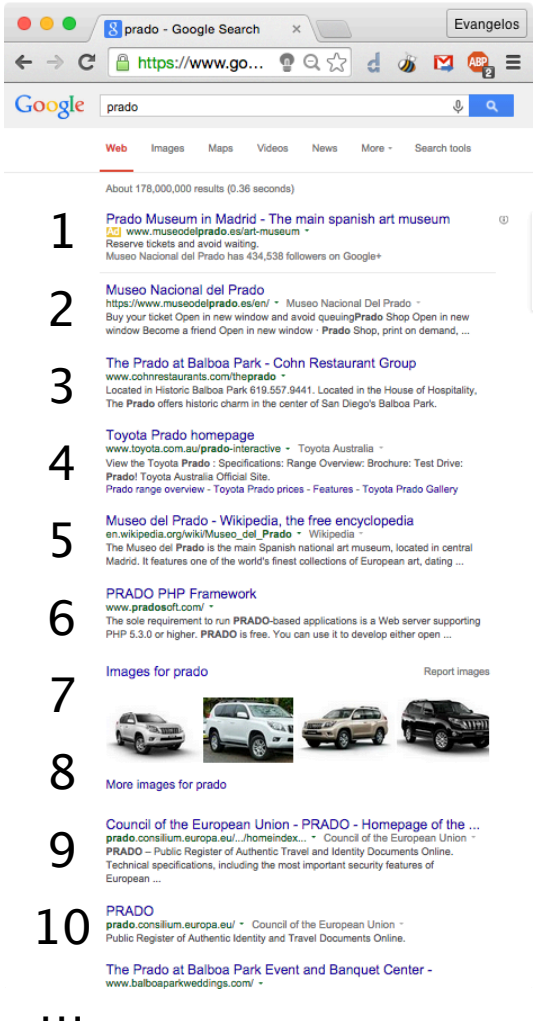
1. Toss a biased coin (θ)
2. If HEADS, observe next document
3. IF TAILS, stop

Ranked Biased Precision Model



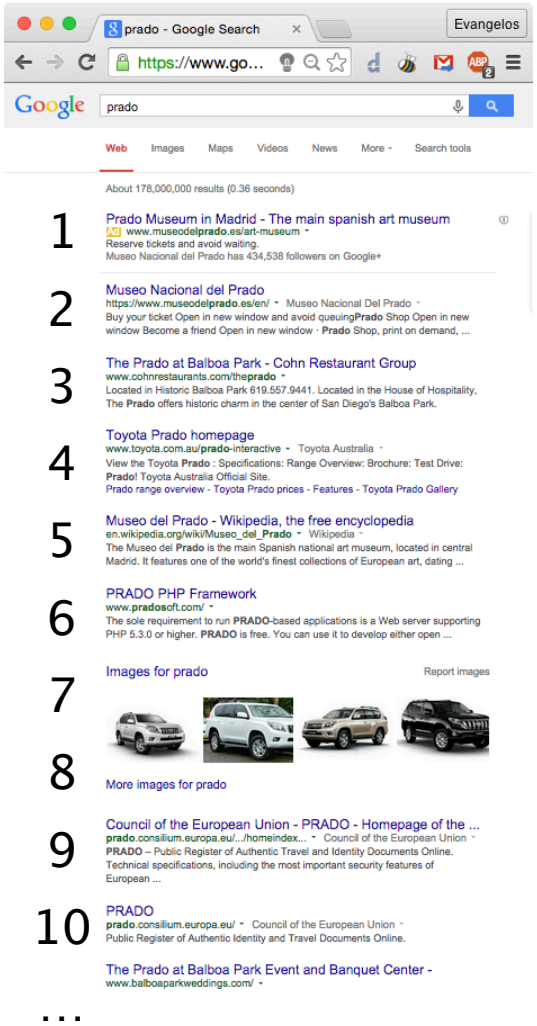


RBP – Expected Utility at Stopping



$$\mathrm{RBP} = \sum_{k=1}^{\infty} rel_k \vartheta^{k-1}(1-\vartheta)$$

Model Browsing Behavior



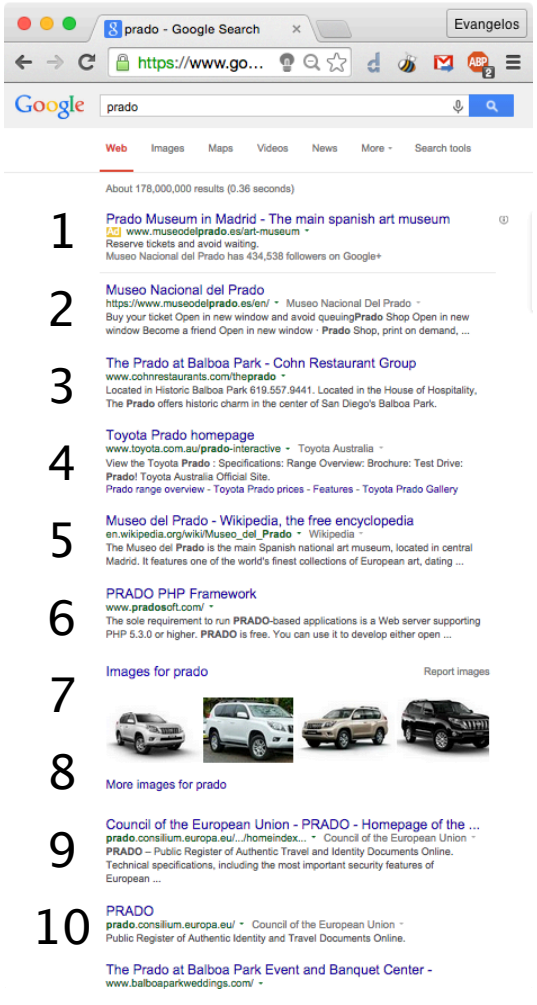
2 Cascade-based models

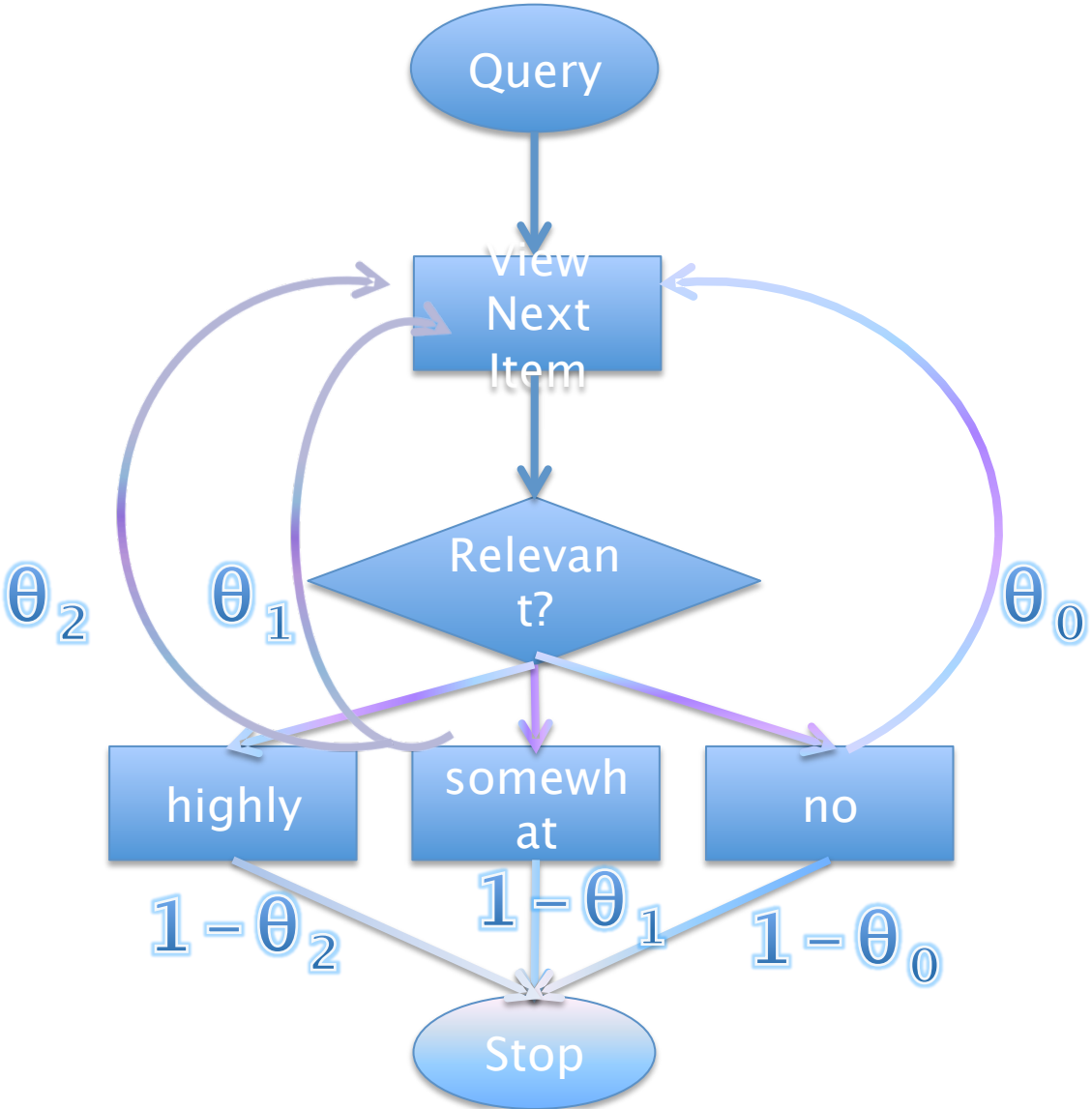
Algorithm 1 The cascade user model

Require: $R_1, \ldots, R_{10}$ the *relevance* of the 10 URLs on the result page.

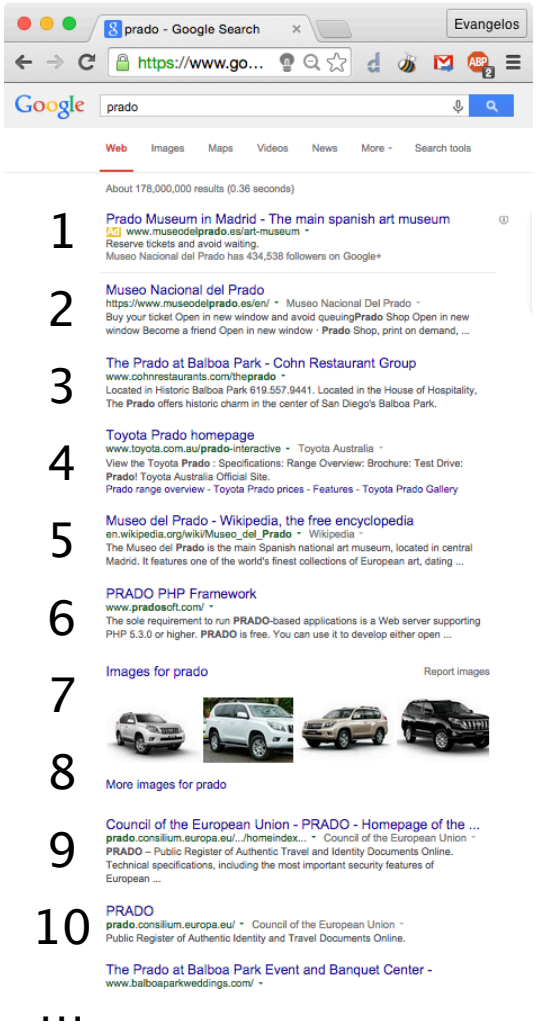
1: $i = 1$
2: User examines position i.
3: **if** random(0,1) $\leq R_i$ **then**
4: User is satisfied with the document in position i and stops.
5: **else**
6: $i \leftarrow i + 1$; go to 2
7: **end if**

Expected Reciprocal Rank Model



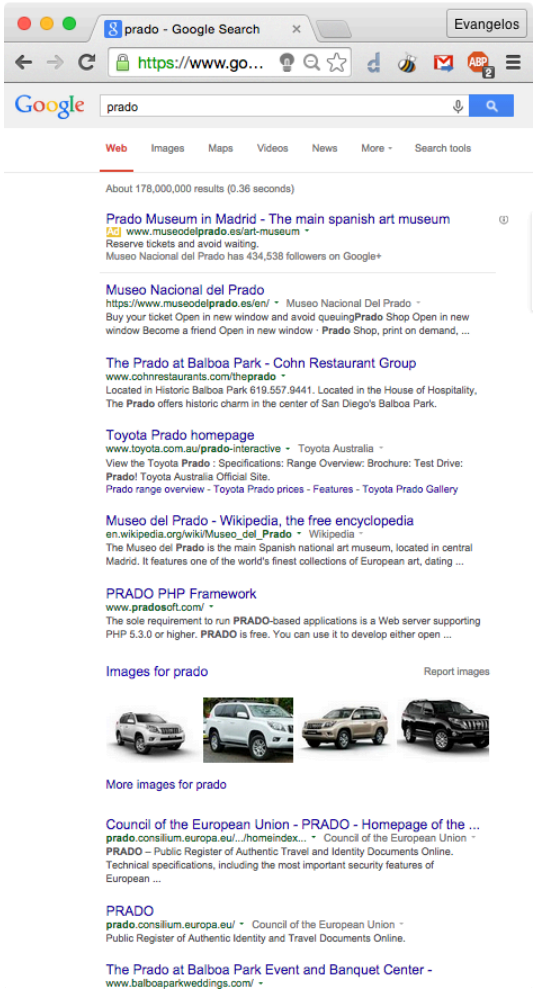


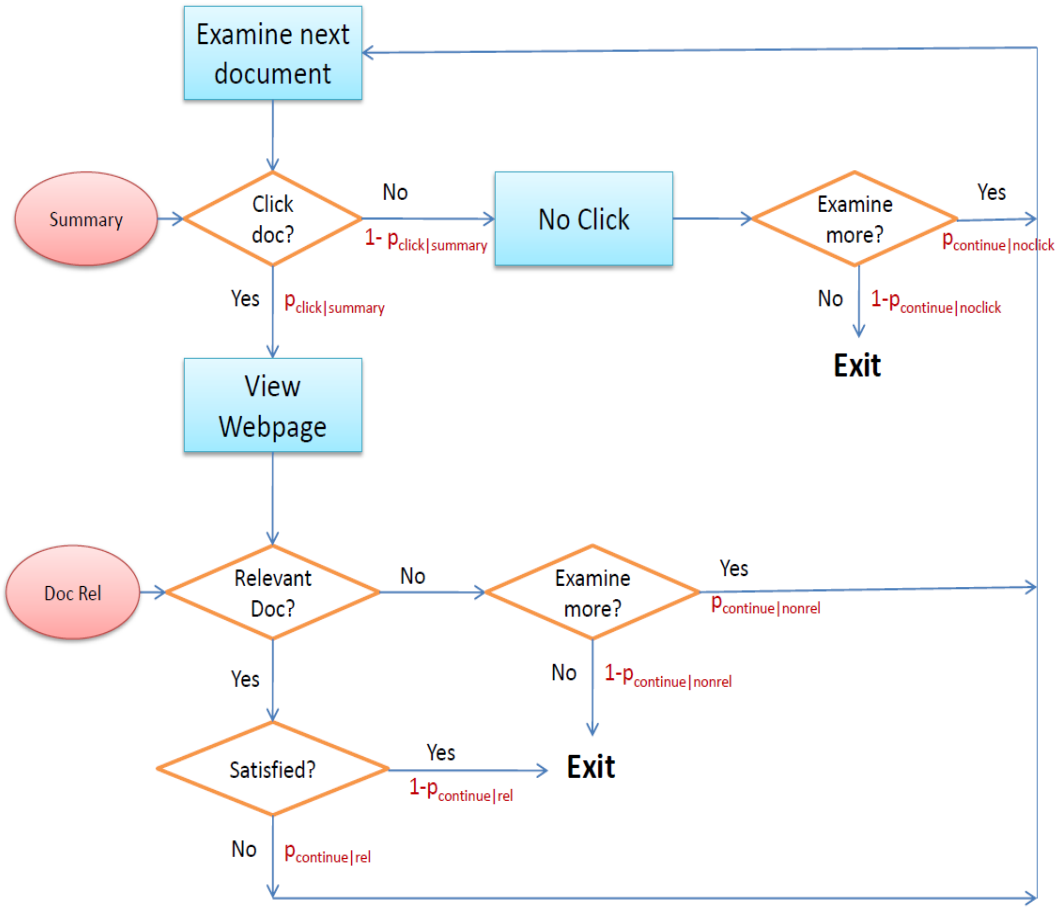
ERR – Expected Effort



$$\mathrm{ERR} = \sum_{r=1}^{n} \left(\prod_{i=1}^{r-1} (1 - \vartheta_i) \vartheta_r \right) \frac{1}{r}$$

Expected Browsing Utility Model





Click model-based Measures

Click Model-Based Information Retrieval Metrics

Aleksandr Chuklin
Yandex & ISLA, University of Amsterdam
Moscow, Russia
A.Chuklin@uva.nl

Pavel Serdyukov
Yandex
Moscow, Russia
pavser@yandex-team.ru

Maarten de Rijke
ISLA, University of Amsterdam
Amsterdam, The Netherlands
deRijke@uva.nl

Click Models for Web Search

Aleksandr Chuklin
University of Amsterdam and Google Switzerland

Ilya Markov
University of Amsterdam

Maarten de Rijke
University of Amsterdam

Time-based Measures

Time-Based Calibration of Effectiveness Measures

Mark D. Smucker
Department of Management Sciences
University of Waterloo, Canada
mark.smucker@uwaterloo.ca

Charles L. A. Clarke
School of Computer Science
University of Waterloo, Canada
claclark@plg.uwaterloo.ca

Choosing Parameter Values

- All user models have parameters
  - Metrics evaluated at fixed parameter values
  - Evaluation w.r.t. an average user

- Different approaches:
  - Predefine parameters
    - ERR: $\vartheta_g = \frac{2^g - 1}{2^{g_{\max}}}, \quad g \in \{0, \ldots, g_{\max}\}$
  - Use click log; fit a model to gaps between clicks (Zhang et al., IRJ, 2010)
  - Minimize variance in evaluation (Kanoulas & Aslam, CIKM '09)

Choosing Parameter Values

Simulating Simple User Behavior for System Effectiveness Evaluation

Ben Carterette*, Evangelos Kanoulas‡, Emine Yilmaz†
carteret@cis.udel.edu, e.kanoulas@shef.ac.uk, {eminey@microsoft.com, eyilmaz@ku.edu.tr}
* Department of Computer & Information Sciences, University of Delaware, Newark, DE
‡ Information School, University of Sheffield, Sheffield, UK
† Microsoft Research, Cambridge, UK & Koc University, Istanbul, Turkey

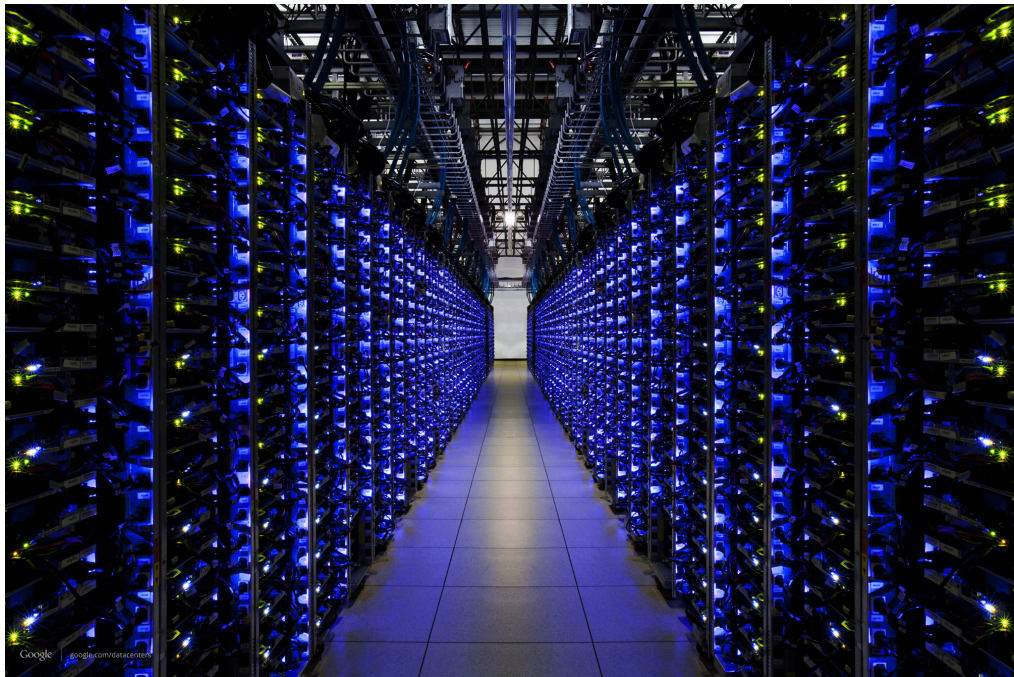
Incorporating Variability in User Behavior into Systems Based Evaluation

Ben Carterette*, Evangelos Kanoulas‡, Emine Yilmaz†
carteret@cis.udel.edu, ekanoulas@gmail.com, {eminey@microsoft.com, eyilmaz@ku.edu.tr}
* Department of Computer & Information Sciences, University of Delaware, Newark, DE
‡ Information School, University of Sheffield, Sheffield, UK
† Microsoft Research, Cambridge, UK & Koc University, Istanbul, Turkey

- Users behave very differently when they search
  - Distribution of parameters (users) need to be considered
- A different approach
  - Mine Web Query logs
  - Learn a distribution of the parameters
  - Use this distribution to evaluate the quality of systems

Patience Distribution for RBP

- Goal: produce a posterior distribution for θ
- Start with a uniform distribution for θ
- Update it based on logged data



Posterior Distribution of Patience θ for RBP

$$P(\theta \mid E) = P(\theta \mid c) = \sum_{r=0}^{\infty} P(\theta \mid r, c) P(r \mid c)$$

$$P(\theta \mid r, c) \propto P(c \mid \theta, r) P(\theta \mid r)$$

The probability that user skips r document

$$P(c \mid \theta, r) = NB(r, \theta)$$

Probability distribution of the number of successes in a sequence of Bernoulli trials before **r** failures occurs.

$$P(\theta \mid r) = Beta(\alpha, \beta)$$

- Start with uniform prior ($\alpha = \beta = 1$)

Posterior Distribution of Patience θ for RBP

$$P(\theta \mid r, E) \propto P(E \mid \theta, r) P(\theta \mid r)$$

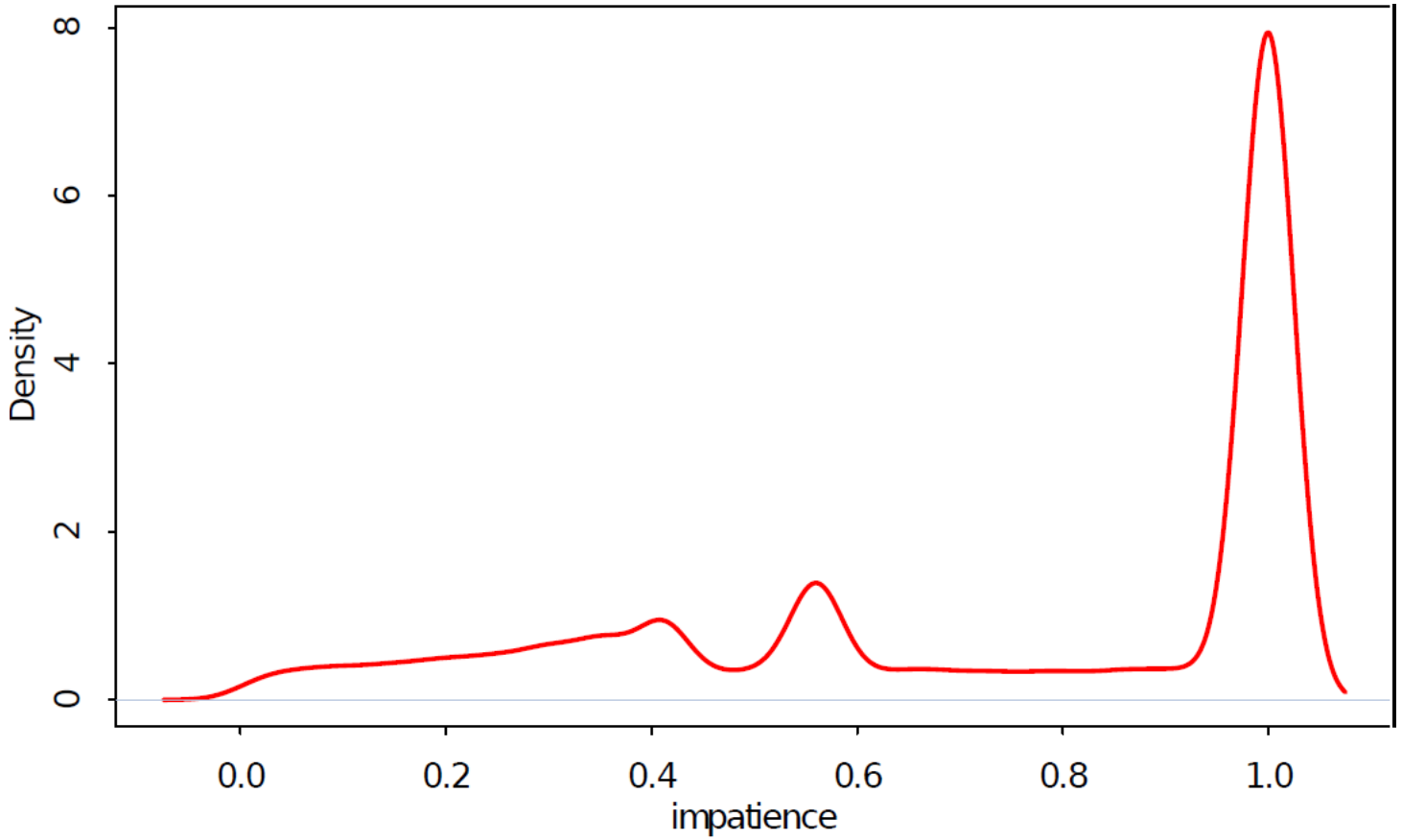
$$P(E \mid \theta, r) = NB(r, \theta) \qquad P(\theta \mid r) = Beta(\alpha, \beta)$$

- If there are m queries, with r number of failures, and c_i number of successes, i=1..m

$$P(\theta \mid r, E) = Beta(\alpha + \sum_{i=1}^{m} c_i, \beta + mr)$$

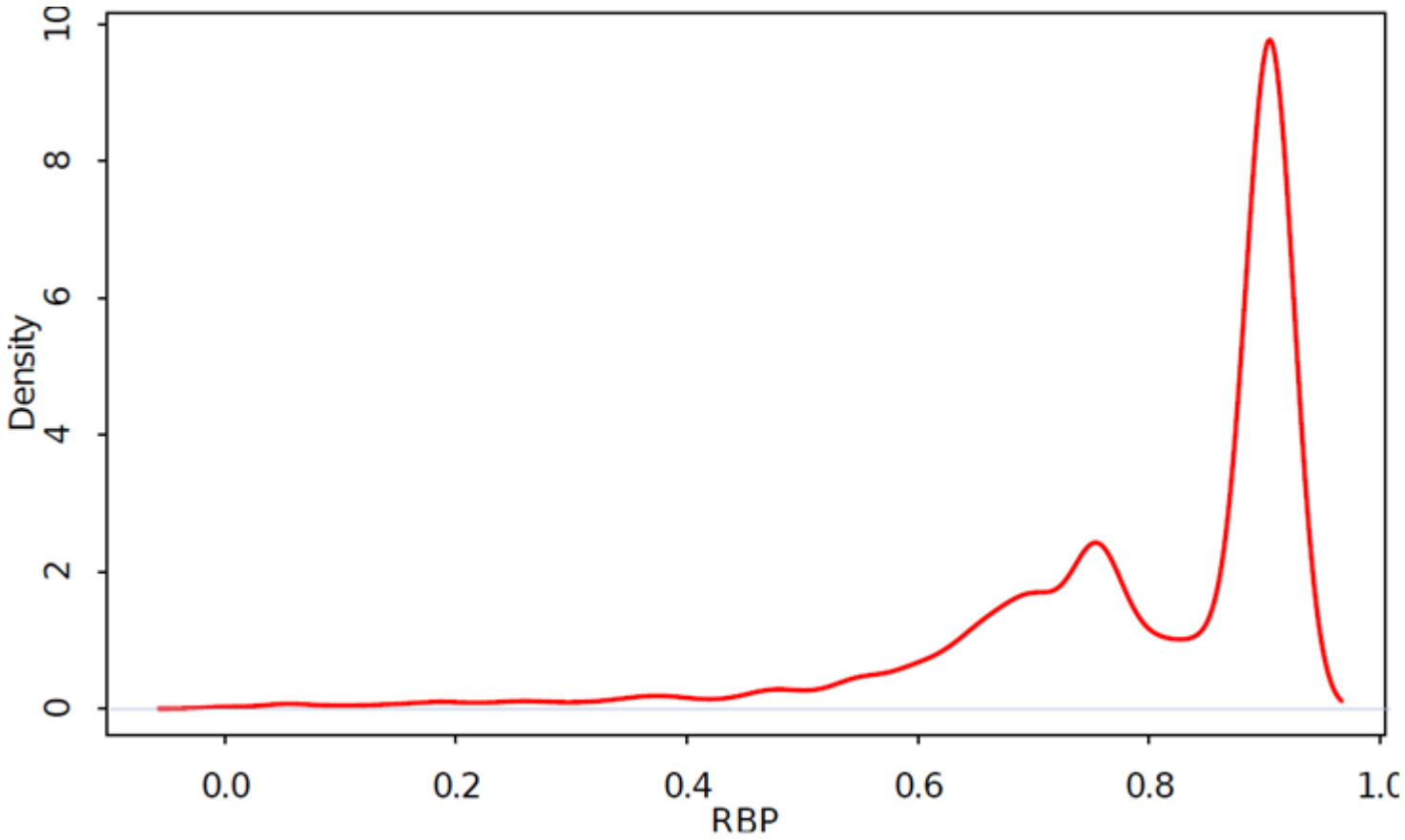
Posterior Distribution for Impatience: RBP

- Using the AOL log



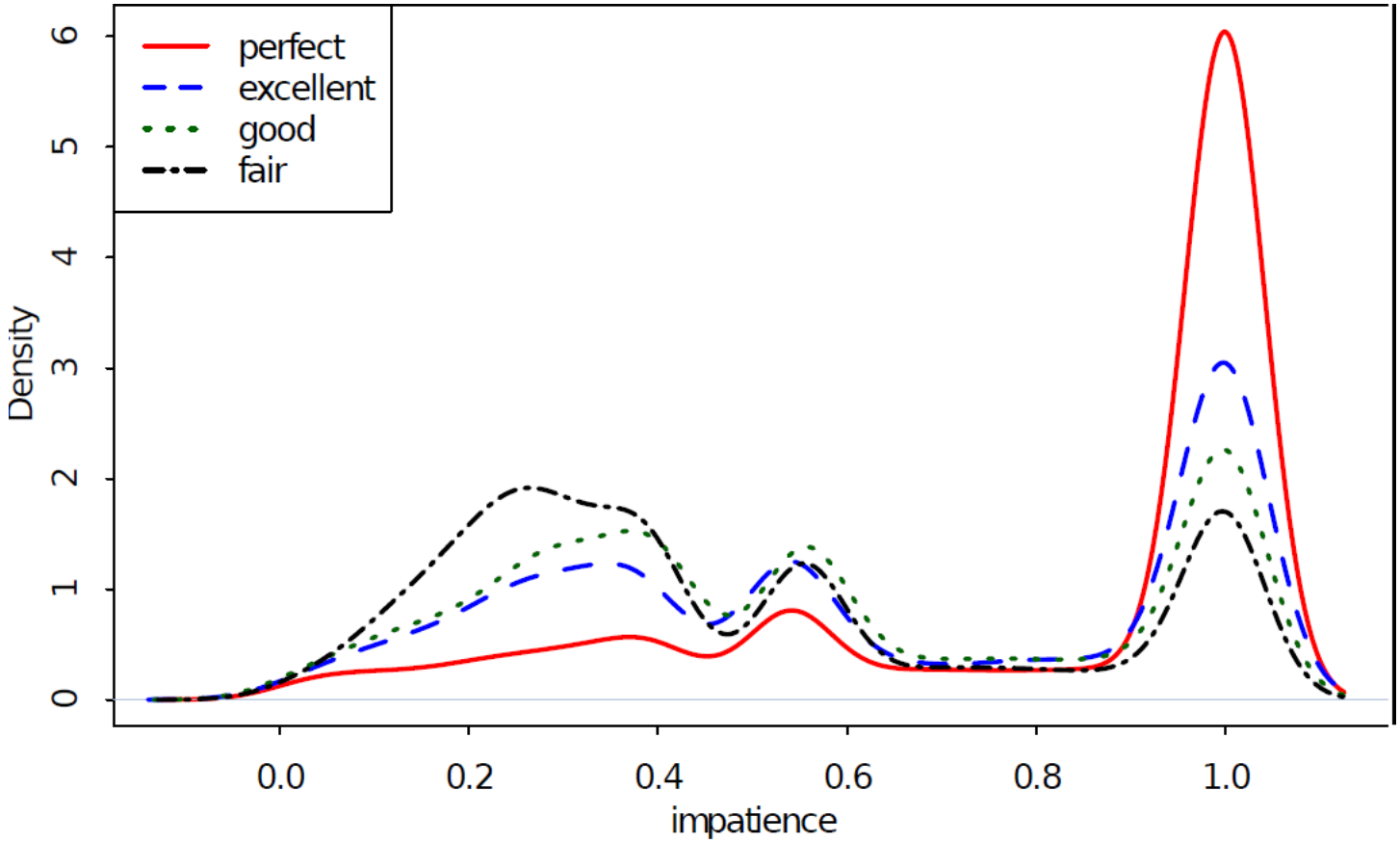
Distribution of RBP

- For some TREC system



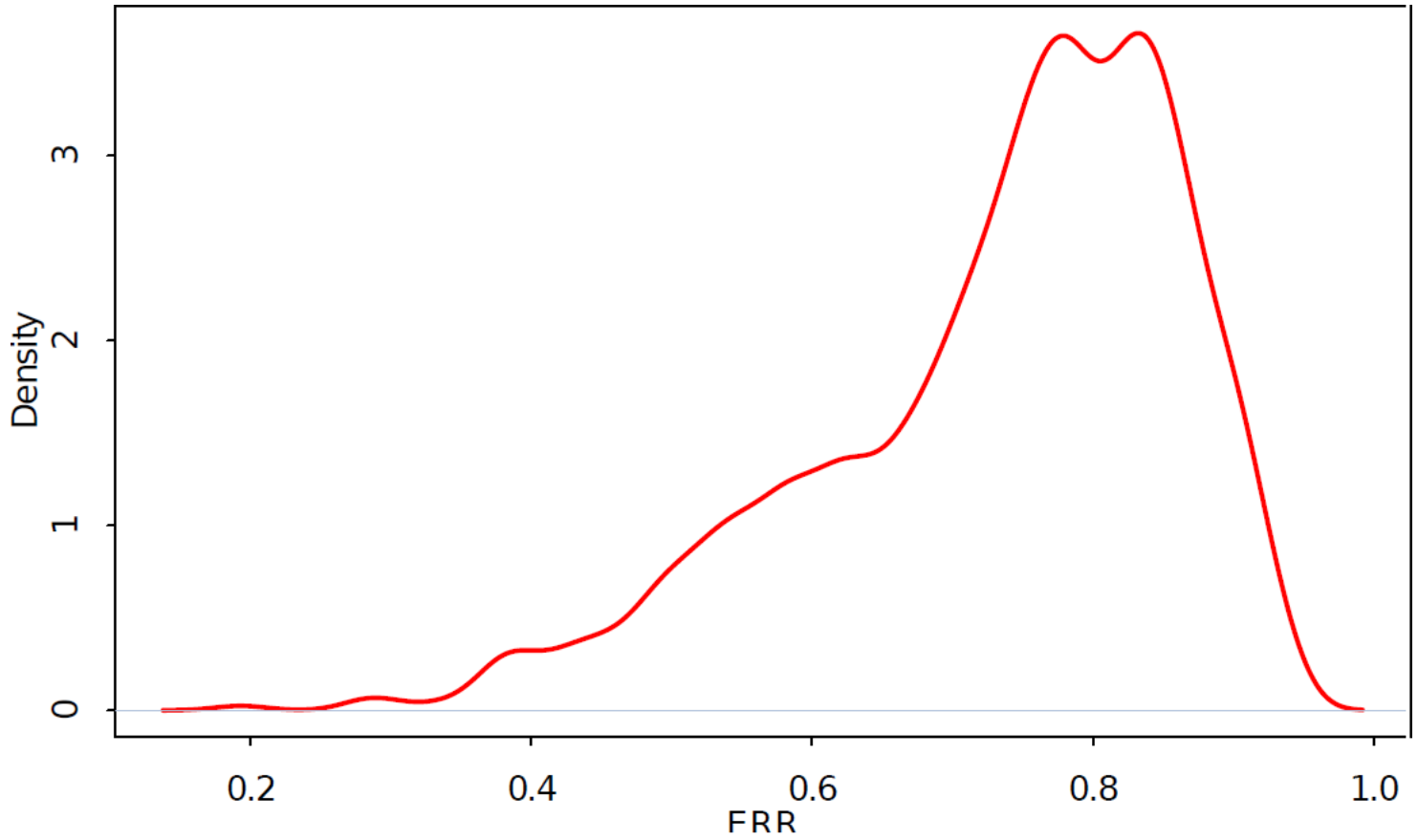
Posterior Distribution for Impatience: ERR

- Using a Bing log



Distribution of ERR

- For some TREC system



Marginal Distribution Analysis

- Given two systems, over all choices of θ
  - What is $P(M_1 > M_2)$?
  - What is $P((M_1 - M_2) > t)$?

Marginal Distribution Analysis

- Marginal distribution over values of a measure

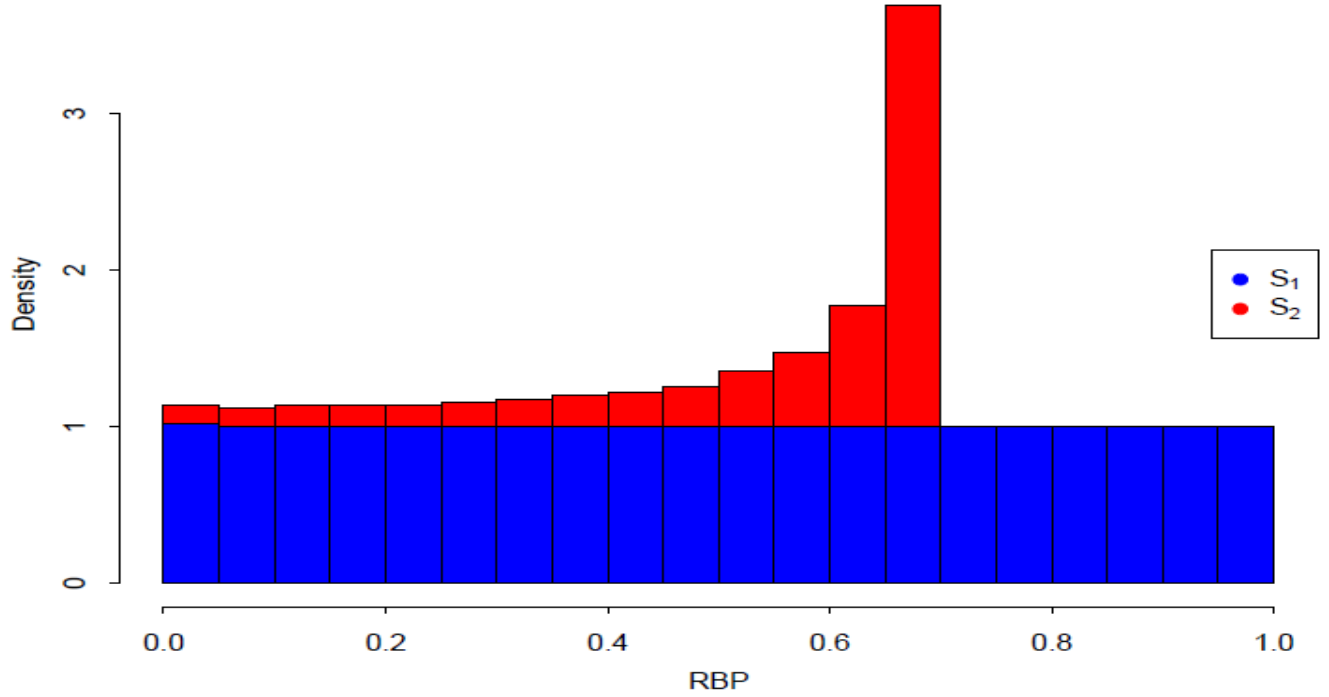
$$P(M = m) = \int P(M = m|\theta)P(\theta)d\theta$$

$$P(M = m \mid \theta) = \begin{cases} 1, \ if \ M = m \\ 0, \ o.w. \end{cases}$$

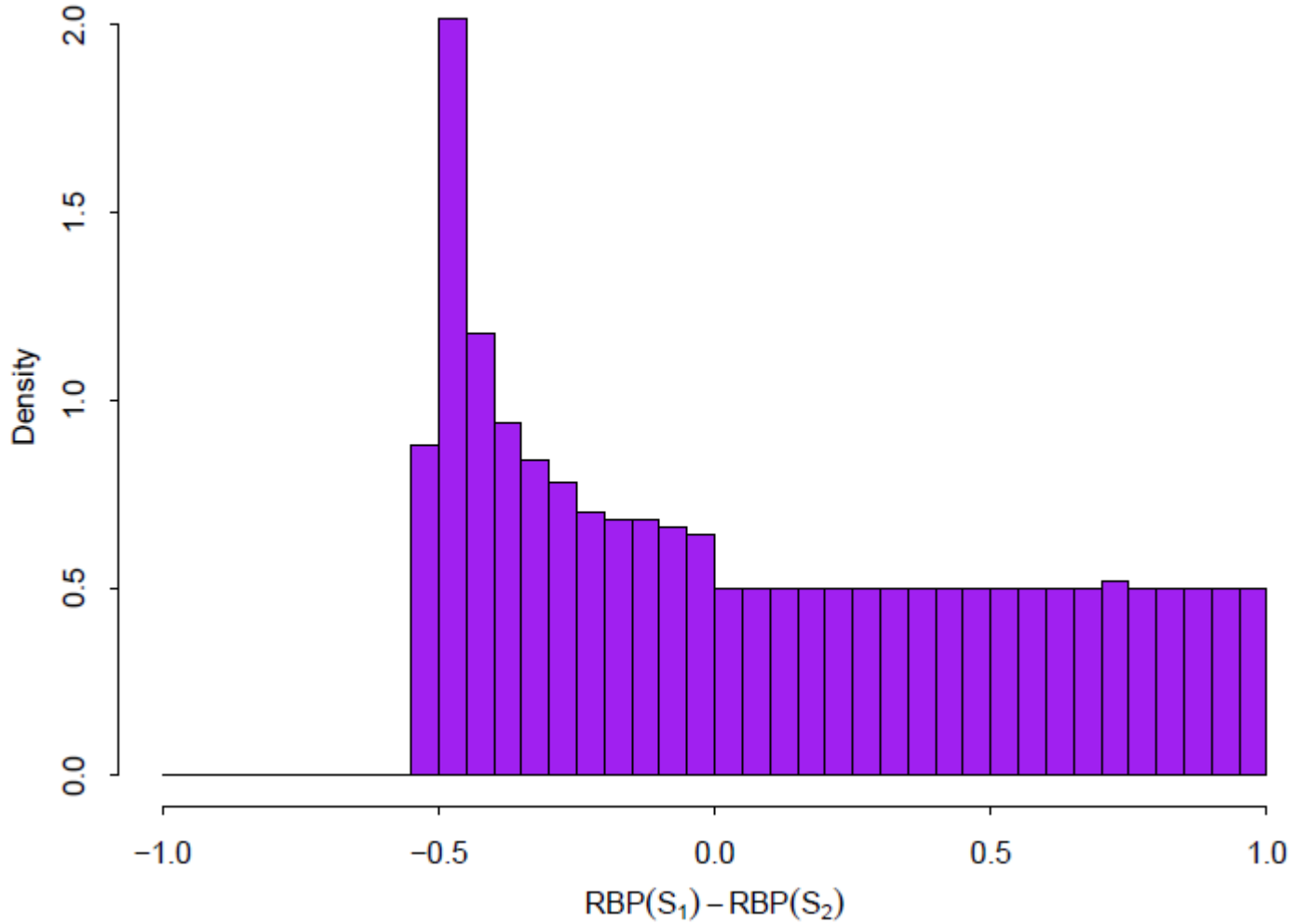
- Monte Carlo simulation to estimate the distribution
  - Sample values of θ, calculate M
  - Infer P(M) over many trials

Marginal Distribution Analysis

- Patience uniformly distributed
- S_1=[R N N N N N N N N N]
- S_2=[N R R R R R R R R R]



Marginal Distribution Analysis



Part 1: Evaluation Measures

- Evaluation measures
  1. Traditional evaluation measures
  2. Model-based measures
  3. Novelty and Diversity-based measures
  4. Session-based measures

- Evaluation of evaluation measures

european economic crisis

https://www.google.com/s…

EU employment guidelines can help member states confr…
TheParliamentMagazine.eu - Jun 15, 2015
Tackling unemployment in the **EU** will require more investment and ... In the short run, the social effects of the **crisis** need to be targeted in an ...

All **EU economies** set to expand for first time since **crisis**
Financial Times - May 27, 2015
All the economies of **EU** member states are set to expand this year for the first time since the **financial crisis** as a combination of cheap oil, the ...

Yahoo News

US Urges **European** Leaders to Solve Greek **Crisis** Quickly
New York Times - May 27, 2015
"Failure to agree on a path forward would create immediate hardship for Greece and broad uncertainties for **Europe** and the global **economy**," ...

• Ewald Nowotny: There is no ECB flexibility for Greece
In-Depth - Telegraph.co.uk - May 28, 2015

Explore in depth (1,407 more articles)

Greece, Germany and France agree to intensify talks – as…
The Guardian - Jun 9, 2015
European leaders realize #Greece needs a viable solution. ..... income, currency and commodities (FICC) markets ahead of the **financial crisis**.

If You Think Greece's **Crisis** Will End Soon, Think Again
Bloomberg - Jun 8, 2015

Explore in depth (4,249 more articles)

Berès: "I'm not convinced measures to tackle **crisis** have …
European Parliament (press release) - 17 hours ago
The question of how to resolve **Europe's economic** woes continues to divide the **European** Union. The **economic** committee approved on 16 ...

The Role of the **European** Parliament in Managing **Europ**…
Katoikos.eu (satire) (registration) (blog) - May 19, 2015
The political reaction to the sovereign debt **crisis** has focused on making **European economic** governance more efficient. But this has ...

Despite rise in growth, eurozone **economy** is still below 2008 level **...**
The Republic - May 20, 2015

Explore in depth (74 more articles)

Daily Express blames Greek **economic crisis** on **EU** and …
Left Foot Forward - Jun 1, 2015
On point 2: to blame migration from what he calls 'war-torn Somalia and Syria' for Greece's **economic crisis** is a hateful lie. People

Greek **crisis**: Merkel says a

https://news.google.com/news/stor…

Top Stories
News near you
Suggested for you
World
U.S.
Business
Technology
Entertainment
Sports
Science
Health
Spotlight

Search this story
european cris Go

Sorted by relevance
Sorted by date

Reset options

Full coverage

Greek **crisis**: Merkel says a deal is still possible - live updates
The Guardian - 34 minutes ago
Following Merkel's speech, Gregor Gysi, of the Left Party, launched a blistering attack on the German government's **eurozone** policies. Now Thomas Oppermann, who heads up the Social Democrat group in the Bundestag, has warned that "time is running out" ...

Greek **crisis**: Protesters demand end to austerity, with **EU** future in doubt - as it **...**
The Guardian - Jun 17, 2015
The warning came as the governor of the Greek central bank said that Greece was on the brink of an "uncontrollable **crisis**" and warned that leaving the **eurozone** would also mean "exit … most likely from the **European** Union". But the chances of a ...

Crisis deepens as Greek central bank warns of **euro** exit
Yahoo News - 14 hours ago
Athens (AFP) - The Greek central bank warned for the first time on Wednesday that the country could suffer a "painful" exit from the **eurozone** and even the **EU** if it fails to reach a bailout deal with international creditors.

Greek **crisis**: **Europe** has nothing to fear from Syriza's belligerence
The Guardian - Jun 16, 2015
In short, the ECB bond-buying programme has transformed the ECB from a passive observer of the **euro crisis**, paralysed by the outdated legalistic constraints of the Maastricht Treaty, into a proper lender of last resort. With powers to monetise government ...

Greece **crisis**: US urges compromise after Greek PM attacks IMF - as it happened
The Guardian - Jun 15, 2015
EU officials wereon Tuesday night making preparations for a **crisis** meeting of leaders on Sunday if, as now expected, the talks between finance ministers on Thursday prove fruitless. Amid the third straight day of sharp declines on the Athens stock market, **EU** ...

Grexit **crisis** deepens as Greek central bank warns of 'painful' **euro** exit
Sydney Morning Herald - 4 hours ago
The Greek central bank warned for the first time on Wednesday that the country could suffer a "painful" exit from the **eurozone** and even the **EU** if it fails to reach a bailout deal with international creditors. The warning came as negotiations over the release of ...

The Next Great **European** Financial **Crisis** Has Begun
ETF Daily News - Jun 16, 2015
europe The Greek financial system is in the process of totally imploding, and the rest of **Europe** will soon follow. Neither the Greeks nor the Germans are willing to give in, and that means that there is very little chance that a debt deal is going to happen by the ...

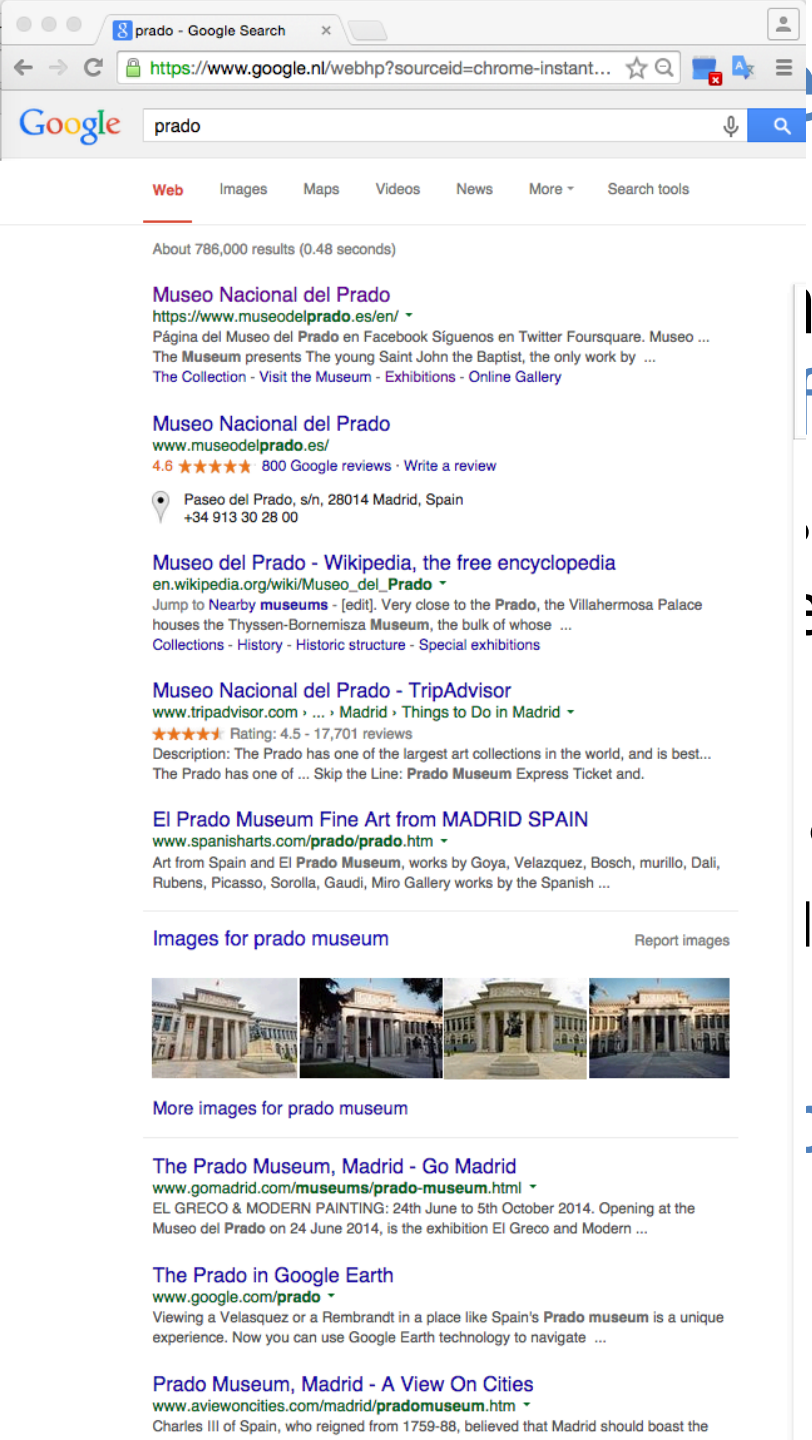
Could the **EU** fall apart completely?
BBC News - 8 hours ago
The **European** prime ministers and presidents who gather in Brussels

ersity?

ierently faceted

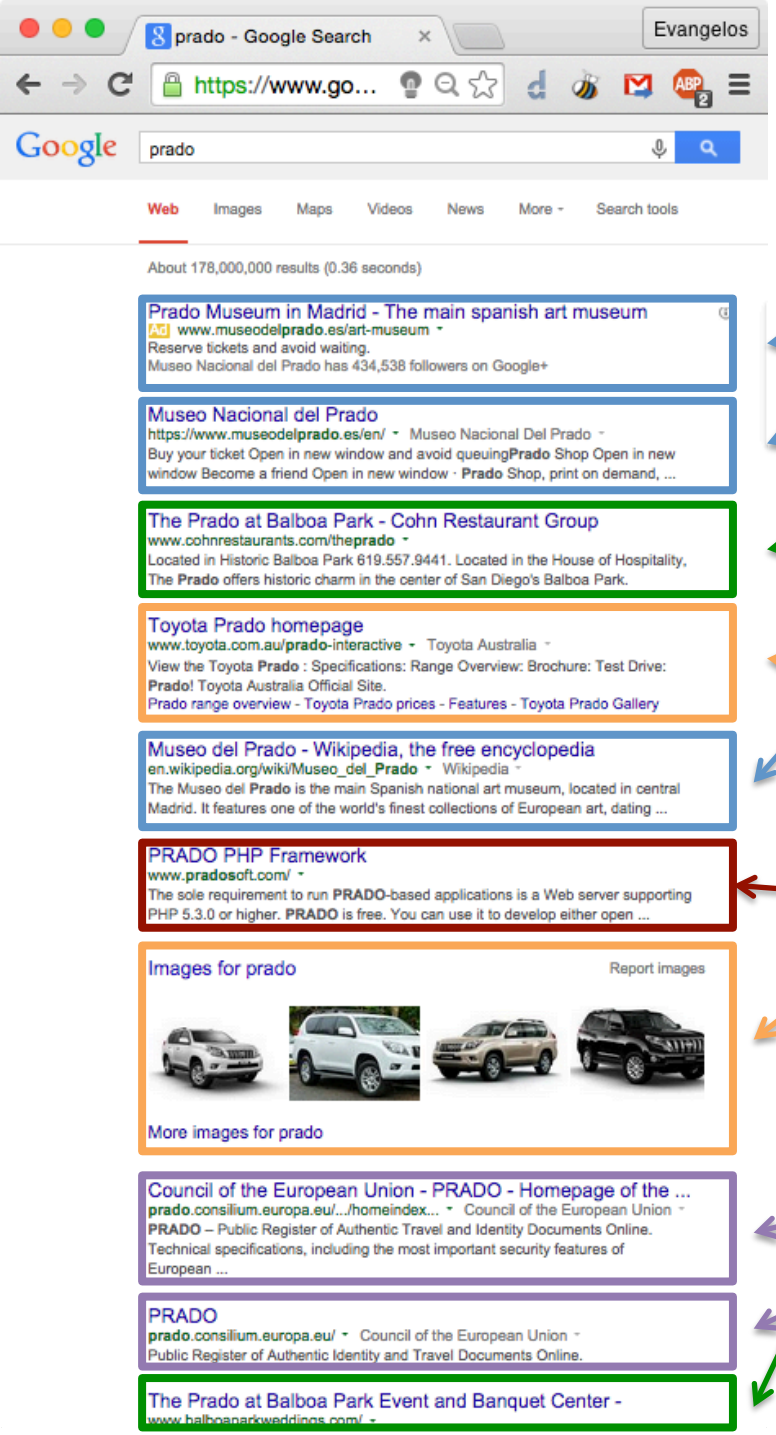
ystem can e user's

attempts lts that t to a ole



Intent-Aware Measures

- Assume there is a probability distribution *P(i|Q)* over intents for a query Q
  - Probability that a randomly-sampled user means intent i when submitting query Q
- The intent-aware version of a measure is its weighted average over this distribution



P(Prado Museum | Q) = 0.35

P(Prado Balboa | Q) = 0.10

P(Toyota Prado| Q) = 0.45

P(Prado PHP| Q) = 0.08

P(Prado EU| Q) = 0.02

Precision@10-IA = 0.35*0.3 + 0.10*0.2
+ 0.45*0.2 + 0.08*0.1 + 0.02*0.2 = 0.227

α-nDCG

- α-nDCG is a generalization of nDCG that accounts for both novelty and diversity
- α is a geometric penalization for redundancy
  - Redefine the gain of a document:
    - +1 for each intent it is relevant to
    - ×(1-α) for each document higher in the ranking that intent already appeared in
- Discount is the same as usual

P(Prado Museum | Q)

P(Prado Balboa | Q)

P(Toyota Prado| Q)

P(Prado PHP| Q)

P(Prado EU| Q)

Finding intents and probabilities

AOL-user-ct-collection — less — 116×53

```
710766  wwwpeoplesearch.comwww.reviewplace.seardh       2006-05-30 22:10:13
710766  wwwpeoplesearch.comwww.reviewplace.seardh       2006-05-30 22:10:33
711391  can not sleep with snoring husband      2006-03-01 01:24:00
711391  cannot sleep with snoring husband       2006-03-01 01:24:07     9       http://www.wjla.com
711391  cannot sleep with snoring husband       2006-03-01 01:24:07     9       http://www.wjla.com
711391  cannot sleep with snoring husband       2006-03-01 01:33:06     1       http://www.epinions.com
711391  jackie zeaman nude      2006-03-01 15:26:27
711391  jackie zeman nude       2006-03-01 15:26:38
711391  strange cosmos  2006-03-01 16:07:15     1       http://www.strangecosmos.com
711391  mansfield first assembly        2006-03-01 16:09:20     1       http://www.mansfieldfirstassembly.org
711391  mansfield first assembly        2006-03-01 16:09:20     3       http://netministries.org
711391  reverend harry myers    2006-03-01 16:10:07
711391  reverend harry myers    2006-03-01 16:10:30
711391  national enquirer       2006-03-01 17:13:14     1       http://www.nationalenquirer.com
711391  how to kill mockingbirds        2006-03-01 17:18:11
711391  how to kill mockingbirds        2006-03-01 17:18:33
711391  how to kill annoying birds in your yards        2006-03-01 17:18:58
711391  how to kill annoying birds in your yards        2006-03-01 17:19:53     2       http://www.sortprice.com
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:23:08     3       http://shopping.msn.com
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:23:08     10      http://www.bergen.org
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:24:35     15      http://www.saferbrand.com
711391  how do i get mocking birds out of my yard       2006-03-01 17:27:17
711391  how do i get mockingbirds out of my yard        2006-03-01 17:27:36     9       http://www.asri.org
711391  how do i get mockingbirds out of my yard        2006-03-01 17:30:14
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     3       http://www.bird-x.com
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     1       http://forums2.gardenweb.com
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     10      http://www.birding.com
711391  mansfield first assembly        2006-03-01 18:31:36     3       http://netministries.org
711391  beth moore      2006-03-01 19:42:41     1       http://www.lproof.org
711391  judy baker ministries   2006-03-01 19:49:03     2       http://www.embracinggrace.com
711391  god will fulfill your hearts desires    2006-03-01 19:59:06     10      http://www.pureintimacy.org
711391  online friendships can be very special  2006-03-01 23:09:37
711391  online friendships can be very special  2006-03-01 23:09:57
711391  online friendships      2006-03-01 23:10:24
711391  cypress fairbanks isd   2006-03-02 07:56:53     1       http://www.cfisd.net
711391  people are not always how they seem over the internet   2006-03-02 08:31:51
711391  friends online can be different in person       2006-03-02 08:32:42
711391  friends online can be different in person       2006-03-02 08:33:04     13      http://www.salon.com
711391  boston butts    2006-03-02 09:47:36
711391  community christian church houston tx   2006-03-02 16:07:53
711391  gay churches in houston tx      2006-03-02 16:08:23
711391  community gospel church in houston tx   2006-03-02 16:08:45     2       http://www.communitygospel.org
711391  houston tx is one hot place     2006-03-02 18:04:44
711391  houston tx is one hot place to live     2006-03-02 18:04:55     9       http://travel.yahoo.com
711391  houston tx is one hot place to live     2006-03-02 18:16:05     1       http://www.houston-texas-online.com
711391  texas hill country and sights around san antonio tx     2006-03-02 18:19:00     5       http://www.answers.c
om
711391  can liver problems cause you to loose your hair 2006-03-02 18:27:04
711391  can liver problems cause you to loose your hair 2006-03-02 18:27:30     1       http://www.askdoctrish.com
711391  strange cosmos  2006-03-02 19:29:31     1       http://www.strangecosmos.com
711391  white hard dry skin on face     2006-03-02 20:31:29
711391  white hard dry skin on face     2006-03-02 20:32:24
:
```

Inferring Query Intent from Reformulations and Clicks

Filip Radlinski
Microsoft Research
Cambridge, UK
filiprad@microsoft.com

Martin Szummer
Microsoft Research
Cambridge, UK
szummer@microsoft.com

Nick Craswell
Microsoft
Redmond, WA, USA
nickcr@microsoft.com

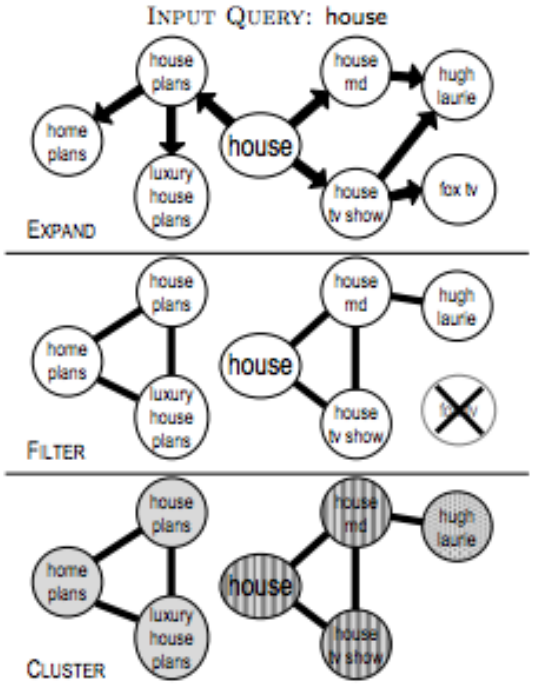


Figure 1: Three Steps of Intent Identification

Finding intents and probabilities

AOL-user-ct-collection — less — 116×53

```
710766  wwwpeoplesearch.comwww.reviewplace.seardh       2006-05-30 22:10:13
710766  wwwpeoplesearch.comwww.reviewplace.seardh       2006-05-30 22:10:33
711391  can not sleep with snoring husband      2006-03-01 01:24:00
711391  cannot sleep with snoring husband       2006-03-01 01:24:07     9       http://www.wjla.com
711391  cannot sleep with snoring husband       2006-03-01 01:24:07     9       http://www.wjla.com
711391  cannot sleep with snoring husband       2006-03-01 01:33:06     1       http://www.epinions.com
711391  jackie zeaman nude      2006-03-01 15:26:27
711391  jackie zeman nude       2006-03-01 15:26:38
711391  strange cosmos  2006-03-01 16:07:15     1       http://www.strangecosmos.com
711391  mansfield first assembly        2006-03-01 16:09:20     1       http://www.mansfieldfirstassembly.org
711391  mansfield first assembly        2006-03-01 16:09:20     3       http://netministries.org
711391  reverend harry myers    2006-03-01 16:10:07
711391  reverend harry myers    2006-03-01 16:10:30
711391  national enquirer       2006-03-01 17:13:14     1       http://www.nationalenquirer.com
711391  how to kill mockingbirds        2006-03-01 17:18:11
711391  how to kill mockingbirds        2006-03-01 17:18:33
711391  how to kill annoying birds in your yards        2006-03-01 17:18:58
711391  how to kill annoying birds in your yards        2006-03-01 17:19:53     2       http://www.sortprice.com
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:23:08     3       http://shopping.msn.com
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:23:08     10      http://www.bergen.org
711391  how to rid your yard of noisy annoying birds    2006-03-01 17:24:35     15      http://www.saferbrand.com
711391  how do i get mocking birds out of my yard       2006-03-01 17:27:17
711391  how do i get mockingbirds out of my yard        2006-03-01 17:27:36     9       http://www.asri.org
711391  how do i get mockingbirds out of my yard        2006-03-01 17:30:14
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     3       http://www.bird-x.com
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     1       http://forums2.gardenweb.com
711391  how to get rid of noisy loud birds      2006-03-01 17:30:52     10      http://www.birding.com
711391  mansfield first assembly        2006-03-01 18:31:36     3       http://netministries.org
711391  beth moore      2006-03-01 19:42:41     1       http://www.lproof.org
711391  judy baker ministries   2006-03-01 19:49:03     2       http://www.embracinggrace.com
711391  god will fulfill your hearts desires    2006-03-01 19:59:06     10      http://www.pureintimacy.org
711391  online friendships can be very special  2006-03-01 23:09:37
711391  online friendships can be very special  2006-03-01 23:09:57
711391  online friendships      2006-03-01 23:10:24
711391  cypress fairbanks isd   2006-03-02 07:56:53     1       http://www.cfisd.net
711391  people are not always how they seem over the internet   2006-03-02 08:31:51
711391  friends online can be different in person       2006-03-02 08:32:42
711391  friends online can be different in person       2006-03-02 08:33:04     13      http://www.salon.com
711391  boston butts    2006-03-02 09:47:36
711391  community christian church houston tx   2006-03-02 16:07:53
711391  gay churches in houston tx      2006-03-02 16:08:23
711391  community gospel church in houston tx   2006-03-02 16:08:45     2       http://www.communitygospel.org
711391  houston tx is one hot place     2006-03-02 18:04:44
711391  houston tx is one hot place to live     2006-03-02 18:04:55     9       http://travel.yahoo.com
711391  houston tx is one hot place to live     2006-03-02 18:16:05     1       http://www.houston-texas-online.com
711391  texas hill country and sights around san antonio tx     2006-03-02 18:19:00     5       http://www.answers.c
om
711391  can liver problems cause you to loose your hair 2006-03-02 18:27:04
711391  can liver problems cause you to loose your hair 2006-03-02 18:27:30     1       http://www.askdoctrish.com
711391  strange cosmos  2006-03-02 19:29:31     1       http://www.strangecosmos.com
711391  white hard dry skin on face     2006-03-02 20:31:29
711391  white hard dry skin on face     2006-03-02 20:32:24
:
```

Random Walks on the Click Graph

Nick Craswell and Martin Szummer
Microsoft Research Cambridge
7 JJ Thomson Ave
Cambridge, UK
{nickcr,szummer}@microsoft.com

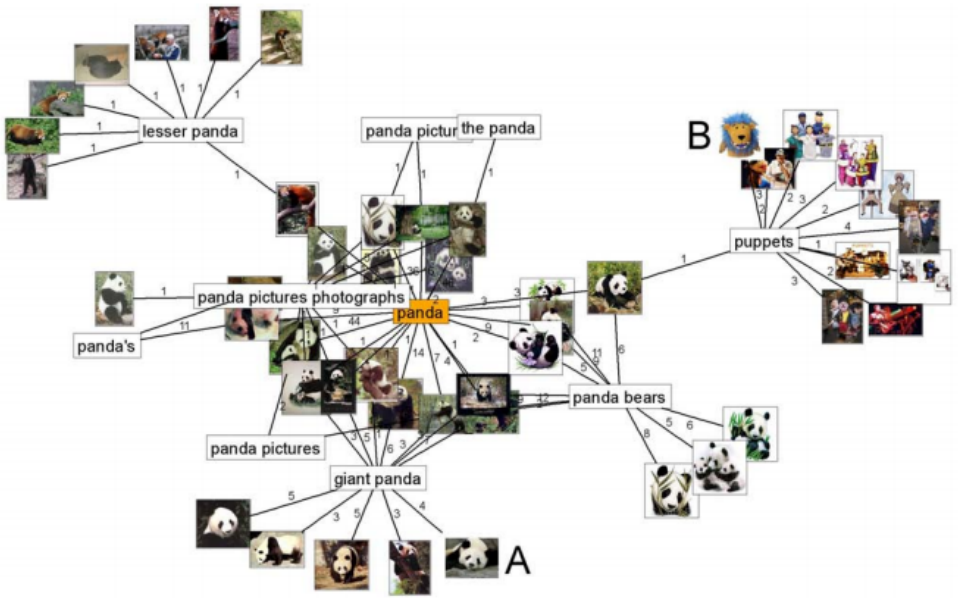


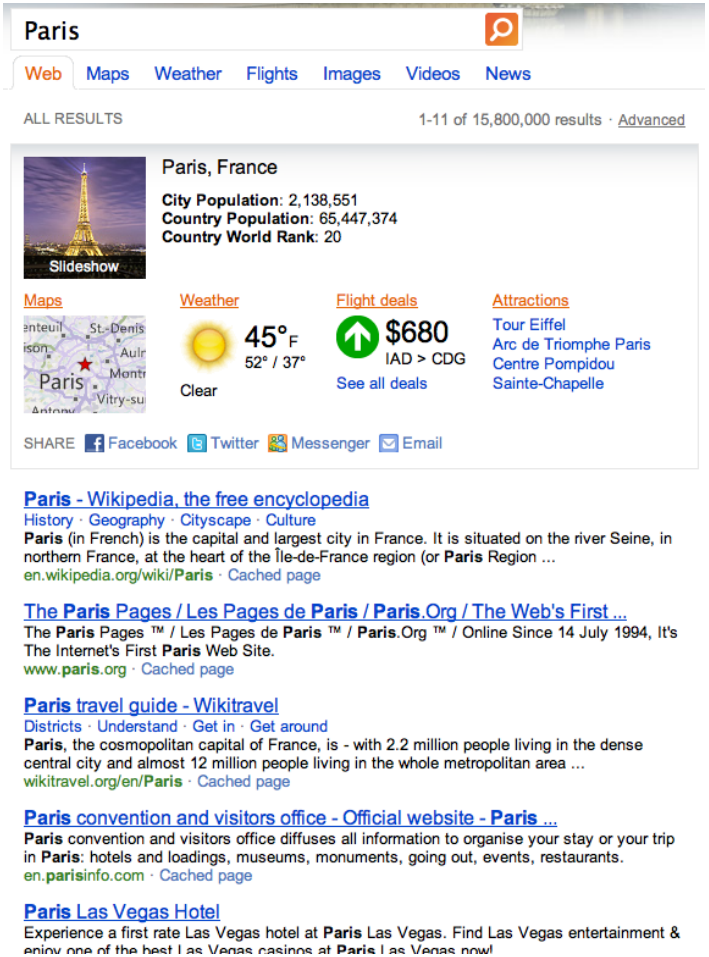
Figure 1: Click graph. Nodes are queries or images, edges indicate clicks. Images A and B are equidistant from the query 'panda' (distance=3), so retrieval based on a naïve shortest-path algorithm could not distinguish them. Our Markov random walk approach sums over paths, so image A benefits from having 7 distinct paths of length 3. In other words, nodes A and 'panda' are connected by a large "volume" of paths.

Part 1: Evaluation Measures

- Evaluation measures
  1. Traditional evaluation measures
  2. Model-based measures
  3. Novelty and Diversity-based measures
  4. Session-based measures

- Evaluation of evaluation measures

Paris

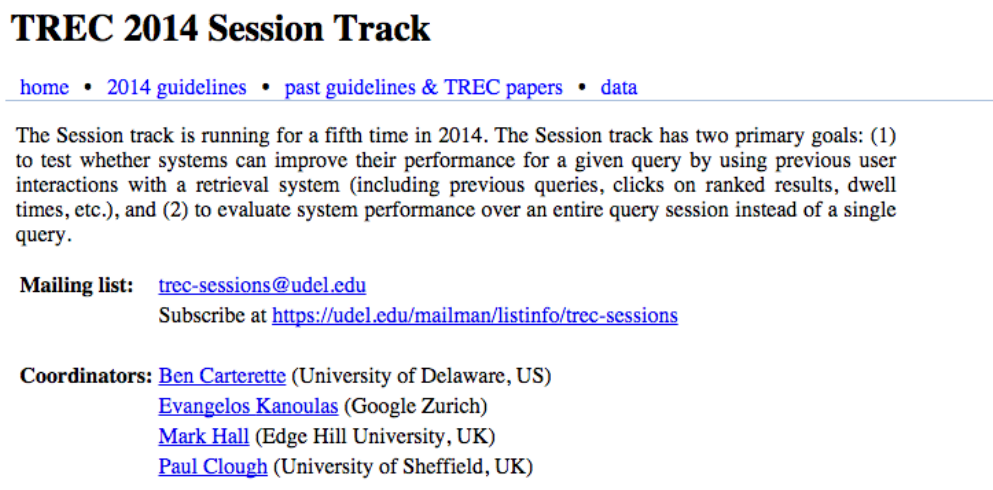

Paris
Web Maps Weather Flights Images Videos News
ALL RESULTS
1-11 of 15,800,000 results · Advanced
Slideshow
Paris, France
City Population: 2,138,551
Country Population: 65,447,374
Country World Rank: 20
Maps
Weather
45°F
52° / 37°
Clear
Flight deals
$680
IAD > CDG
See all deals
Attractions
Tour Eiffel
Arc de Triomphe Paris
Centre Pompidou
Sainte-Chapelle
SHARE Facebook Twitter Messenger Email
Paris - Wikipedia, the free encyclopedia
History · Geography · Cityscape · Culture
Paris (in French) is the capital and largest city in France. It is situated on the river Seine, in northern France, at the heart of the Île-de-France region (or Paris Region ...
en.wikipedia.org/wiki/Paris · Cached page
The Paris Pages / Les Pages de Paris / Paris.Org / The Web's First ...
The Paris Pages ™ / Les Pages de Paris ™ / Paris.Org ™ / Online Since 14 July 1994, It's The Internet's First Paris Web Site.
www.paris.org · Cached page
Paris travel guide - Wikitravel
Districts · Understand · Get in · Get around
Paris, the cosmopolitan capital of France, is - with 2.2 million people living in the dense central city and almost 12 million people living in the whole metropolitan area ...
wikitravel.org/en/Paris · Cached page
Paris convention and visitors office - Official website - Paris ...
Paris convention and visitors office diffuses all information to organise your stay or your trip in Paris: hotels and loadings, museums, monuments, going out, events, restaurants.
en.parisinfo.com · Cached page
Paris Las Vegas Hotel
Experience a first rate Las Vegas hotel at Paris Las Vegas. Find Las Vegas entertainment &

Session-based Evaluation

- Focus on serving the best results for a one-shot query
- Users frequently reformulate their initial query
- Questions:
  - Can we measure the effectiveness of a retrieval system over a sequence of reformulations (session)?
  - Can we optimize systems to provide better results over a session?

From single queries to sessions

- TREC Session Track 2010-2014
  - Goal:
    - Extend the evaluation framework from one query evaluation to multi-query session evaluation
    - Construct appropriate test collections
    - Establish new evaluation measures



TREC 2014 Session Track

home • 2014 guidelines • past guidelines & TREC papers • data

The Session track is running for a fifth time in 2014. The Session track has two primary goals: (1) to test whether systems can improve their performance for a given query by using previous user interactions with a retrieval system (including previous queries, clicks on ranked results, dwell times, etc.), and (2) to evaluate system performance over an entire query session instead of a single query.

Mailing list: trec-sessions@udel.edu
Subscribe at https://udel.edu/mailman/listinfo/trec-sessions

Coordinators: Ben Carterette (University of Delaware, US)
Evangelos Kanoulas (Google Zurich)
Mark Hall (Edge Hill University, UK)
Paul Clough (University of Sheffield, UK)

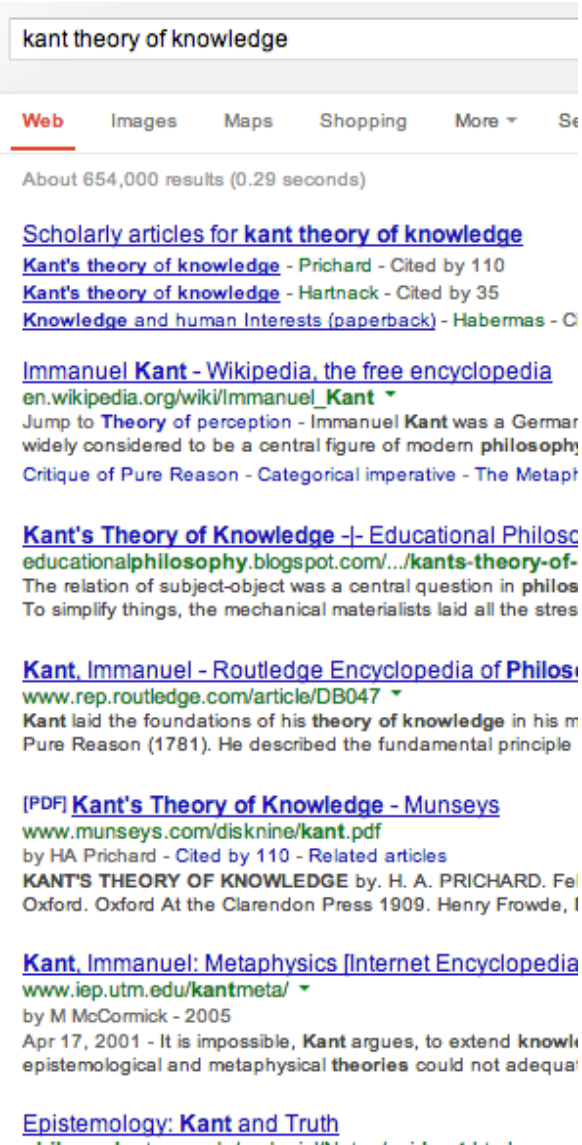
Construct appropriate test collections

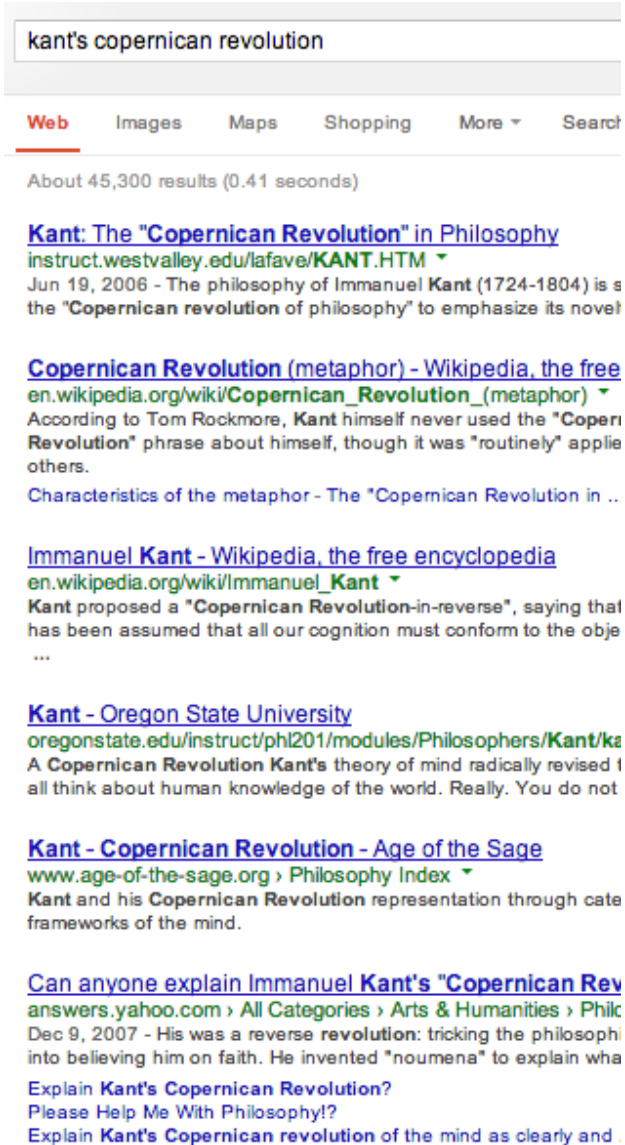
Rethink of evaluation measures

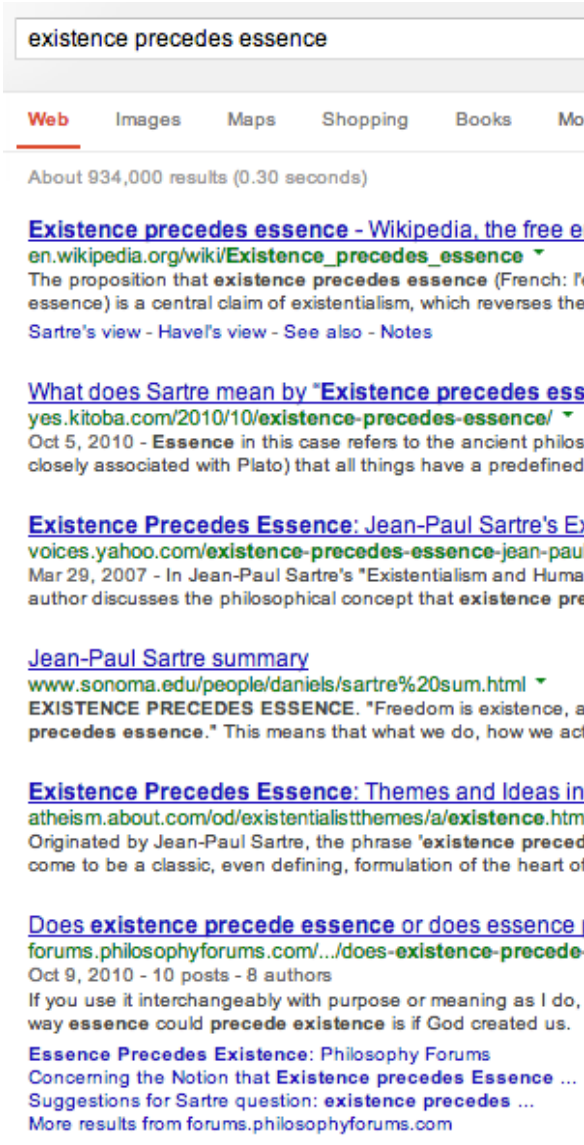
Construct appropriate test collections

Rethink of evaluation measures

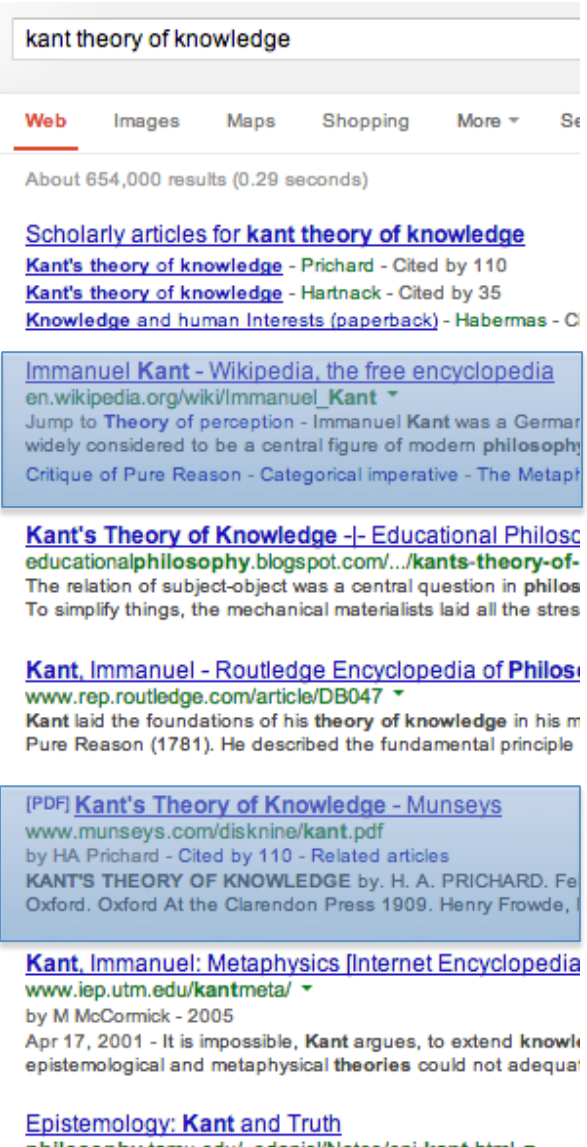
Experiment

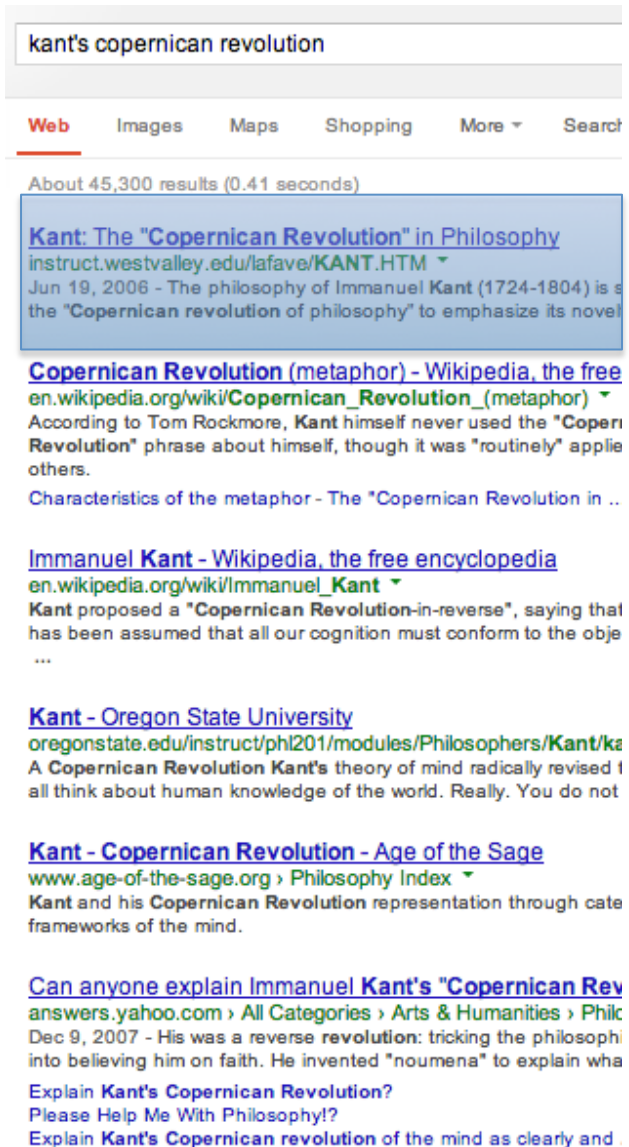

kant theory of knowledge
Web
Images
Maps
Shopping
More
About 654,000 results (0.29 seconds)
Scholarly articles for kant theory of knowledge
Kant's theory of knowledge - Prichard - Cited by 110
Kant's theory of knowledge - Hartnack - Cited by 35
Knowledge and human Interests (paperback) - Habermas
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Jump to Theory of perception - Immanuel Kant was a Germar
widely considered to be a central figure of modern philosophy
Critique of Pure Reason - Categorical imperative - The Metapt
Kant's Theory of Knowledge -|- Educational Philosc
educationalphilosophy.blogspot.com/.../kants-theory-of-
The relation of subject-object was a central question in philos
To simplify things, the mechanical materialists laid all the stres
Kant, Immanuel - Routledge Encyclopedia of Philos
www.rep.routledge.com/article/DB047
Kant laid the foundations of his theory of knowledge in his m
Pure Reason (1781). He described the fundamental principle
[PDF] Kant's Theory of Knowledge - Munseys
www.munseys.com/disknine/kant.pdf
by HA Prichard - Cited by 110 - Related articles
KANT'S THEORY OF KNOWLEDGE by. H. A. PRICHARD. Fe
Oxford. Oxford At the Clarendon Press 1909. Henry Frowde,
Kant, Immanuel: Metaphysics [Internet Encyclopedia
www.iep.utm.edu/kantmeta/
by M McCormick - 2005
Apr 17, 2001 - It is impossible, Kant argues, to extend knowl
epistemological and metaphysical theories could not adequat
Epistemology: Kant and Truth

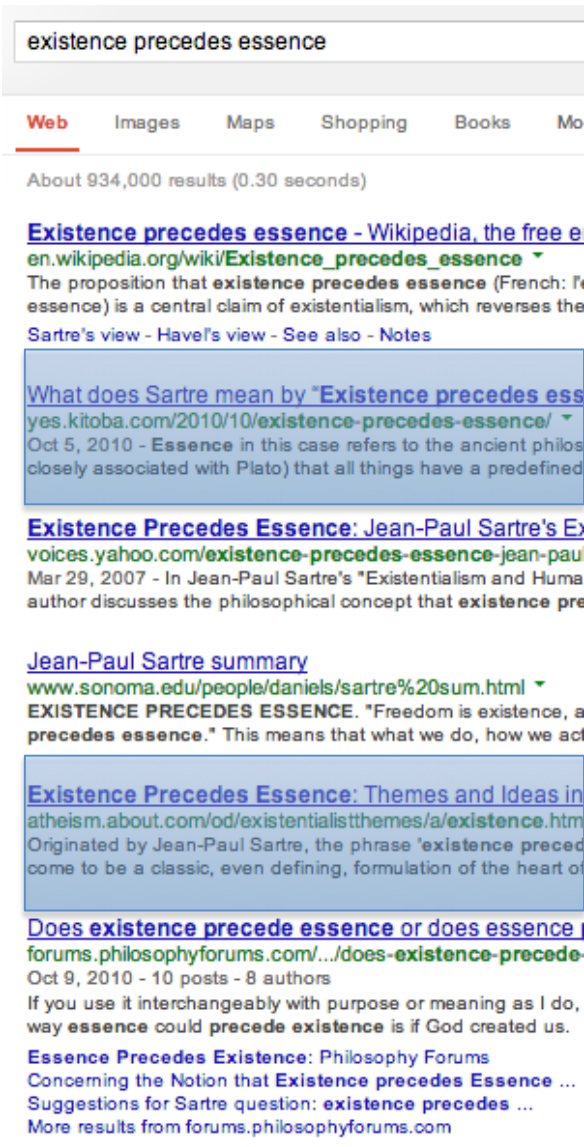

kant's copernican revolution
Web
Images
Maps
Shopping
More
Search
About 45,300 results (0.41 seconds)
Kant: The "Copernican Revolution" in Philosophy
instruct.westvalley.edu/lafave/KANT.HTM
Jun 19, 2006 - The philosophy of Immanuel Kant (1724-1804) is s
the "Copernican revolution of philosophy" to emphasize its novel
Copernican Revolution (metaphor) - Wikipedia, the free
en.wikipedia.org/wiki/Copernican_Revolution_(metaphor)
According to Tom Rockmore, Kant himself never used the "Coperr
Revolution" phrase about himself, though it was "routinely" applie
others.
Characteristics of the metaphor - The "Copernican Revolution in ..
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Kant proposed a "Copernican Revolution-in-reverse", saying that
has been assumed that all our cognition must conform to the obje
...
Kant - Oregon State University
oregonstate.edu/instruct/phl201/modules/Philosophers/Kant/ka
A Copernican Revolution Kant's theory of mind radically revised t
all think about human knowledge of the world. Really. You do not
Kant - Copernican Revolution - Age of the Sage
www.age-of-the-sage.org › Philosophy Index
Kant and his Copernican Revolution representation through cate
frameworks of the mind.
Can anyone explain Immanuel Kant's "Copernican Rev
answers.yahoo.com › All Categories › Arts & Humanities › Philc
Dec 9, 2007 - His was a reverse revolution: tricking the philosophi
into believing him on faith. He invented "noumena" to explain wha
Explain Kant's Copernican Revolution?
Please Help Me With Philosophy!?
Explain Kant's Copernican revolution of the mind as clearly and


existence precedes essence
Web
Images
Maps
Shopping
Books
About 934,000 results (0.30 seconds)
Existence precedes essence - Wikipedia, the free e
en.wikipedia.org/wiki/Existence_precedes_essence
The proposition that existence precedes essence (French:
essence) is a central claim of existentialism, which reverses the
Sartre's view - Havel's view - See also - Notes
What does Sartre mean by "Existence precedes ess
yes.kitoba.com/2010/10/existence-precedes-essence/
Oct 5, 2010 - Essence in this case refers to the ancient philos
closely associated with Plato) that all things have a predefined
Existence Precedes Essence: Jean-Paul Sartre's Ex
voices.yahoo.com/existence-precedes-essence-jean-paul
Mar 29, 2007 - In Jean-Paul Sartre's "Existentialism and Huma
author discusses the philosophical concept that existence pre
Jean-Paul Sartre summary
www.sonoma.edu/people/daniels/sartre%20sum.html
EXISTENCE PRECEDES ESSENCE. "Freedom is existence, a
precedes essence." This means that what we do, how we act
Existence Precedes Essence: Themes and Ideas in
atheism.about.com/od/existentialistthemes/a/existence.htm
Originated by Jean-Paul Sartre, the phrase 'existence preced
come to be a classic, even defining, formulation of the heart of
Does existence precede essence or does essence
forums.philosophyforums.com/.../does-existence-precede-
Oct 9, 2010 - 10 posts - 8 authors
If you use it interchangeably with purpose or meaning as I do,
way essence could precede existence is if God created us.
Essence Precedes Existence: Philosophy Forums
Concerning the Notion that Existence precedes Essence ...
Suggestions for Sartre question: existence precedes ...
More results from forums.philosophyforums.com

Experiment


kant theory of knowledge
Web
Images
Maps
Shopping
More
About 654,000 results (0.29 seconds)
Scholarly articles for kant theory of knowledge
Kant's theory of knowledge - Prichard - Cited by 110
Kant's theory of knowledge - Hartnack - Cited by 35
Knowledge and human Interests (paperback) - Habermas
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Jump to Theory of perception - Immanuel Kant was a Germar
widely considered to be a central figure of modern philosophy
Critique of Pure Reason - Categorical imperative - The Metaph
Kant's Theory of Knowledge -|- Educational Philosc
educationalphilosophy.blogspot.com/.../kants-theory-of-
The relation of subject-object was a central question in philos
To simplify things, the mechanical materialists laid all the stres
Kant, Immanuel - Routledge Encyclopedia of Philos
www.rep.routledge.com/article/DB047
Kant laid the foundations of his theory of knowledge in his m
Pure Reason (1781). He described the fundamental principle
[PDF] Kant's Theory of Knowledge - Munseys
www.munseys.com/disknine/kant.pdf
by HA Prichard - Cited by 110 - Related articles
KANT'S THEORY OF KNOWLEDGE by. H. A. PRICHARD. Fe
Oxford. Oxford At the Clarendon Press 1909. Henry Frowde,
Kant, Immanuel: Metaphysics [Internet Encyclopedia
www.iep.utm.edu/kantmeta/
by M McCormick - 2005
Apr 17, 2001 - It is impossible, Kant argues, to extend knowl
epistemological and metaphysical theories could not adequat
Epistemology: Kant and Truth


kant's copernican revolution
Web
Images
Maps
Shopping
More
Search
About 45,300 results (0.41 seconds)
Kant: The "Copernican Revolution" in Philosophy
instruct.westvalley.edu/lafave/KANT.HTM
Jun 19, 2006 - The philosophy of Immanuel Kant (1724-1804) is s
the "Copernican revolution of philosophy" to emphasize its novel
Copernican Revolution (metaphor) - Wikipedia, the free
en.wikipedia.org/wiki/Copernican_Revolution_(metaphor)
According to Tom Rockmore, Kant himself never used the "Coperi
Revolution" phrase about himself, though it was "routinely" applie
others.
Characteristics of the metaphor - The "Copernican Revolution in ..
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Kant proposed a "Copernican Revolution-in-reverse", saying that
has been assumed that all our cognition must conform to the obje
...
Kant - Oregon State University
oregonstate.edu/instruct/phl201/modules/Philosophers/Kant/ka
A Copernican Revolution Kant's theory of mind radically revised t
all think about human knowledge of the world. Really. You do not
Kant - Copernican Revolution - Age of the Sage
www.age-of-the-sage.org › Philosophy Index
Kant and his Copernican Revolution representation through cate
frameworks of the mind.
Can anyone explain Immanuel Kant's "Copernican Rev
answers.yahoo.com › All Categories › Arts & Humanities › Philc
Dec 9, 2007 - His was a reverse revolution: tricking the philosophi
into believing him on faith. He invented "noumena" to explain wha
Explain Kant's Copernican Revolution?
Please Help Me With Philosophy!?
Explain Kant's Copernican revolution of the mind as clearly and


existence precedes essence
Web
Images
Maps
Shopping
Books
About 934,000 results (0.30 seconds)
Existence precedes essence - Wikipedia, the free e
en.wikipedia.org/wiki/Existence_precedes_essence
The proposition that existence precedes essence (French: l'e
essence) is a central claim of existentialism, which reverses the
Sartre's view - Havel's view - See also - Notes
What does Sartre mean by "Existence precedes ess
yes.kitoba.com/2010/10/existence-precedes-essence/
Oct 5, 2010 - Essence in this case refers to the ancient philos
closely associated with Plato) that all things have a predefined
Existence Precedes Essence: Jean-Paul Sartre's Ex
voices.yahoo.com/existence-precedes-essence-jean-paul
Mar 29, 2007 - In Jean-Paul Sartre's "Existentialism and Huma
author discusses the philosophical concept that existence pre
Jean-Paul Sartre summary
www.sonoma.edu/people/daniels/sartre%20sum.html
EXISTENCE PRECEDES ESSENCE. "Freedom is existence, a
precedes essence." This means that what we do, how we act
Existence Precedes Essence: Themes and Ideas in
atheism.about.com/od/existentialistthemes/a/existence.htm
Originated by Jean-Paul Sartre, the phrase 'existence preced
come to be a classic, even defining, formulation of the heart of
Does existence precede essence or does essence
forums.philosophyforums.com/.../does-existence-precede-
Oct 9, 2010 - 10 posts - 8 authors
If you use it interchangeably with purpose or meaning as I do,
way essence could precede existence is if God created us.
Essence Precedes Existence: Philosophy Forums
Concerning the Notion that Existence precedes Essence ...
Suggestions for Sartre question: existence precedes ...
More results from forums.philosophyforums.com

Simulation

ClueWeb09 Category B

kant theory of knowledge Search

Searched for: **kant theory of knowledge**

1. Knowledge, Theory of | LibraryThing
(clueweb09-en0010-54-07918) http://www.librarything.com/subject/Knowledge%2C+Theory+of
Knowledge, Theory of | LibraryThing **...**46) empiricism(134) enlightenment(98) epistemology(1,052) german idealism **kant**(149) **knowledge**(105) locke(55) logic(109) logical positivism(34) metaphysics(231) modern philosophy(143) phi
[cached version]

2. Term Papers on Philosophy - Term Papers Lab
(clueweb09-en0011-00-11977) http://www.termpaperslab.com/term-papers/philosophy/k.html
...feature protagonists in situations out of which arise existentialist values. Existentialism is a philosoph... Save **Kant** Ca Immanuel **Kant** By: Yonna Yelverton... Save **Kant Kant** starts off making two distinctions regarding kinds of **knowle** formal/material. Empirical or experience-based **knowledge** is contrasted with ration... Save **Kant** Compare Mill and **K** better societal order? John Stuart Mill (1808-73) believed in an ethical**...**
[cached version]

3. Educational Theory of Immanuel Kant
(clueweb09-en0011-06-15131) http://www.newfoundations.com/GALLERY/Kant.html
Educational **Theory** of Immanuel **Kant** ©2000 NewFoundations The Educational **Theory** of Immanuel **Kant** Analys 1/4/08 I. **Theory** of Value What **knowledge** and skills are worthwhile learning? What are the goals of education Meta therefore as a stimulus to undertake a critical investigation of metaphysics (Copleston 212) II. **Theory** of **Knowledge** different from belief? The belief that knowledge is intrinsically**...**
[cached version]

4. Baruch Undergrad. Honors [1993]: White, J. Defense of Immanuel Kant's...
(clueweb09-en0002-08-24287) http://newman.baruch.cuny.edu/digital/2000/honors/white_1993.htm
Baruch Undergrad. Honors [1993]: White, J. Defense of Immanuel **Kant's**... DEFENSE OF IMMANUEL KANT'S REASON by JULIAN WHITE© DECEMBER 6, 1993 Submitted**...**that has been subjected to strong critiques which Alternatives to **Kant's theory** are problematic, however. Utilitarian philosophers, like David Hume, claim that moral st
[cached version]

5. The Principles of Friesian Philosophy
(clueweb09-en0008-53-11127) http://www.friesian.com/friesian.htm
...philosophy who forgets that philosophy is about wisdom may still be a real lover of **knowledge**, may still be a great a very good**...**Method. But while Nelson's theory of Socratic Method as "abstraction" is intuitionistic and inadequate, th terms of Popper's **theory** of falsification, which winnows out inconsistent items. Popper's **theory** of scientific method, generated, may thus be coupled with the Socratic injunction, "Say what you believe," to recover what **Kant** called the The theory of non-intuitive immediate knowledge frees the Kantian epistemology of synthetic a priori**...**
[cached version]

Epistemology: **Kant** and Truth

Construct appropriate test collections

Rethink of evaluation measures

What is a good system?

How can we measure “goodness”?

Measuring “goodness”

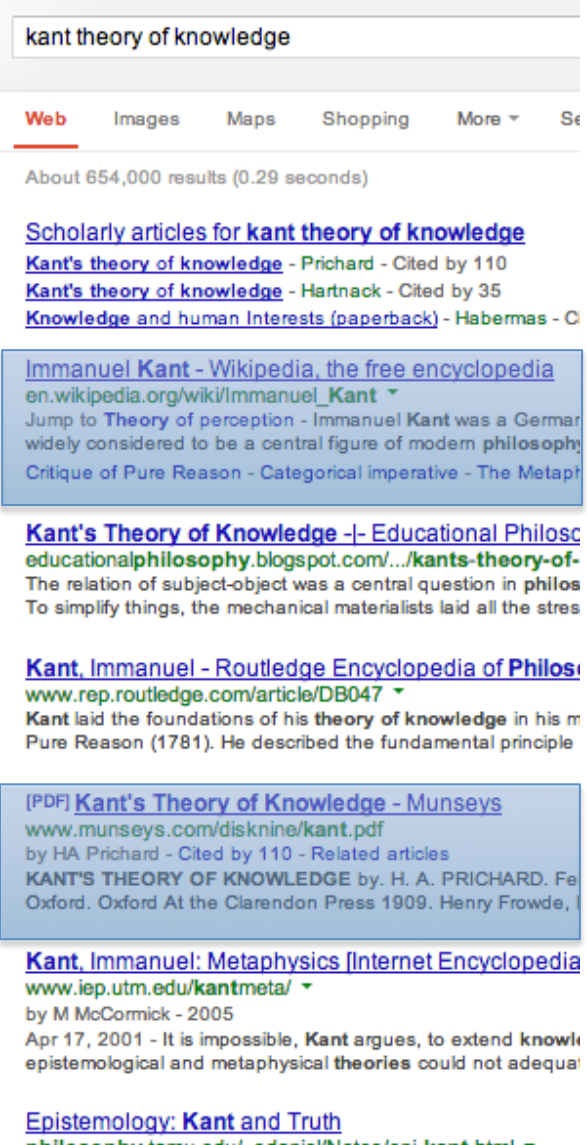
The user steps down a ranked list of documents, observes each one of them until a decision point and either

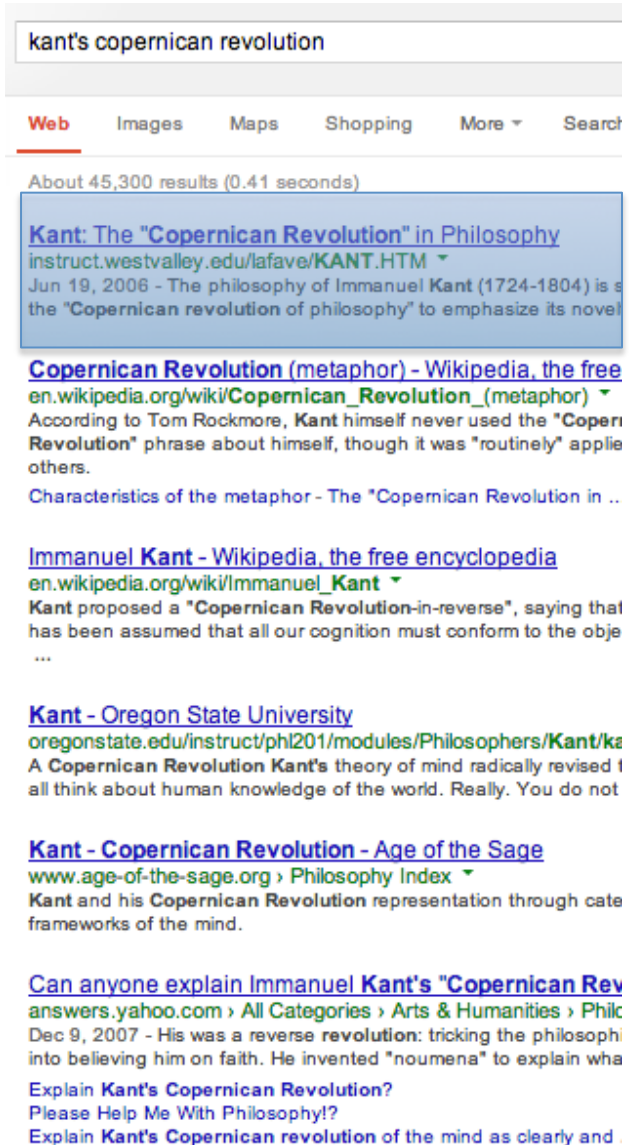
a) abandons the search, or

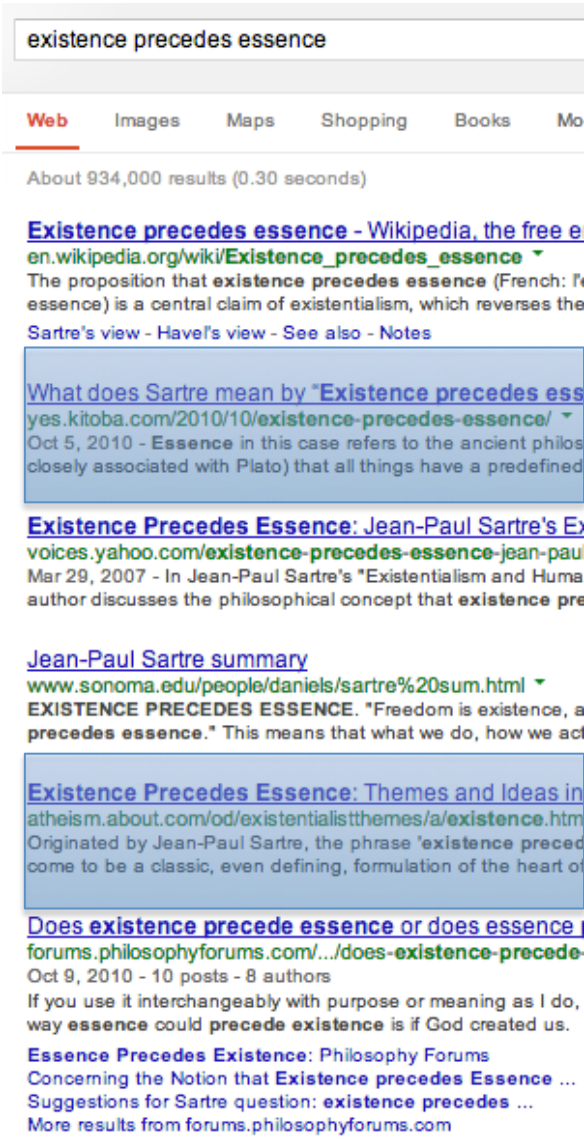
b) reformulates

While stepping down or sideways, the user accumulates utility

Experiment


kant theory of knowledge
Web
Images
Maps
Shopping
More
About 654,000 results (0.29 seconds)
Scholarly articles for kant theory of knowledge
Kant's theory of knowledge - Prichard - Cited by 110
Kant's theory of knowledge - Hartnack - Cited by 35
Knowledge and human Interests (paperback) - Habermas - C
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Jump to Theory of perception - Immanuel Kant was a Germar
widely considered to be a central figure of modern philosophy
Critique of Pure Reason - Categorical imperative - The Metapl
Kant's Theory of Knowledge -|- Educational Philosc
educationalphilosophy.blogspot.com/.../kants-theory-of-
The relation of subject-object was a central question in philos
To simplify things, the mechanical materialists laid all the stres
Kant, Immanuel - Routledge Encyclopedia of Philos
www.rep.routledge.com/article/DB047
Kant laid the foundations of his theory of knowledge in his m
Pure Reason (1781). He described the fundamental principle
[PDF] Kant's Theory of Knowledge - Munseys
www.munseys.com/disknine/kant.pdf
by HA Prichard - Cited by 110 - Related articles
KANT'S THEORY OF KNOWLEDGE by. H. A. PRICHARD. Fe
Oxford. Oxford At the Clarendon Press 1909. Henry Frowde,
Kant, Immanuel: Metaphysics [Internet Encyclopedia
www.iep.utm.edu/kantmeta/
by M McCormick - 2005
Apr 17, 2001 - It is impossible, Kant argues, to extend knowl
epistemological and metaphysical theories could not adequat
Epistemology: Kant and Truth


kant's copernican revolution
Web
Images
Maps
Shopping
More
About 45,300 results (0.41 seconds)
Kant: The "Copernican Revolution" in Philosophy
instruct.westvalley.edu/lafave/KANT.HTM
Jun 19, 2006 - The philosophy of Immanuel Kant (1724-1804) is s
the "Copernican revolution of philosophy" to emphasize its novel
Copernican Revolution (metaphor) - Wikipedia, the free
en.wikipedia.org/wiki/Copernican_Revolution_(metaphor)
According to Tom Rockmore, Kant himself never used the "Coperi
Revolution" phrase about himself, though it was "routinely" applie
others.
Characteristics of the metaphor - The "Copernican Revolution in ..
Immanuel Kant - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Immanuel_Kant
Kant proposed a "Copernican Revolution-in-reverse", saying that
has been assumed that all our cognition must conform to the obje
...
Kant - Oregon State University
oregonstate.edu/instruct/phl201/modules/Philosophers/Kant/ka
A Copernican Revolution Kant's theory of mind radically revised t
all think about human knowledge of the world. Really. You do not
Kant - Copernican Revolution - Age of the Sage
www.age-of-the-sage.org › Philosophy Index
Kant and his Copernican Revolution representation through cate
frameworks of the mind.
Can anyone explain Immanuel Kant's "Copernican Rev
answers.yahoo.com › All Categories › Arts & Humanities › Philc
Dec 9, 2007 - His was a reverse revolution: tricking the philosophi
into believing him on faith. He invented "noumena" to explain wha
Explain Kant's Copernican Revolution?
Please Help Me With Philosophy!?
Explain Kant's Copernican revolution of the mind as clearly and


existence precedes essence
Web
Images
Maps
Shopping
Books
About 934,000 results (0.30 seconds)
Existence precedes essence - Wikipedia, the free e
en.wikipedia.org/wiki/Existence_precedes_essence
The proposition that existence precedes essence (French: l'e
essence) is a central claim of existentialism, which reverses the
Sartre's view - Havel's view - See also - Notes
What does Sartre mean by "Existence precedes ess
yes.kitoba.com/2010/10/existence-precedes-essence/
Oct 5, 2010 - Essence in this case refers to the ancient philos
closely associated with Plato) that all things have a predefined
Existence Precedes Essence: Jean-Paul Sartre's Ex
voices.yahoo.com/existence-precedes-essence-jean-paul
Mar 29, 2007 - In Jean-Paul Sartre's "Existentialism and Huma
author discusses the philosophical concept that existence pre
Jean-Paul Sartre summary
www.sonoma.edu/people/daniels/sartre%20sum.html
EXISTENCE PRECEDES ESSENCE. "Freedom is existence, a
precedes essence." This means that what we do, how we act
Existence Precedes Essence: Themes and Ideas in
atheism.about.com/od/existentialistthemes/a/existence.htm
Originated by Jean-Paul Sartre, the phrase 'existence preced
come to be a classic, even defining, formulation of the heart of
Does existence precede essence or does essence
forums.philosophyforums.com/.../does-existence-precede-
Oct 9, 2010 - 10 posts - 8 authors
If you use it interchangeably with purpose or meaning as I do,
way essence could precede existence is if God created us.
Essence Precedes Existence: Philosophy Forums
Concerning the Notion that Existence precedes Essence ...
Suggestions for Sartre question: existence precedes ...
More results from forums.philosophyforums.com

What are the challenges?

BERLIN
9 575 KM
TOKYO
14 724 KM
RIO DE JANEIRO
6 055 KM
JERUSALEM
7 468 KM
SOUTH POLE
NEW YORK
12 541 KM
NEW DELHI
9 296 KM

Session DCG

Discounted Cumulated Gain Based Evaluation of Multiple-Query IR Sessions

Kalervo Järvelin[1], Susan L. Price[2], Lois M.L. Delcambre[2], and Marianne Lykke Nielsen[3]

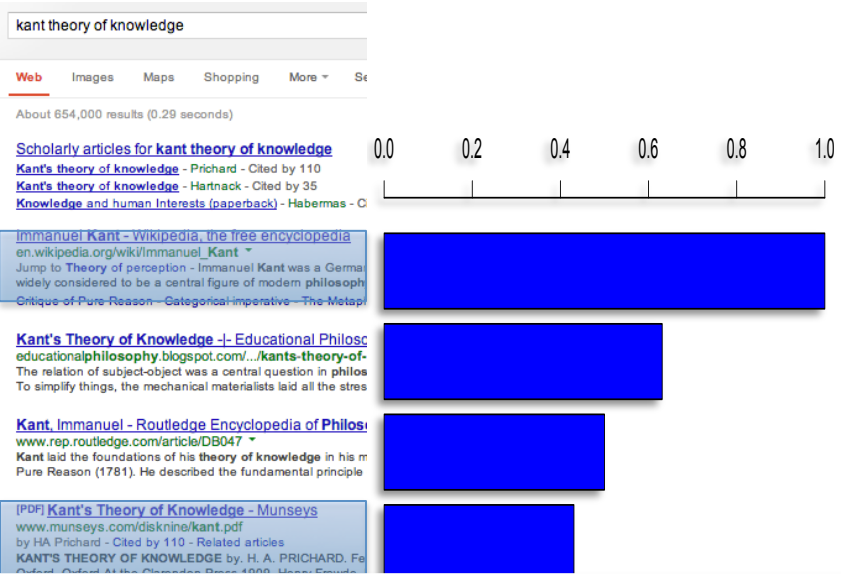
$$\sum_{r=1}^{k} \frac{2^{rel(r)} - 1}{\log_b(r + b - 1)}$$

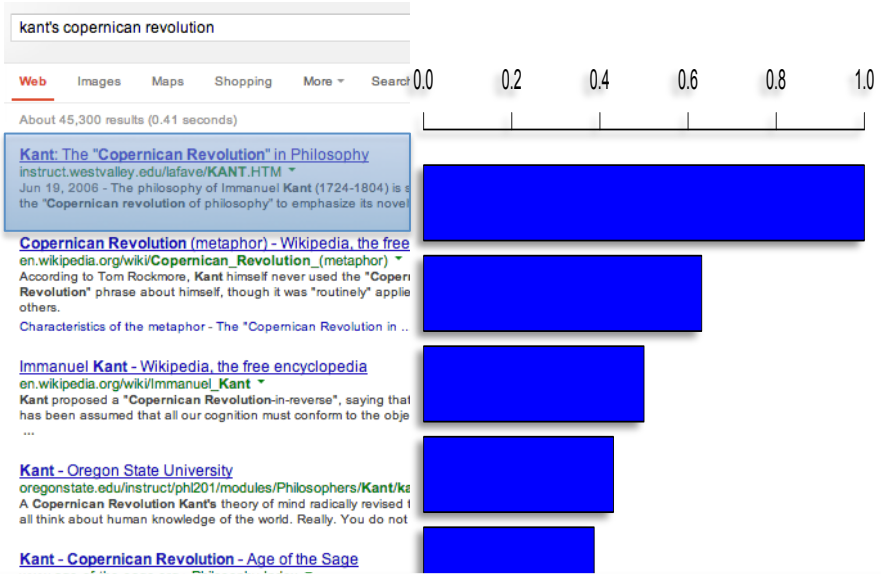
$$\sum_{r=1}^{k} \frac{2^{rel(r)} - 1}{\log_b(r + b - 1)}$$

Session DCG

Discounted Cumulated Gain Based Evaluation of Multiple-Query IR Sessions

Kalervo Järvelin[1], Susan L. Price[2], Lois M.L. Delcambre[2], and Marianne Lykke Nielsen[3]





$$\frac{1}{\log_c(1+c-1)}DCG(RL1)+\frac{1}{\log_c(2+c-1)}\cdot DCG(RL2)$$

Existing measures

- Session DCG [Järvelin et al ECIR 2008]

  The user steps down the ranked list until rank *k* and reformulates [Deterministic; no early abandonment]

Model-based measures

Probabilistic space of users following different paths

- Ω is the space of all paths
- P(ω) is the prob of a user following a path ω in Ω
- U(ω) is the utility of path ω in Ω

$$esM = \sum_{\omega \in \Omega} P(\omega)U(\omega)$$

[Yang and Lad, ICTIR 2009]

Expected Global Utility

Modeling Expected Utility of Multi-session Information Distillation

Yiming Yang and Abhimanyu Lad

1. User steps down ranked results one-by-one
2. Stops browsing documents based on a stochastic process that defines a stopping probability distribution over ranks and reformulates
3. Gains something from relevant documents, accumulating utility

Expected Global Utility

- The probability of a user following a path ω:

$$P(\omega) = P(r_1, r_2, \ldots, r_K)$$

  r_i is the stopping and reformulation point in list i

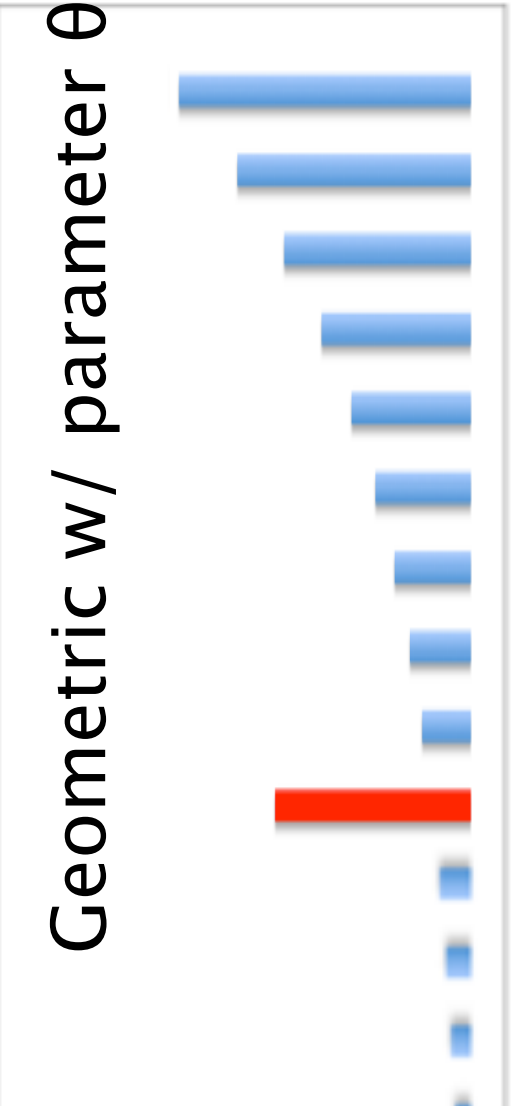
  - Assumption: stopping positions in each list are independent

$$P(r_1, r_2, \ldots, r_K) = P(r_1)P(r_2)\ldots P(r_K)$$

  - Use geometric distribution (RBP) to model the stopping and reformulation behavior

$$P(r_i = r) = (1 - \vartheta)\vartheta^{k-1}$$

Expected Global Utility



Q1	Q2	Q3
N	R	R
N	R	R
N	R	R
N	R	R
N	R	R
N	N	R
N	N	R
N	N	R
N	N	R
N	N	R

Existing measures

- Session DCG [Järvelin et al ECIR 2008]

  The user steps down the ranked list until rank *k* and reformulates [Deterministic; no early abandonment]

- Expected global utility [Yang and Lad ICTIR 2009]

  The user steps down a ranked list of documents until a decision point and reformulates [Stochastic; no early abandonment]

Model-based measures

Probabilistic space of users following different paths

- Ω is the space of all paths
- P(ω) is the prob of a user following a path ω in Ω
- M_ω is a measure over a path ω

$$\mathrm{es}M = \sum_{\omega \in \Omega} P(\omega) M_\omega$$

[Kanoulas et al. SIGIR 2011]

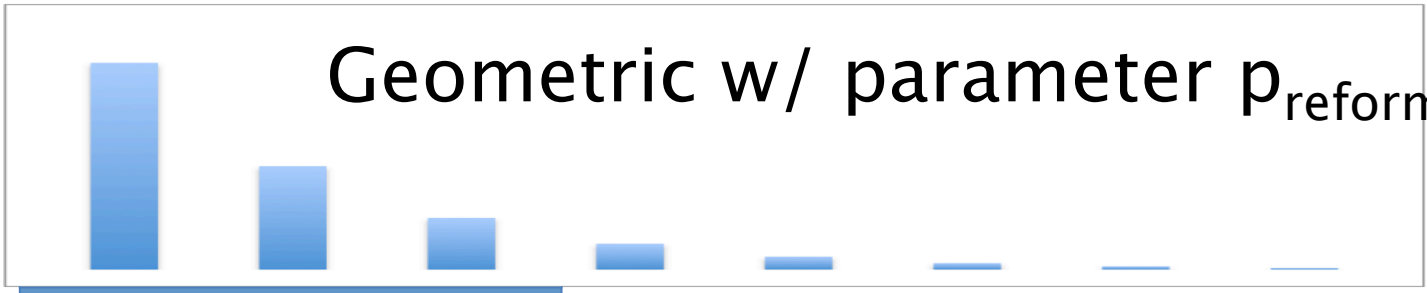
Probability of a path

Q1	Q2	Q3
N	R	R
N	R	R
N	R	R
N	R	R
N	R	R
N	N	R
N	N	R
N	N	R
N	N	R
N	N	R
…	…	…

Probability of abandoning at reform 2

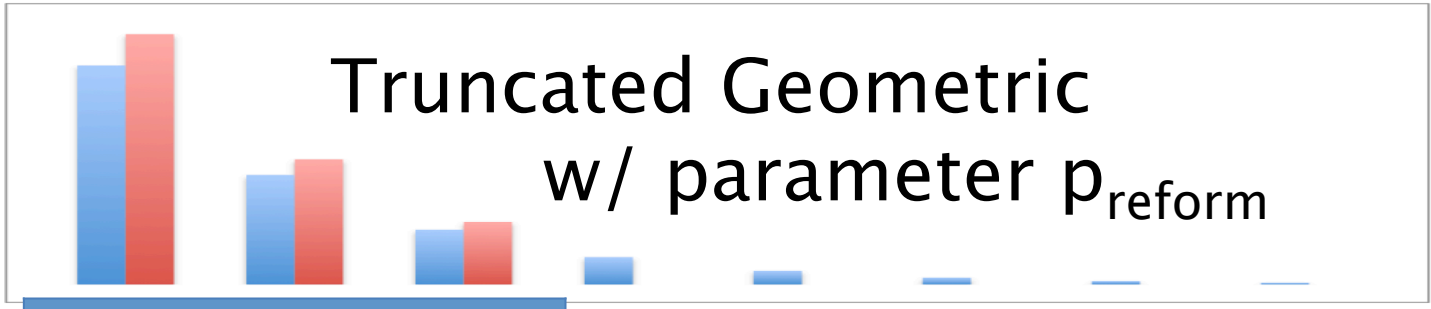
X

Probability of reformulating at rank 3



Q1	Q2	Q3
N	R	R
N	R	R
N	R	R
N	R	R
N	R	R
N	N	R
N	N	R
N	N	R
N	N	R
N	N	R
…	…	…

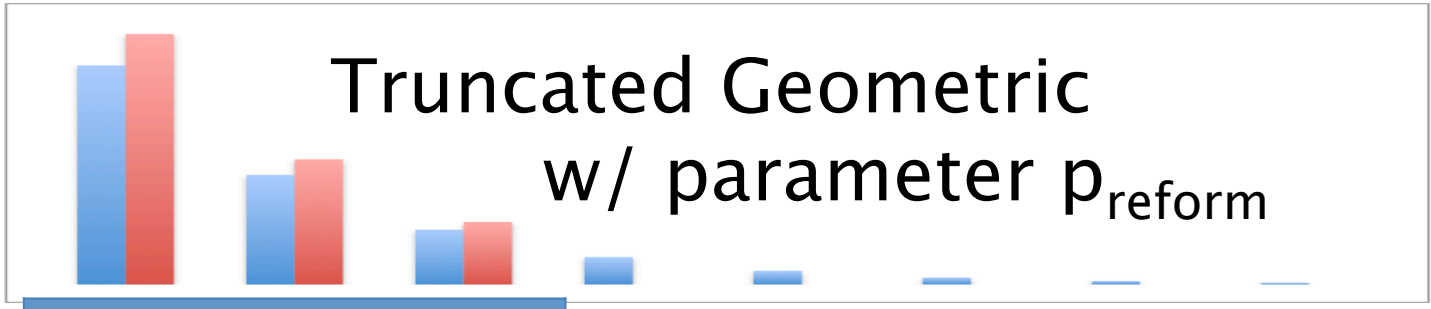
(1) Probability of abandoning the session at reformulation i

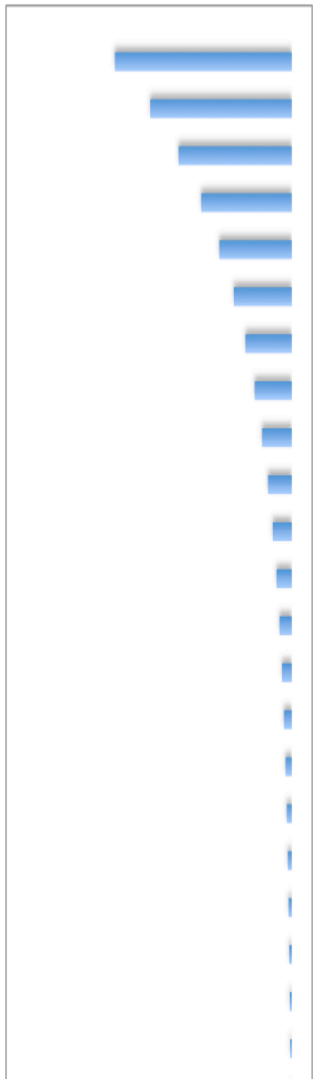


Truncated Geometric w/ parameter p_{reform}

Q1	Q2	Q3
N	R	R
N	R	R
N	R	R
N	R	R
N	R	R
N	N	R
N	N	R
N	N	R
N	N	R
N	N	R
…	…	…

(1) Probability of abandoning the session at reformulation i





Q1	Q2	Q3
N	R	R
N	R	R
N	R	R
N	R	R
N	R	R
N	N	R
N	N	R
N	N	R
N	N	R
N	N	R
…	…	…

(2) Probability of reformulating at rank j

Existing measures

- Session DCG [Järvelin et al ECIR 2008]
  The user steps down the ranked list until rank *k* and reformulates [Deterministic; no early abandonment]

- Expected global utility [Yang and Lad ICTIR 2009]
  The user steps down a ranked list of documents until a decision point and reformulates [Stochastic; no early abandonment]

- Expected session measures [Kanoulas et al. SIGIR 2011]
  The user steps down a ranked list of documents until a decision point and either abandons the query or reformulates [Stochastic; allows early abandonment]

Part 1: Evaluation Measures

- Evaluation measures
  1. Traditional evaluation measures
  2. Model-based measures
  3. Novelty and Diversity-based measures
  4. Session-based measures

- Evaluation of evaluation measures

Meta-evaluation

- Evaluate how good evaluation measures are
  1. Evaluate the models based on query logs
  2. Compare with side-by-side
  3. Compare their discriminative power
  4. Compare their informativeness

1. Query-log evaluation of models

A Comparative Study of Click Models for Web Search

Artem Grotov, Aleksandr Chuklin*, Ilya Markov
Luka Stout, Finde Xumara, and Maarten de Rijke

- Log-likelihood
  - how well a model approximates observed data
- Perplexity
  - how surprised a model is to see a click at rank r in a session s
- Correlation with click-metrics

1. Query-log evaluation of models

Expected Reciprocal Rank for Graded Relevance

Olivier Chapelle
Yahoo! Labs
Santa Clara, CA
chap@yahoo-inc.com

Donald Metzler
Yahoo! Labs
Santa Clara, CA
metzler@yahoo-inc.com

Ya Zhang
Yahoo! Labs
Sunnyvale, CA
yazhang@yahoo-inc.com

Pierre Grinspan
Google Inc
San Bruno, CA
pgrinspan@gmail.com

Table 2: Correlation between click and editorial metrics. Top row: market 1; bottom row: market 2. Left column: depth 10; right column: depth 5.

	DCG	NDCG	AP	RR	ERR
UCTR	0.028	-0.009	0.016	0.156	**0.214**
max RR	0.125	0.046	0.109	0.304	**0.393**
mean RR	0.147	0.063	0.129	0.333	**0.428**
min RR	0.162	0.076	0.142	0.352	**0.451**
PLC	0.138	0.057	0.121	0.321	**0.413**
SS	0.236	0.156	0.165	0.329	**0.384**

	DCG	NDCG	AP	RR	ERR
UCTR	0.131	0.108	0.225	0.238	**0.307**
max RR	0.197	0.199	0.348	0.366	**0.449**
mean RR	0.220	0.217	0.376	0.395	**0.480**
min RR	0.235	0.230	0.394	0.414	**0.500**
PLC	0.210	0.209	0.364	0.382	**0.466**
SS	0.357	0.361	0.402	0.415	**0.482**

	DCG	NDCG	AP	RR	ERR
UCTR	0.374	0.268	0.435	0.592	**0.664**
max RR	0.389	0.348	0.473	0.626	**0.700**
mean RR	0.410	0.347	0.479	0.642	**0.710**
min RR	0.423	0.346	0.479	0.649	**0.712**
PLC	0.399	0.347	0.480	0.635	**0.708**
SS	0.461	0.309	0.249	0.467	**0.501**

	DCG	NDCG	AP	RR	ERR
UCTR	0.298	0.215	0.335	0.356	**0.435**
max RR	0.319	0.253	0.369	0.407	**0.493**
mean RR	0.347	0.268	0.395	0.434	**0.515**
min RR	0.362	0.275	0.408	0.447	**0.524**
PLC	0.334	0.261	0.386	0.424	**0.507**
SS	0.427	0.378	0.373	0.402	**0.465**

2. Side-by-side

- Side-by-side
  - Predict choice

Do User Preferences and Evaluation Measures Line Up?

Mark Sanderson, Monica Lestari Paramita, Paul Clough, Evangelos Kanoulas
Department of Information Studies, University of Sheffield
Regent Court, 211 Portobello St, Sheffield, S1 4DP, UK
+44 114 22 22648
m.sanderson@shef.ac.uk

COOGEE BEACH SYDNEY AUSTRALIA.
Coogee Beach is a great place to eat, surf and swim. Visitors and locals enjoy the great hospitality and exciting night-life of the **Coogee Bay** Hotel, ...
How to get to Coogee Beach - Photos - Event Calendar
members.optusnet.com.au/mons1/ - Cached - Similar

Coogee, New South Wales - Wikipedia, the free encyclopedia
Jump to and **Coogee Bay** along with **Coogee Beach** lie towards the eastern ...: **Coogee** was gazetted as a village ... to become the **Coogee Bay** ...
History - Transport - Geography and landmarks
en.wikipedia.org/wiki/**Coogee**,_New_South_Wales - Cached - Similar

Coogee Beach - Sydney
Coogee Beach in the eastern suburbs of Sydney: **coogee beach** sydney sydney **coogee** bondi **beach** larry cossies.
goaustralia.about.com/cs/sydneysightscity/l/blpiccg1.htm - Cached - Similar

Coogee Beach - Sydney - Reviews of **Coogee Beach** - TripAdvisor
Coogee Beach, Sydney: See 7 reviews, articles, and 2 photos of **Coogee Beach**, ranked No.10 on TripAdvisor among 397 attractions in Sydney.
www.tripadvisor.com/Attraction_Review-g255060-d257492-Reviews-**Coogee_Beach**-Sydney_New_South_Wales.html - Cached - Similar

COOGEE BEACH SYDNEY AUSTRALIA.
Coogee Beach is a great place to eat, surf and swim. Visitors and locals enjoy the great hospitality and exciting night-life of the **Coogee Bay** Hotel, ...
How to get to Coogee Beach - Photos - Event Calendar
members.optusnet.com.au/mons1/ - Cached - Similar

Coogee, New South Wales - Wikipedia, the free encyclopedia
Jump to and **Coogee Bay** along with **Coogee Beach** lie towards the eastern ...: **Coogee** was gazetted as a village ... to become the **Coogee Bay** ...
History - Transport - Geography and landmarks
en.wikipedia.org/wiki/**Coogee**,_New_South_Wales - Cached - Similar

Coogee Beach - Sydney
Coogee Beach in the eastern suburbs of Sydney: **coogee beach** sydney sydney **coogee** bondi **beach** larry cossies.
goaustralia.about.com/cs/sydneysightscity/l/blpiccg1.htm - Cached - Similar

Coogee Beach - Sydney - Reviews of **Coogee Beach** - TripAdvisor
Coogee Beach, Sydney: See 7 reviews, articles, and 2 photos of **Coogee Beach**, ranked No.10 on TripAdvisor among 397 attractions in Sydney.
www.tripadvisor.com/Attraction_Review-g255060-d257492-Reviews-**Coogee_Beach**-Sydney_New_South_Wales.html - Cached - Similar

2. Side-by-side

Users	nDCG		P(10)		ERR	
Agree	160	65%	131	62%	164	66%
Rnk eql	21	9%	18	9%	21	9%
Disgree	66	27%	61	29%	62	25%
	247		**210**		**247**	

Users	α-nDCG		IA-PC	
Agree	50	64%	28	60%
Rnk eq	4	5%	1	2%
Disgree	24	31%	18	38%
	78		**47**	

3. Discriminative power

Evaluating Evaluation Metrics based on the Bootstrap

Tetsuya Sakai
Knowledge Media Laboratory, Toshiba Corporate R&D Center
tetsuya.sakai@toshiba.co.jp

Evaluating Diversified Search Results Using Per-intent Graded Relevance

Tetsuya Sakai
Microsoft Research Asia, Beijing, P.R.C.
tetsuyasakai@acm.org

Ruihua Song
Microsoft Research Asia, Beijing, P.R.C.
Song.Ruihua@microsoft.com

4. Informativeness

- A ranked list of relevance judgments defines a search engine performance

- How much does a measure reduce one's uncertainty in the underlying list of relevance?
  - Informative measures: large reduction in uncertainty
  - Non-informative measures: little or no reduction in uncertainty

The Maximum Entropy Method for Analyzing Retrieval Measures

Javed A. Aslam*, Emine Yilmaz, Virgiliu Pavlu
College of Computer and Information Science
Northeastern University
360 Huntington Ave, #202 WVH
Boston, MA 02115
{jaa,emine,vip}@ccs.neu.edu

Informativeness

A system retrieves *N* documents. What is the probability of relevance of each document in the list?

Informativeness

A system retrieves *N* documents. What is the probability of relevance of each document in the list?

No additional information. Most natural answer *1/2*.

Informativeness

A system retrieves *N* documents.
R of them are relevant.
What is the probability of
relevance of each document
in the list?

Informativeness

A system retrieves N documents.
R of them are relevant.
What is the probability of relevance of each document in the list?

Most natural answer R/N.

Informativeness

A system retrieves *N* documents. *R* of them are relevant. The average precision of the system is *ap*. What is the probability of relevance of each document in the list?

Informativeness

A system retrieves *N* documents. *R* of them are relevant. The average precision of the system is *ap*. What is the probability of relevance of each document in the list?

If *ap* is high, then top documents are more likely to be relevant. But how can we find the most reasonable distribution?

The Maximum Entropy Method

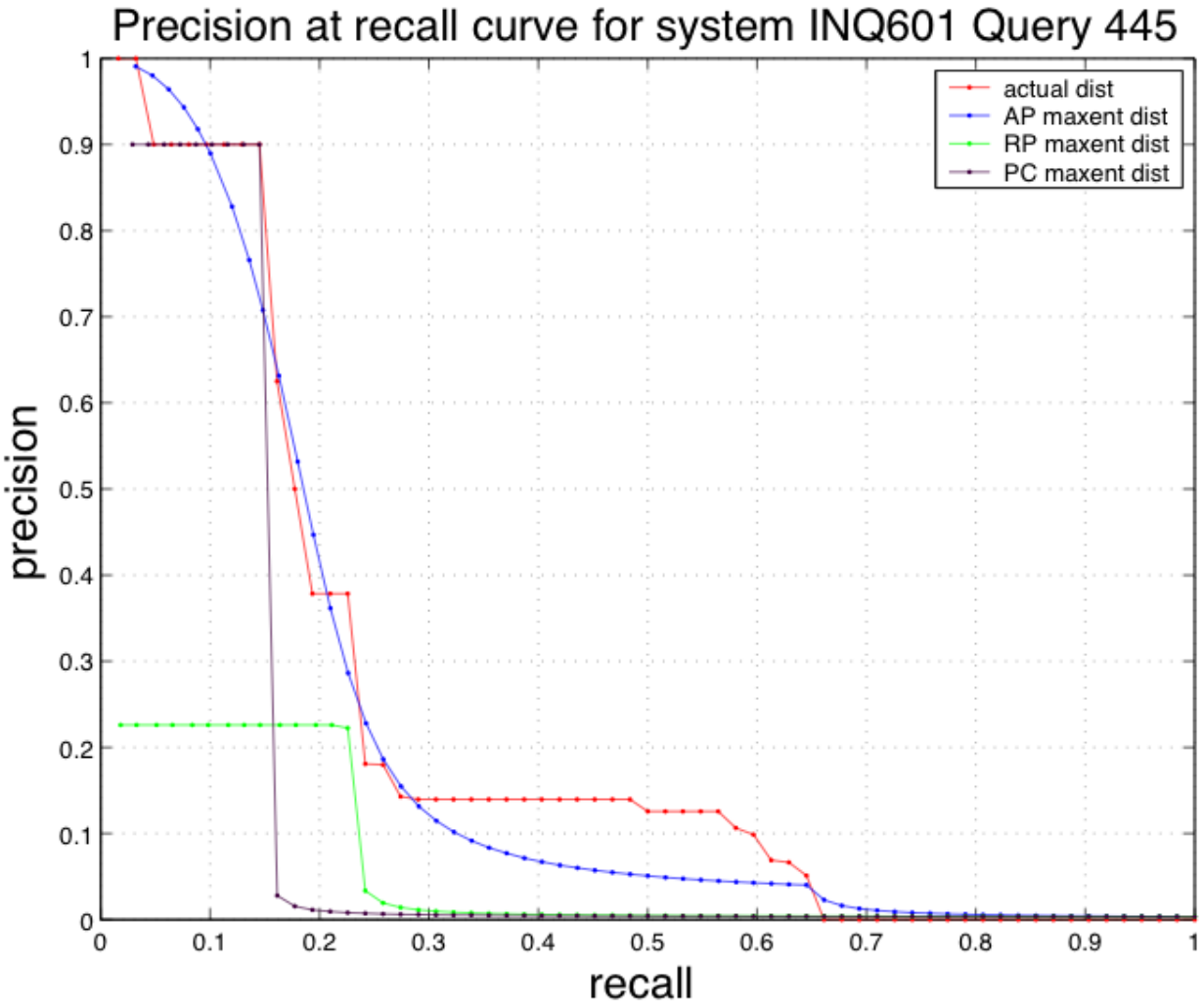
- The most "reasonable" distribution satisfying the given constraints
- Principle of Maximal Ignorance
  - Pick distribution which is least predictable (most random) subject to constraints
- How to measure randomness? Entropy
- Thus, max entropy distribution subject to constraints

Setup for AP Constraint

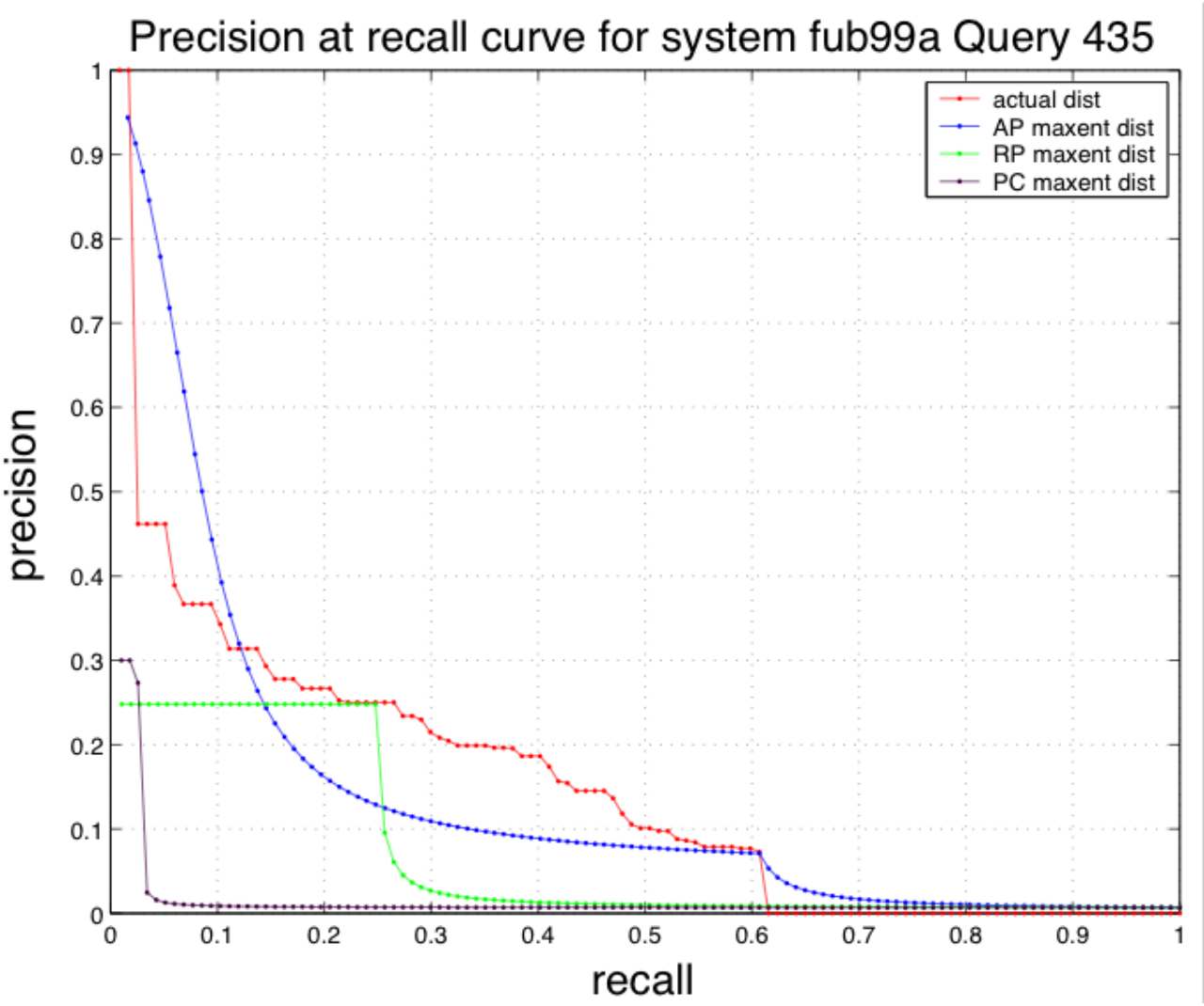
- Goal:
  - Given the average precision value (ap) of a list, infer probability of relevance of document at rank i

- Maximum entropy setup:
  - Maximize
    - $\sum_{i=1}^{N} H(p_i)$
  - Subject to
    - $E[AP] = \frac{1}{R} \sum_{i=1}^{N} \left(\frac{p_i}{i} \left(1 + \sum_{j=1}^{i-1} \right) \right) = ap$
    - $E[R] = \sum_{i=1}^{N} p_i = R$

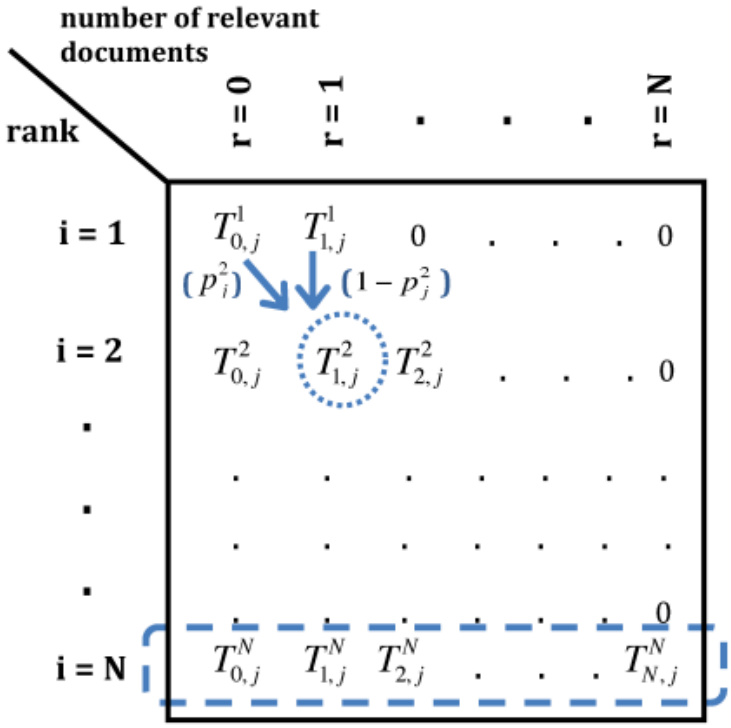
Actual and Inferred P-R Curves



Actual and Inferred P-R Curves



Setup for Cascade and Intent-Aware Measures



Example :

Given p^2_j is the probability of relevance value to the intent j for the document at rank i=2, the gain value from the previous rank would increase by $\dot{G}^2_j$ with a probability of p^2_j, or it would remain unchanged with a probability of $1-p^2_j$. These two alternatives affect the expected value of the measure at rank i=2 as follows:

$$T^2_{1,j} = p^2_j(T^1_{0,j} + \dot{G}^2_j) + (1-p^2_j)T^1_{1,j}$$

Distribution of the expected value of the measure for the top **N** Documents

On the Informativeness of Cascade and Intent-Aware Effectiveness Measures

Azin Ashkan and Charles L.A. Clarke
Cheriton School of Computer Science
University of Waterloo, Canada
{aashkan, claclark}@cs.uwaterloo.ca

Setup for Cascade and Intent-Aware Measures

Maximize $\sum_{i=1}^{N} H(P^i)$

Subject to:

1. $\sum_{i=1}^{N} p^i = R_{ret}$
2. $T_r^i = p^i[T_{r-1}^{i-1} + \frac{(1-\alpha)^{r-1}}{i}] + (1-p^i)T_r^{i-1} = ERR_{ret}$

   $T_r^i = 0 \quad \forall i < r$

(a) ERR

Maximize $\sum_{i=1}^{N} H(P^i)$

Subject to:

1. $\sum_{i=1}^{N} p^i = R_{ret}$
2. $T_r^i = p^i[T_{r-1}^{i-1} + (1-\alpha)^{r-1}\beta^{i-1}] + (1-p^i)T_r^{i-1} = RBP_{ret}$

   $T_r^i = 0 \quad \forall i < r$

(b) RBP

Maximize $\sum_{i=1}^{N}\sum_{j=1}^{M} H(p_j^i)$

Subject to:

1. $\forall\ 1 \le j \le M, \quad \sum_{i=1}^{N} p_j^i = R_{ret}^j$
2. $T_{r,j}^i = p_j^i[T_{r-1,j}^{i-1} + \frac{(1-\alpha)^{r-1}}{i}] + (1-p_j^i)T_{r,j}^{i-1}$

   $T_r^i = 0 \quad \forall i < r$

   $\frac{1}{M}\sum_{j=1}^{M} T_{R_{ret}^j,j}^N = ERR_IA_{ret}$

(a) ERR_IA

Maximize $\sum_{i=1}^{N}\sum_{j=1}^{M} H(p_j^i)$

Subject to:

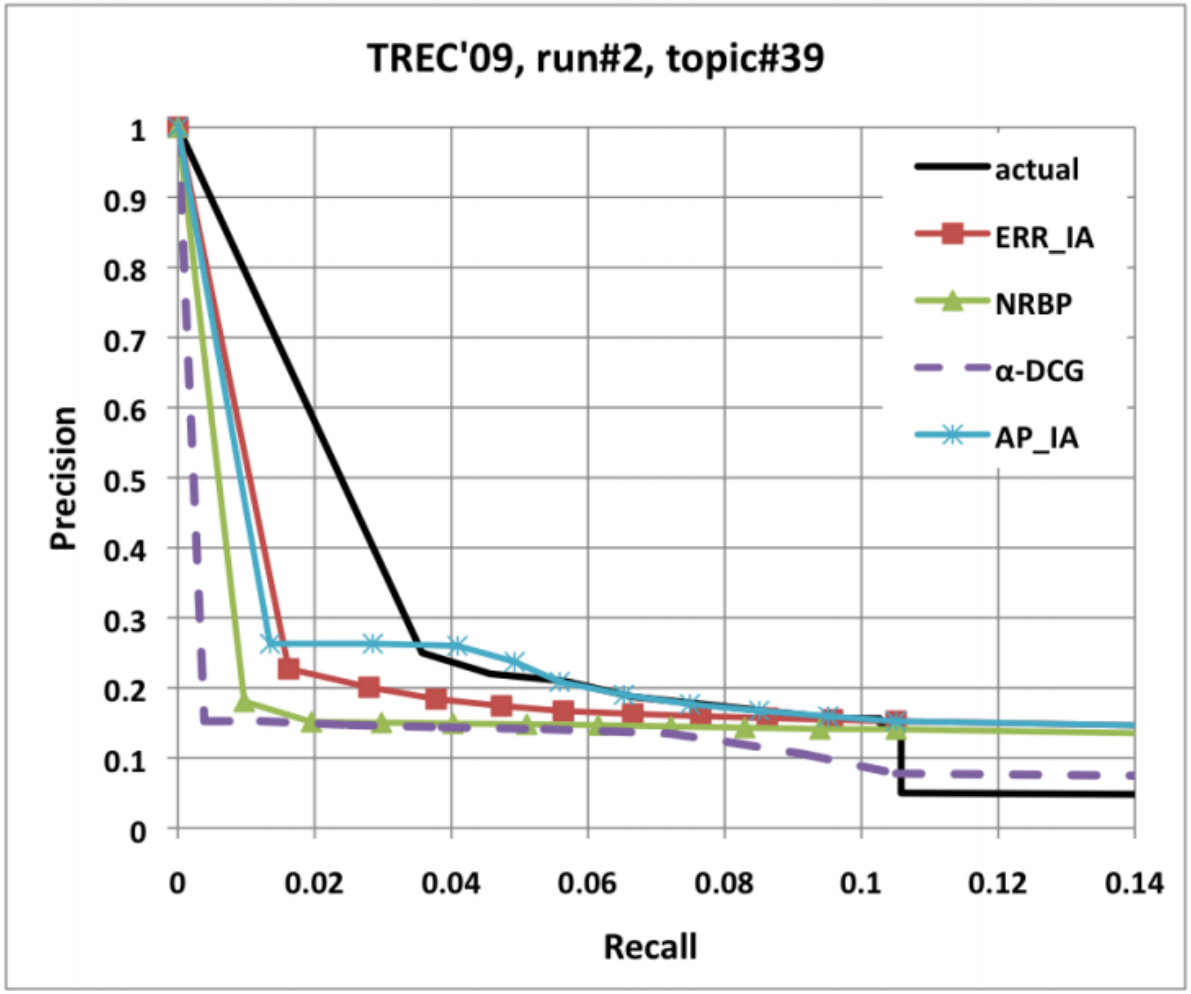
1. $\forall\ 1 \le j \le M, \quad \sum_{i=1}^{N} p_j^i = R_{ret}^j$
2. $T_{r,j}^i = p_j^i[T_{r-1,j}^{i-1} + \frac{(1-\alpha)^{r-1}}{\log_2 i}] + (1-p_j^i)T_{r,j}^{i-1}$

   $T_{r,j}^i = 0 \quad \forall i < r$

   $\frac{1}{M}\sum_{j=1}^{M} T_{R_{ret}^j,j}^N = \alpha DCG_{ret}$

(c) α-DCG

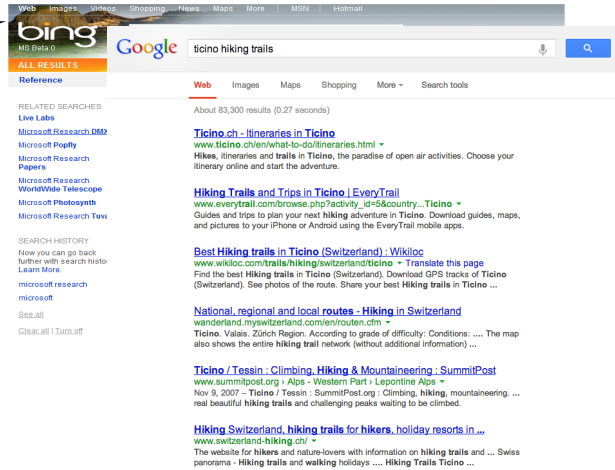
Actual and Inferred P-R Curves



Part 2. Relevance Judgments

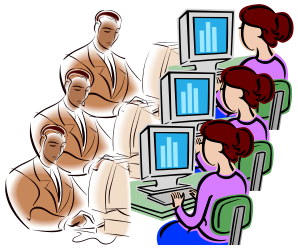
- uses of alternative dispute resolution
- job search vancouver washington
- poem of arrival of columbus

Results



Search Algorithms

Judges



QUERY: job search vancouver washington

DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

How many documents to judge?

- Many measures are based on
  - Recall : "*out of all good documents in the collection, how many did the algorithm find?*"
  - all good documents in the collection need to be identified

Choosing Query-Document Pairs

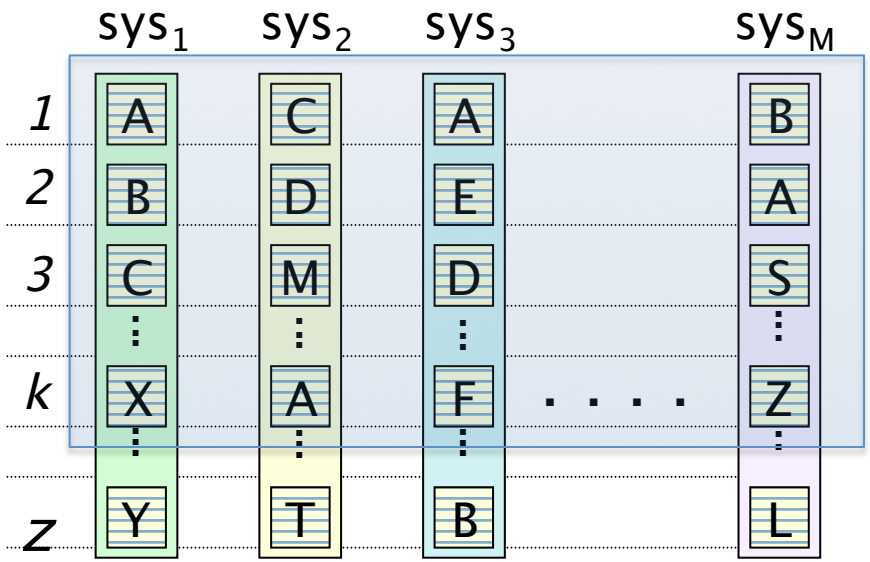
- Deterministic

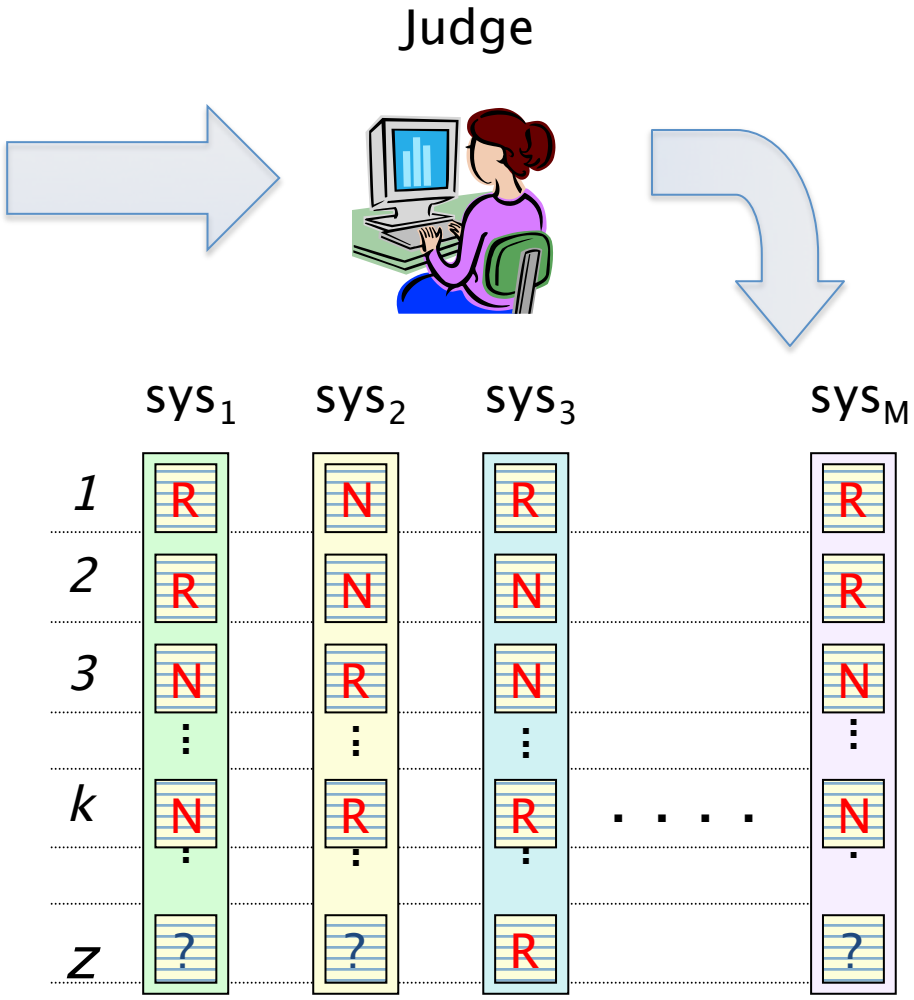
  1. Depth-k pooling (standard method)
  2. Evaluating without judgments (automatic eval)
  3. Finding relevance documents as quickly as possible
  4. Nugget-based evaluation

- Stochastic

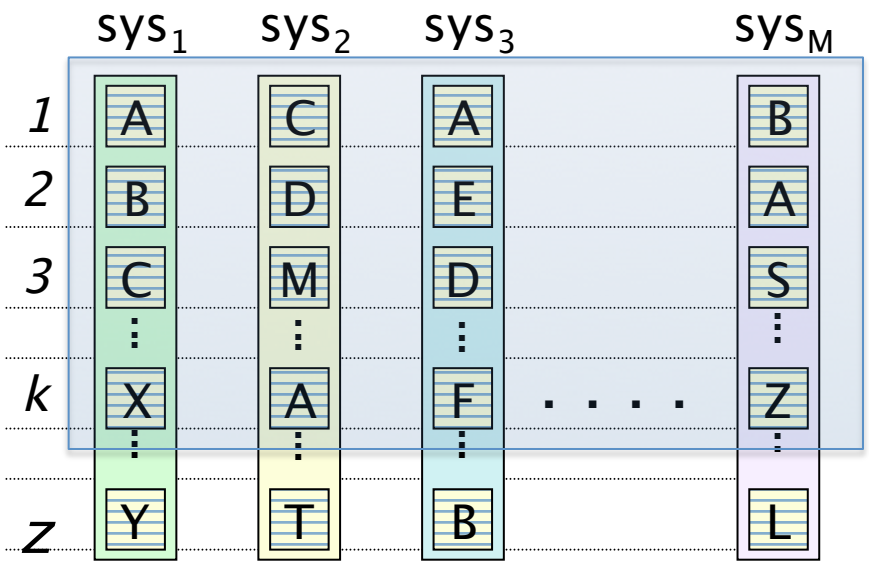
  5. Estimating measures
  6. Inferring relevance judgments

Depth-k pooling
(TREC Standard Setup)

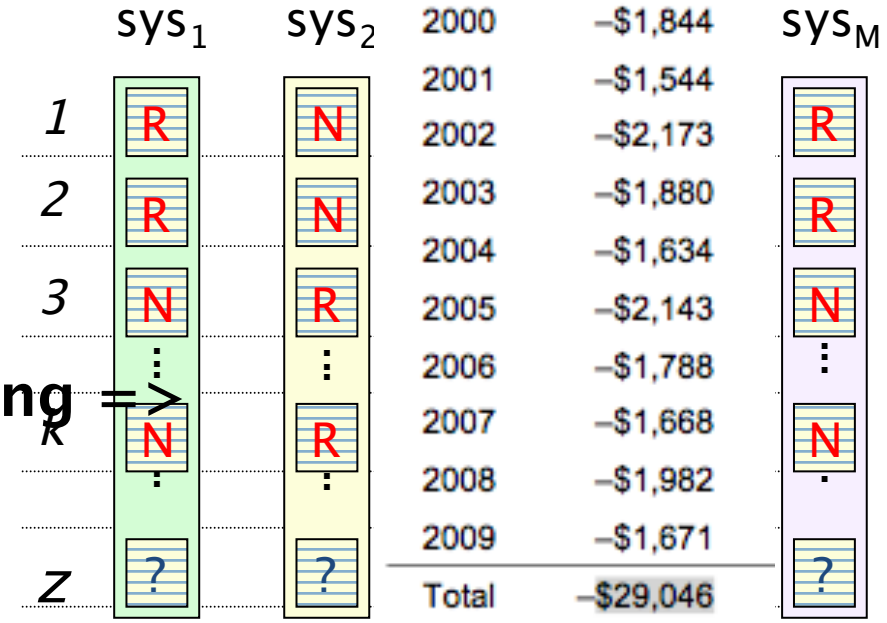




Depth-k pooling (TREC Standard Setup)



Year	Total TREC Investment Costs (thousands $2009)
1991	–$753
1992	–$713
1993	–$674
1994	–$1,522
1995	–$1,282
1996	–$2,129
1997	–$61
1998	–$1,739
1999	–$1,848
2000	–$1,844
2001	–$1,544
2002	–$2,173
2003	–$1,880
2004	–$1,634
2005	–$2,143
2006	–$1,788
2007	–$1,668
2008	–$1,982
2009	–$1,671
Total	–$29,046



TREC 8 test collection

- **50 topics, depth-100 pooling => 86,830 judgments**
- **30 sec per judgment => 724 hours => 18 weeks**

Choosing Query-Document Pairs

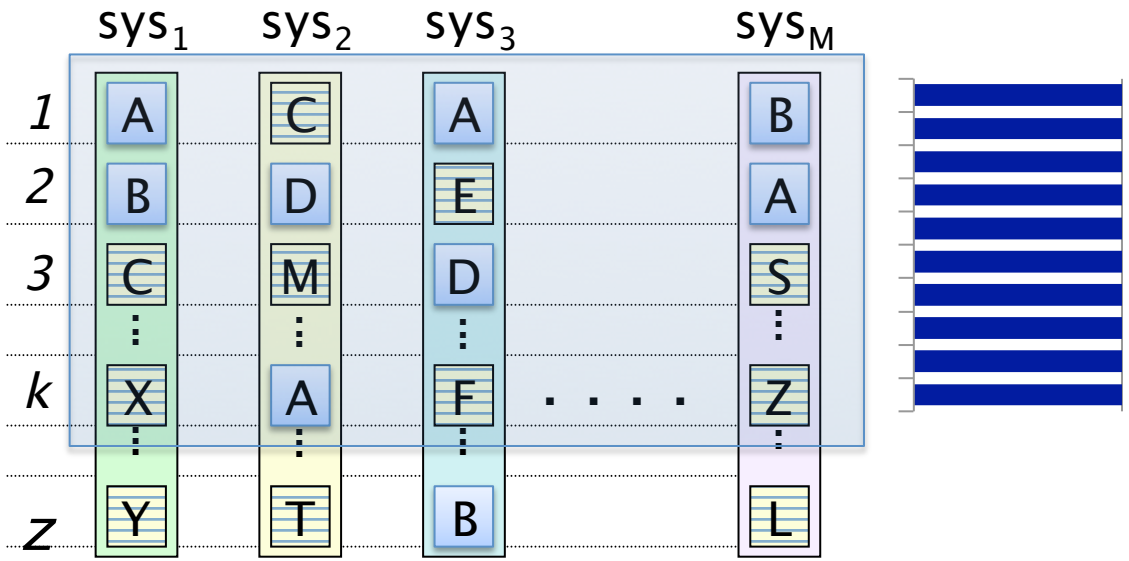
- Deterministic

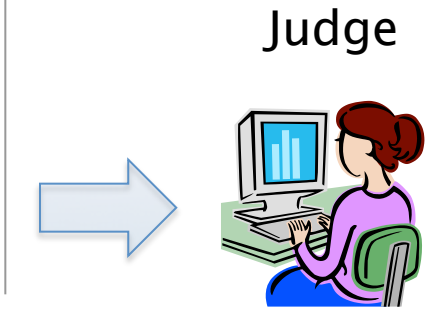
  1. Depth-k pooling (standard method)
  2. Evaluating without judgments (automatic eval)
  3. Finding relevance documents as quickly as possible
  4. Nugget-based evaluation

- Stochastic

  5. Sampling and Estimating measures
  6. Sampling and Inferring relevance judgments

Uniform random sampling

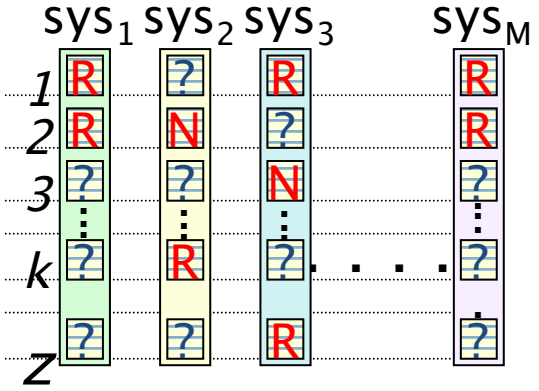




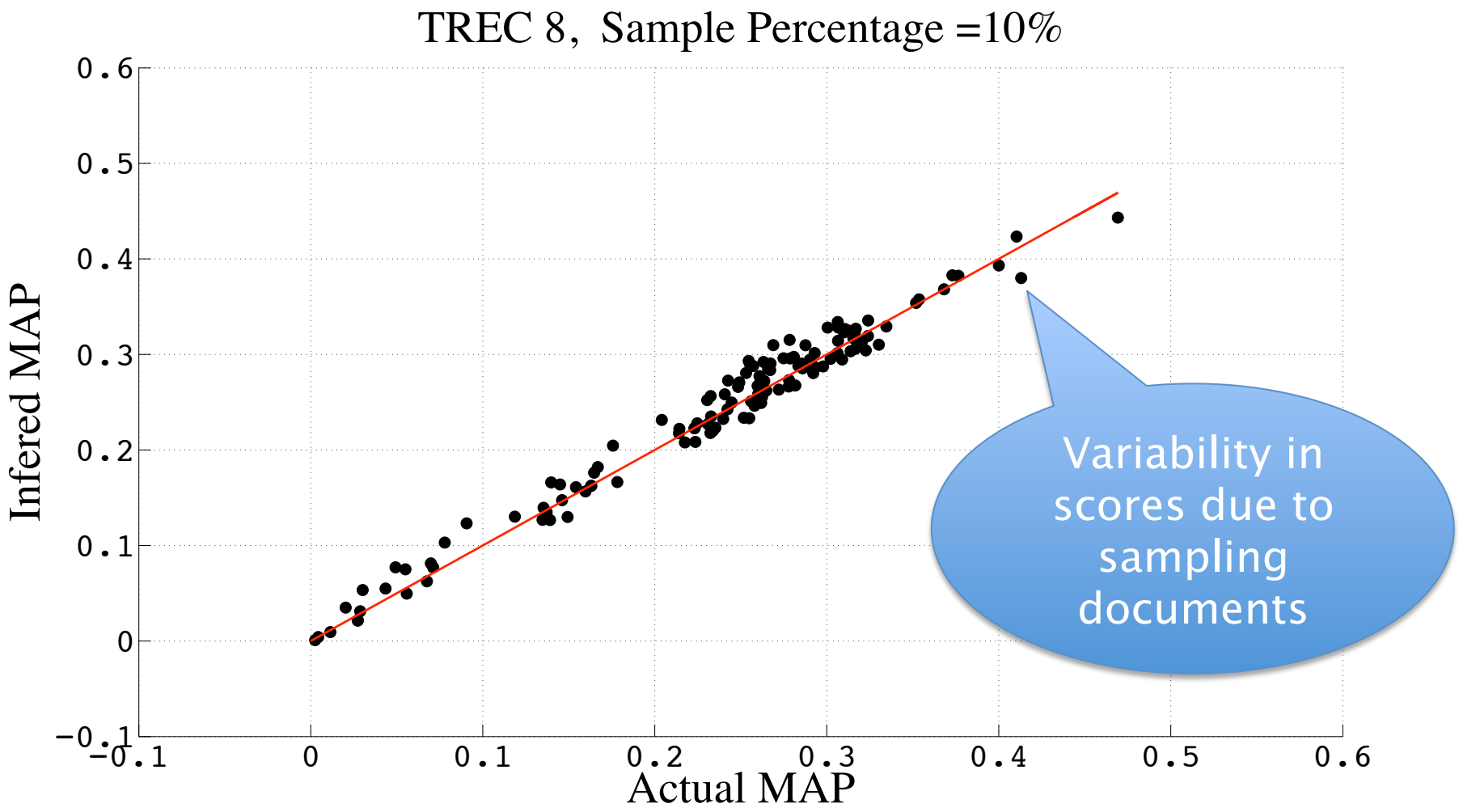
Inferred AP : Estimating Average Precision with Incomplete Judgments

Emine Yilmaz, Javed A. Aslam*
College of Computer and Information Science
Northeastern University
360 Huntington Ave, #202 WVH
Boston, MA 02115

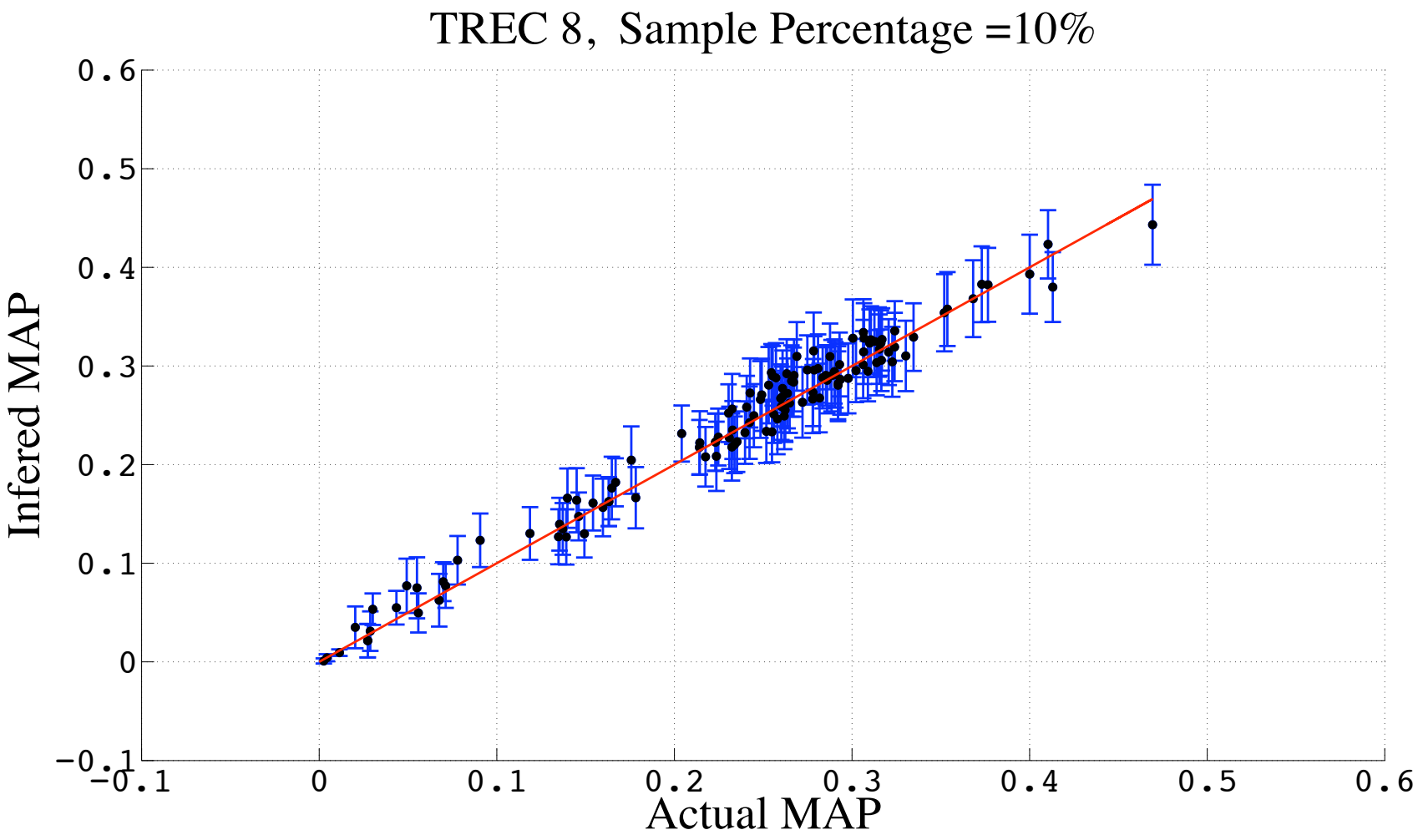
{emine, jaa}@ccs.neu.edu



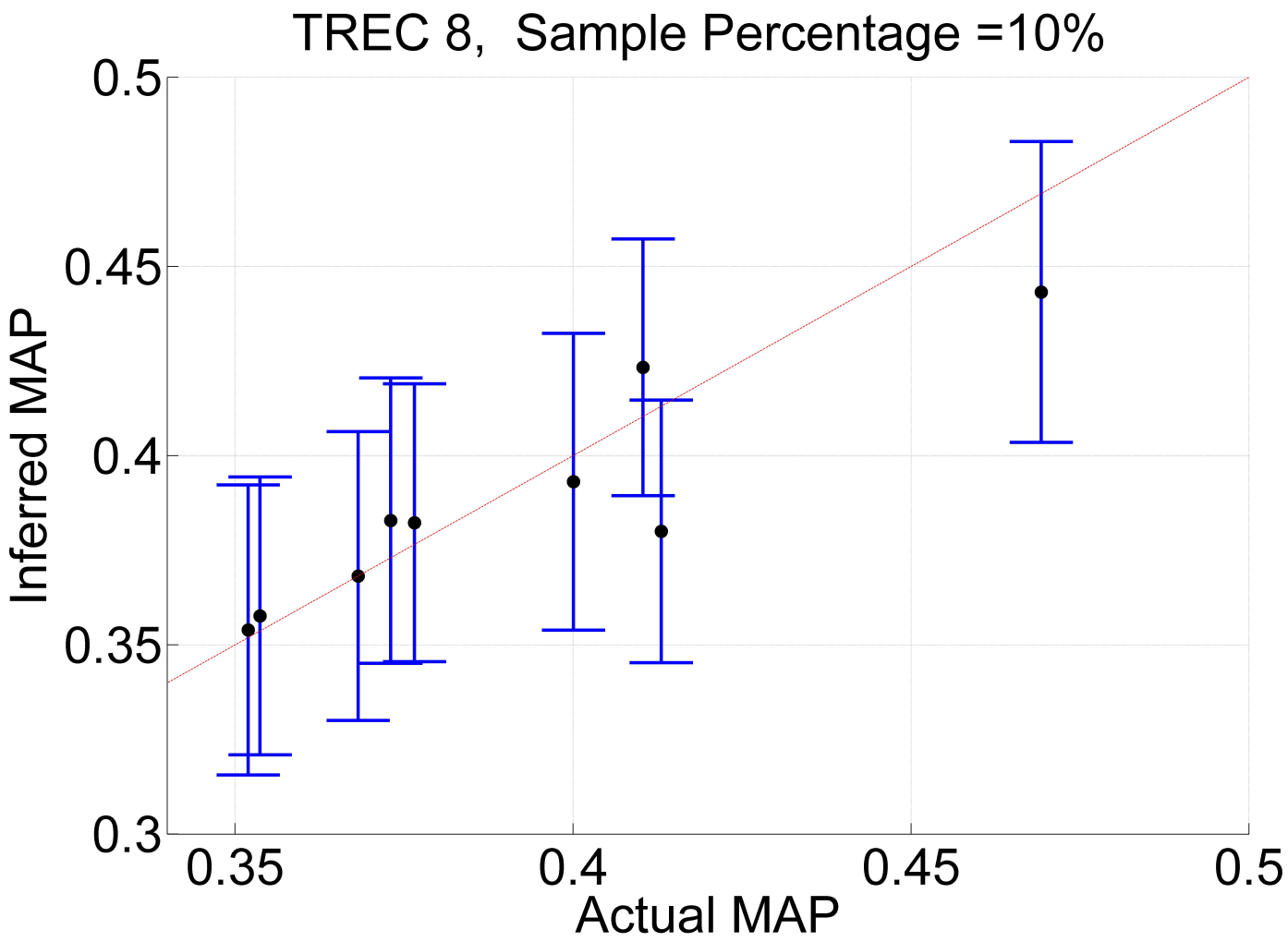
Inferred Measures



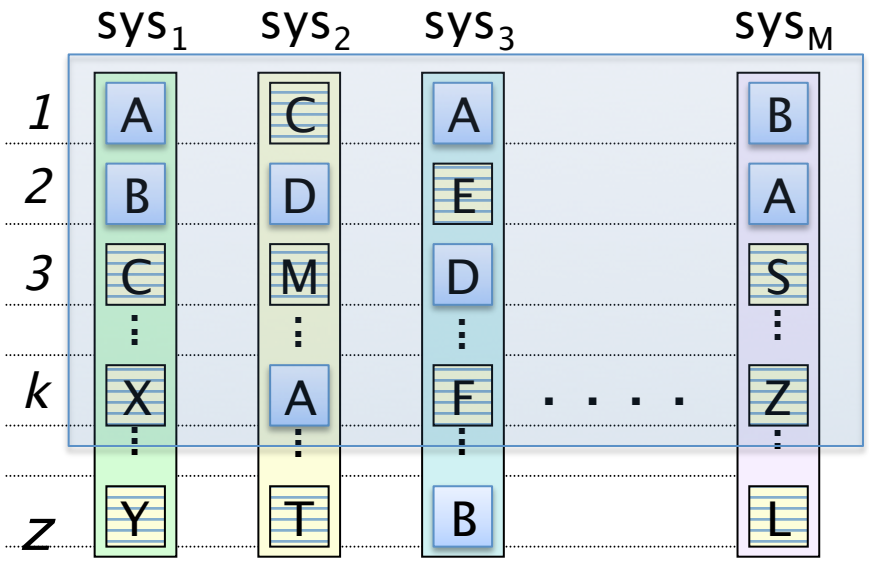
Confidence Intervals for Inferred Measures

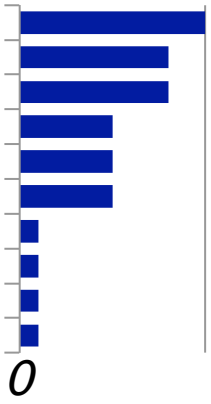


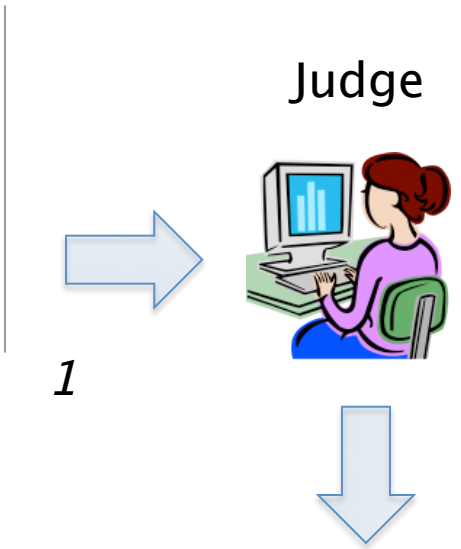
Confidence Intervals for Inferred Measures



Stratified Random Sampling







A Simple and Efficient Sampling Method for Estimating AP and NDCG

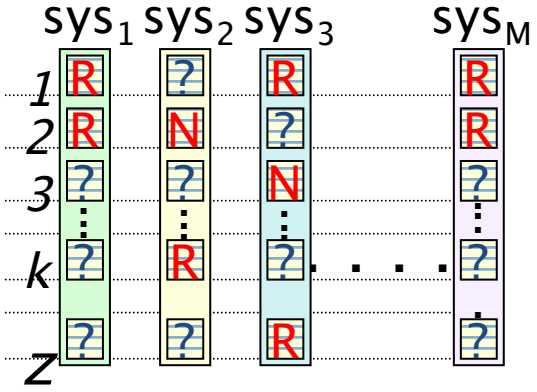
Emine Yilmaz* eminey@microsoft.com

Evangelos Kanoulas† ekanou@ccs.neu.edu

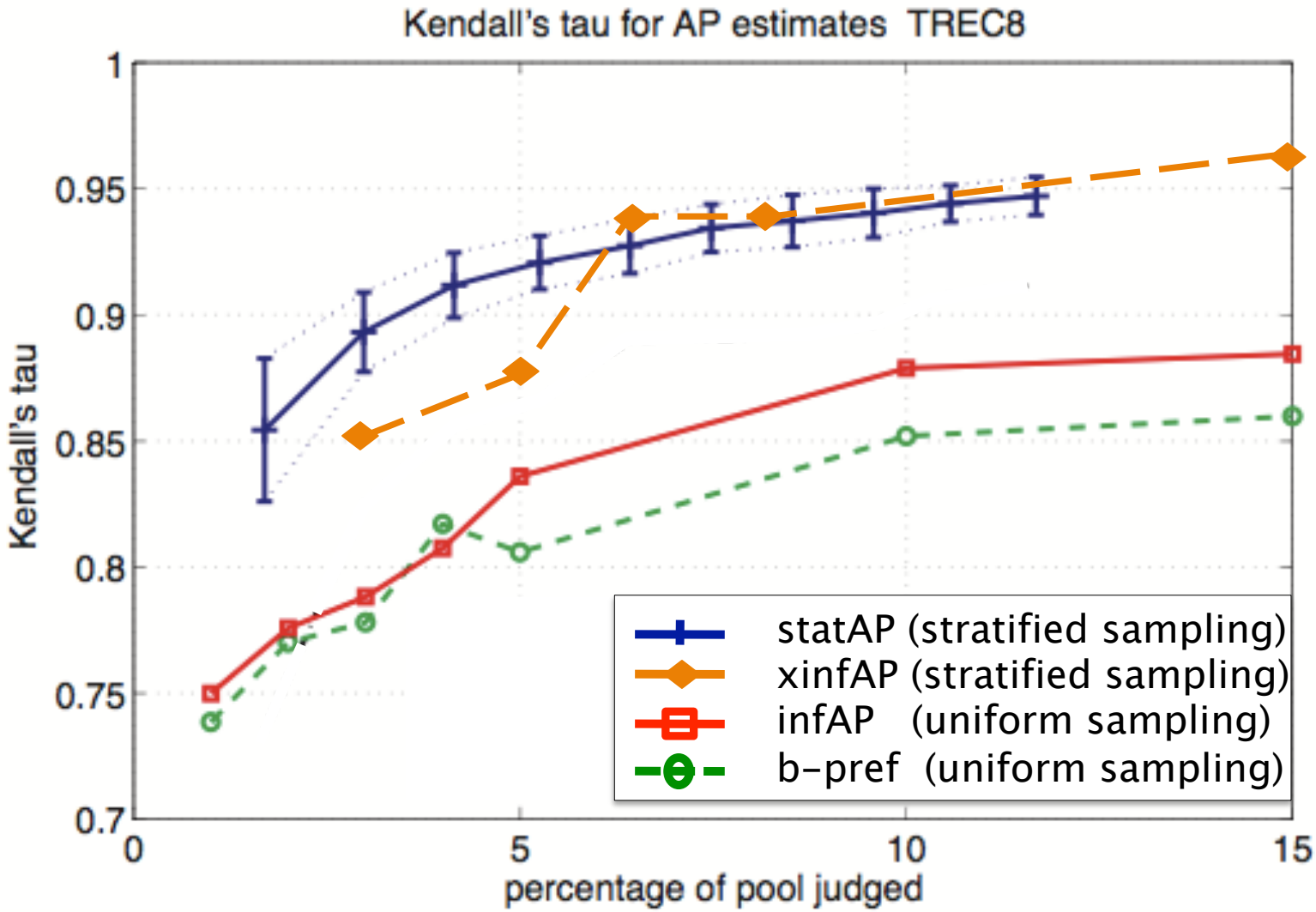
Javed A. Aslam†* jaa@ccs.neu.edu

*Microsoft Research
7 JJ Thomson Avenue
Cambridge CB3 0FB, UK

†College of Computer and Information Science
Northeastern University
360 Huntington Ave, #202 WVH
Boston, MA 02115



Comparison of the measures: Kendall's Tau



Choosing Query-Document Pairs

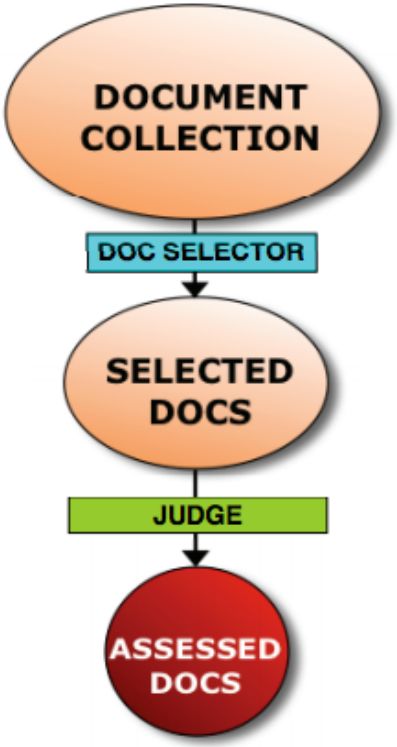
- Deterministic

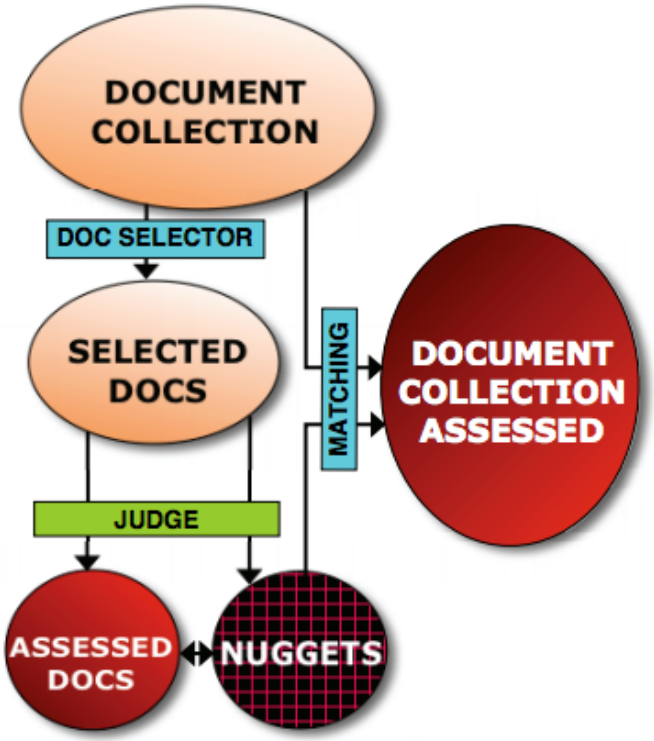
  1. Depth-k pooling (standard method)
  2. Evaluating without judgments (automatic eval)
  3. Finding relevance documents as quickly as possible
  4. Nugget-based evaluation

- Stochastic

  5. Sampling and Estimating measures
  6. Sampling and Inferring relevance judgments

Nugget-based Evaluation





IR System Evaluation using Nugget-based Test Collections

Virgil Pavlu Shahzad Rajput Peter B. Golbus Javed A. Aslam
College of Computer and Information Science, Northeastern University
{vip, rajput, pgolbus, jaa}@ccs.neu.edu*

Nugget-based Evaluation

Algorithm:

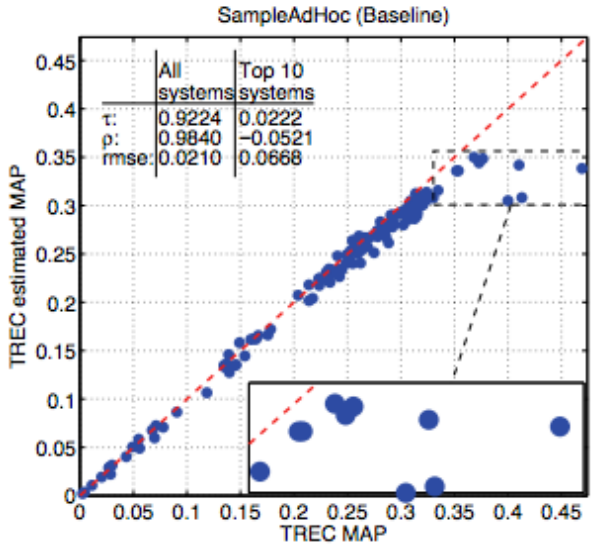
1. Sample documents (stratified sampling)
2. Extract nuggets
   - Smallest portion of text that constitutes relevant information
3. Infer relevance
   - Matching algorithm between nuggets and documents
   - Threshold on relevance

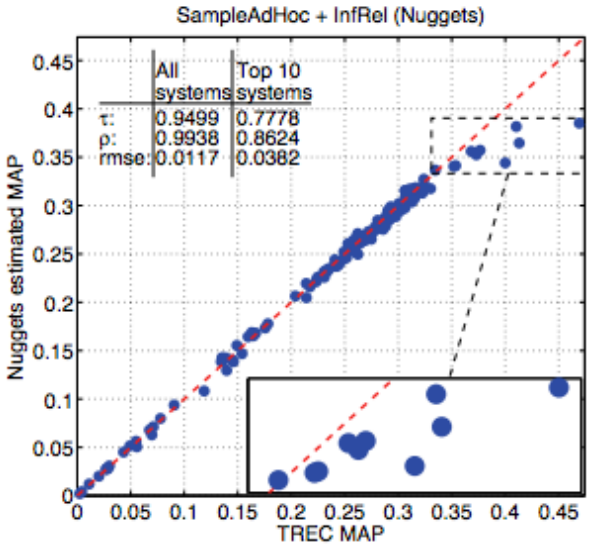
TREC Temporal Summarization

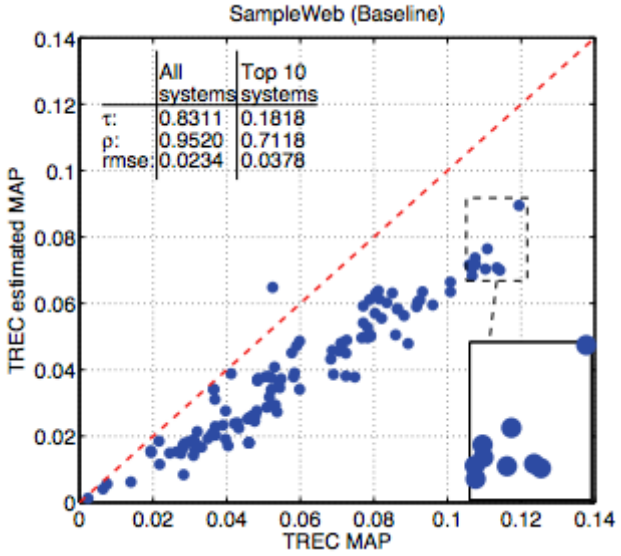
One Click Access

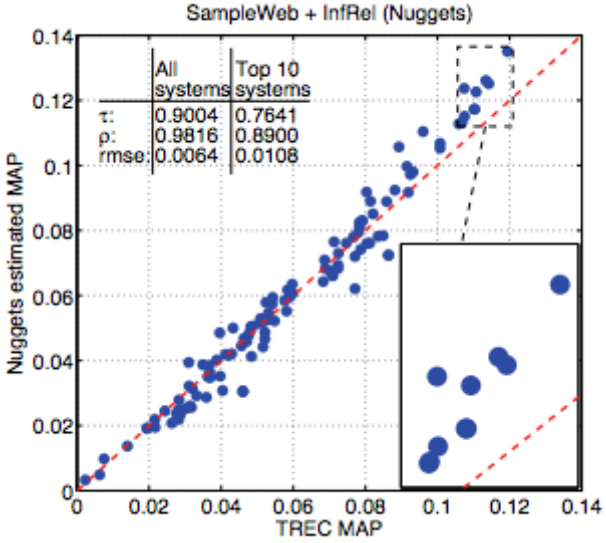
One Click Access evaluation at NTCIR

Nugget-based Evaluation

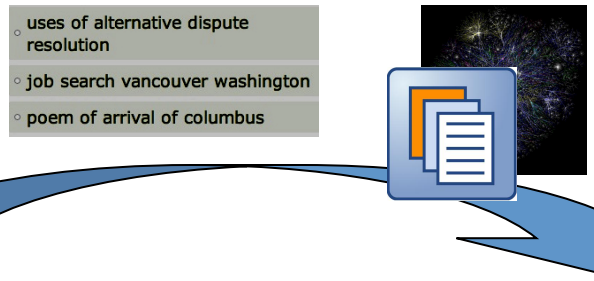




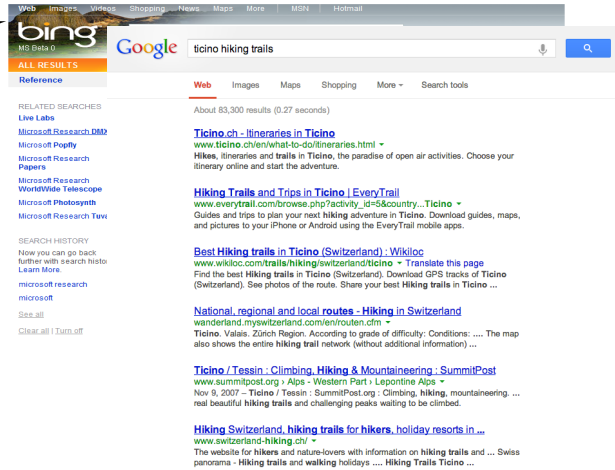




Test Collection Construction

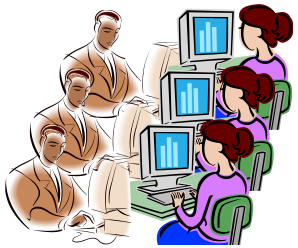


Results



Search Algorithms

Judges



QUERY: job search vancouver washington

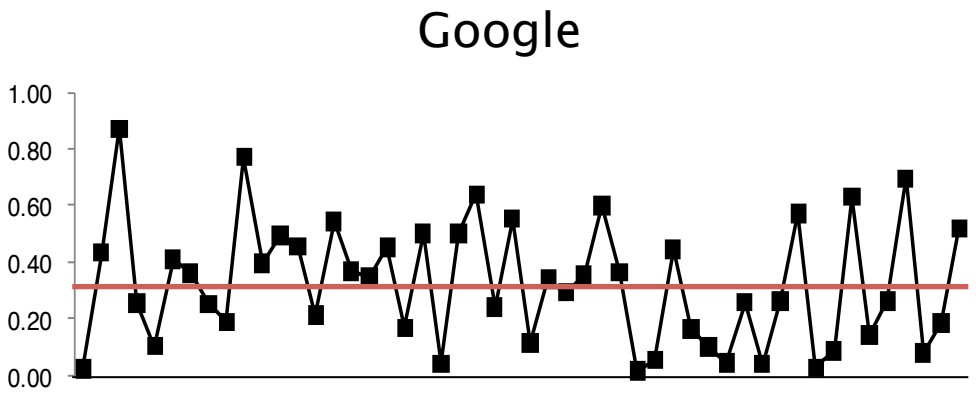
DESCRIPTION: I would like to find web page that aggregate job opportunities in the IT industry in Vancouver, Washington.

2. Comparative Evaluation

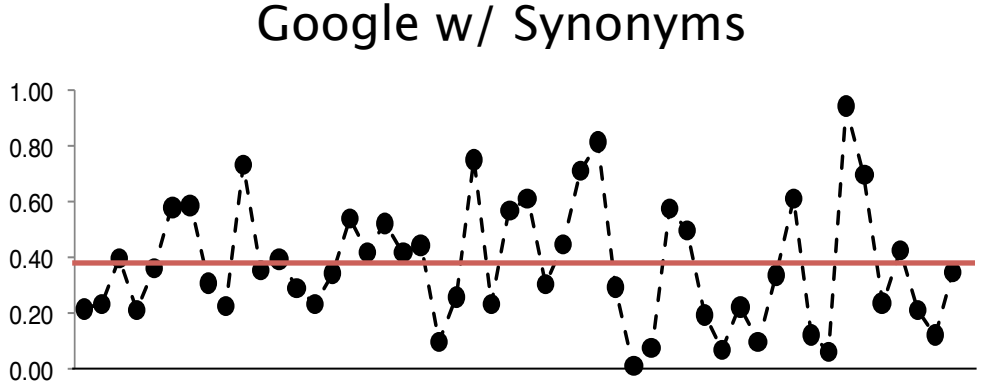
Comparing Retrieval Systems

- Hypothesis: Expanding a short keyword query with synonyms will improve search engine effectiveness

- Google:
  - Mean NDCG = 0.324



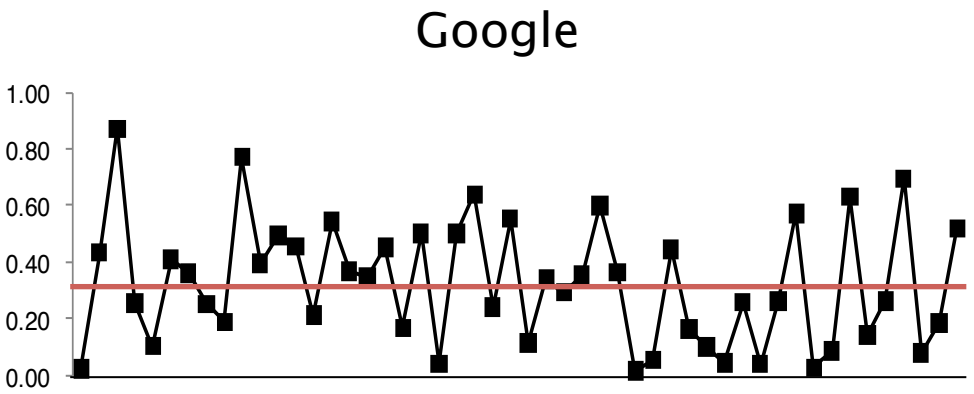
- Google w/ Synonyms
  - Mean NDCG = 0.373



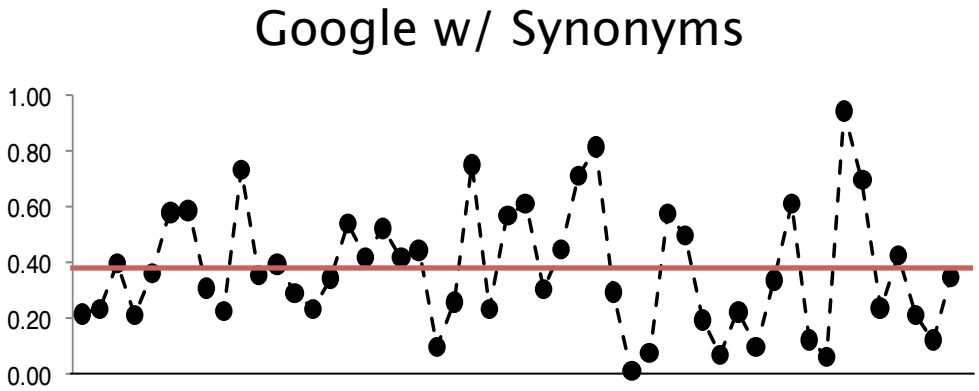
So what?

- Do these results support my hypothesis?

- Is it possible that my results are just random?



statistical significance testing



Hypotheses and Experiments

- Hypothesis:
  - Using an SVM for classification will give better accuracy than using Naïve Bayes
  - Expanding a short keyword query with synonyms will improve search engine effectiveness

- Experiment:
  - Build a baseline system
  - Modify it based on your hypothesis
  - Try both systems on one or more data sets

Statistical Significance Testing

- Two hypotheses, e.g.
  - H_0: B−A = 0
  - H_a: B−A ≠ 0 or B−A>0
- System performance measurements over a sample of queries
- A test statistic T computed from those measurements
- The distribution of the test statistic under H_0

Statistical Significance Testing

- We want to prove the null hypothesis wrong

- A p-value is
  - the probability of sampling the test statistic T …
  - … from a distribution obtained by assuming H_0 is true
    - If the p-value is low, conclude H_0 is false

What Are Tests Telling Us?

- Fisher: p-value is the likelihood of the data under H_0
  - The p-value is a conclusion about this particular experiment only
  - Nothing more, nothing less

- Neyman-Pearson: $p<0.05$ means we can reject H_0 as being unlikely to be true
  - p-values lead to inference about population
  - The p-value itself is not interesting; the inference is
  - Note that we do not accept that H_1 is true!

What Are Tests NOT Telling Us?

- NOT the probability of H_0 is true
- NOT the probability that the results are due to chance
- NOT whether the experiment is reliable

Commonly used tests

A Comparison of Statistical Significance Tests for Information Retrieval Evaluation

Mark D. Smucker, James Allan, and Ben Carterette
Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts Amherst
{smucker, allan, carteret}@cs.umass.edu

- Non-parametric
  - Sign test/binomial test
  - Wilcoxon signed rank test
- Parametric
  - Student's t-test
- Distribution-free
  - Randomization test
  - Bootstrap test

Student's t-test

- Assumptions
  - follow normal distribution
- Statistic : $t = \frac{\overline{B-A}}{\sigma_{B-A}} . \sqrt{N}$
- t-test performs well even when the normality assumption is violated [Hull SIGIR93]

Student's t-test

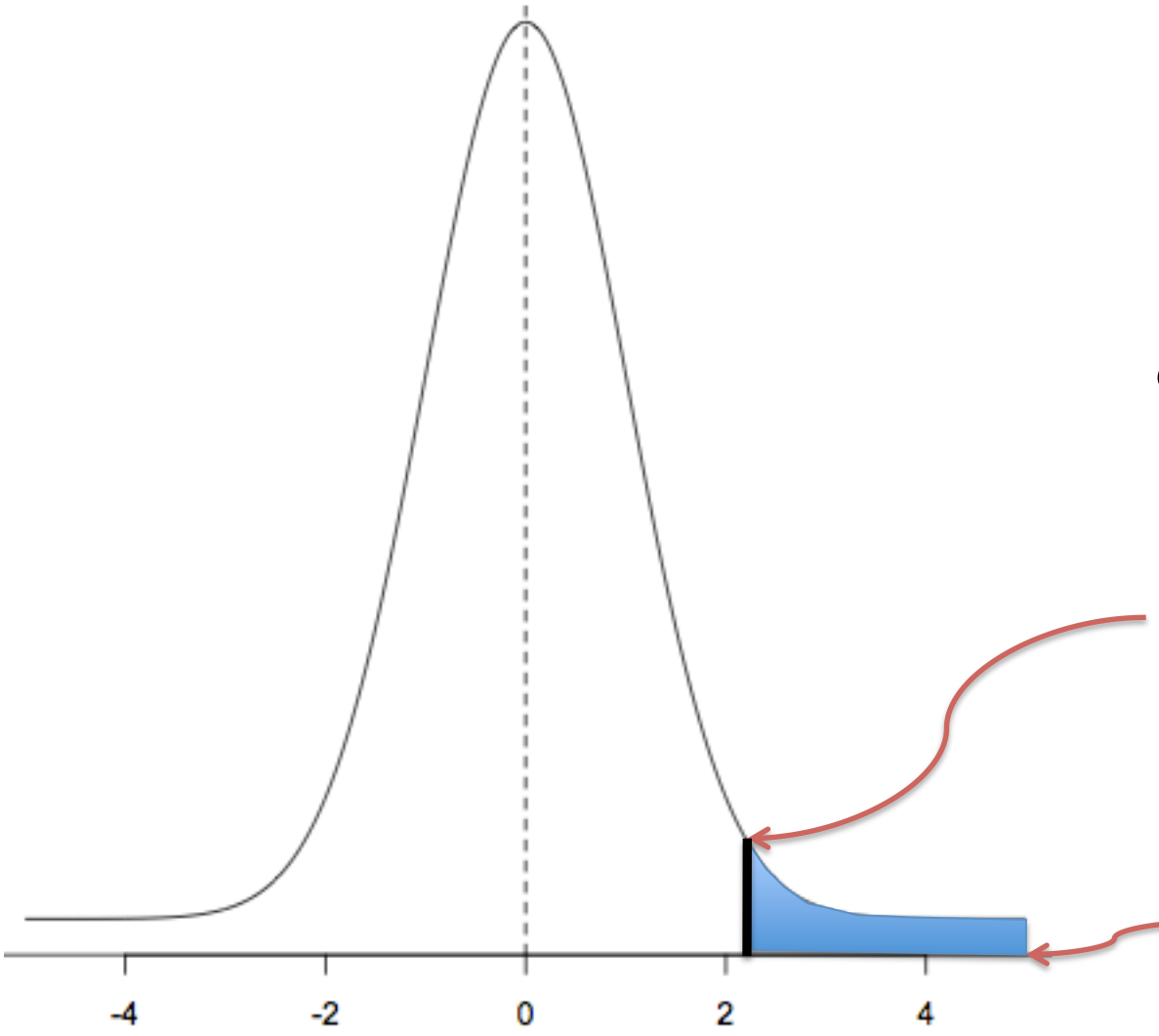
Query	A	B	B-A
1	.25	.35	+.10
2	.43	.84	+.41
3	.39	.15	-.24
4	.75	.75	0
5	.43	.68	+.25
6	.15	.85	+.70
7	.20	.80	+.60
8	.52	.50	-.02
9	.49	.58	+.09
10	.50	.75	+.25

$$\hat{\mu} = \overline{B - A} = 0.214$$

$$\hat{\sigma}_{B-A} = 0.291$$

$$t = \frac{\hat{\mu}}{\hat{\sigma}_{B-A}} \sqrt{n} = 2.33$$

Student's t-test



$$\hat{\mu} = \overline{B - A} = 0.214$$

$$\hat{\sigma}_{B-A} = 0.291$$

$$t = \frac{\hat{\mu}}{\hat{\sigma}_{B-A}}\sqrt{n} = 2.33$$

$$p - value = 0.02$$

Randomization test

- Loop for many times {
  - Load query scores for 2 systems
  - Randomly swap values per topic
  - Compute average for each syster
  - Compute difference between averages
  - Add difference to array

  }
- Sort array
- If actual difference outside 95% differences in array
  - Two systems are significantly different

Query	A		B
1	.25		.35
2	.43	⟷	.84
3	.39	⟷	.15
4	.75	⟷	.75
5	.43		.68
6	.15	⟷	.85
7	.20	⟷	.80
8	.52		.50
9	.49		.58
10	.50		.75

Bootstrap test

- Loop for many times {
  - Load topic scores for 2 systems
  - Randomly sample pair of scores w/ replacement
  - Compute average for each system
  - Compute difference between averages
  - Add difference to array

  }
- Sort array
- If actual difference outside 95% differences in array
  - Two systems are significantly different

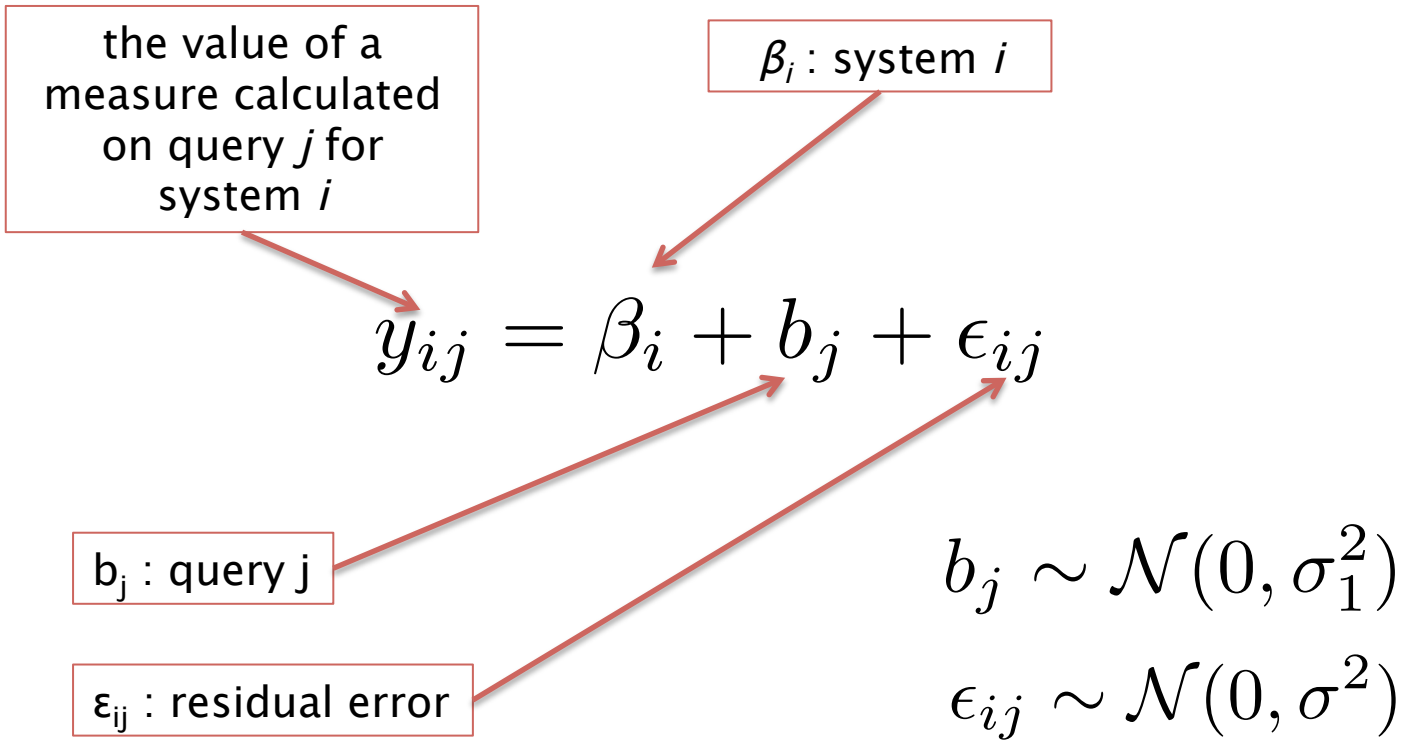
The Linear Model

- Statistical tests are classifier
  - Like classifiers, they are based on an underlying mode
  - Unlike classifiers, we cannot evaluate them directly

- The t-test is based on a linear regression model

Mixed Effects Models

- T-test



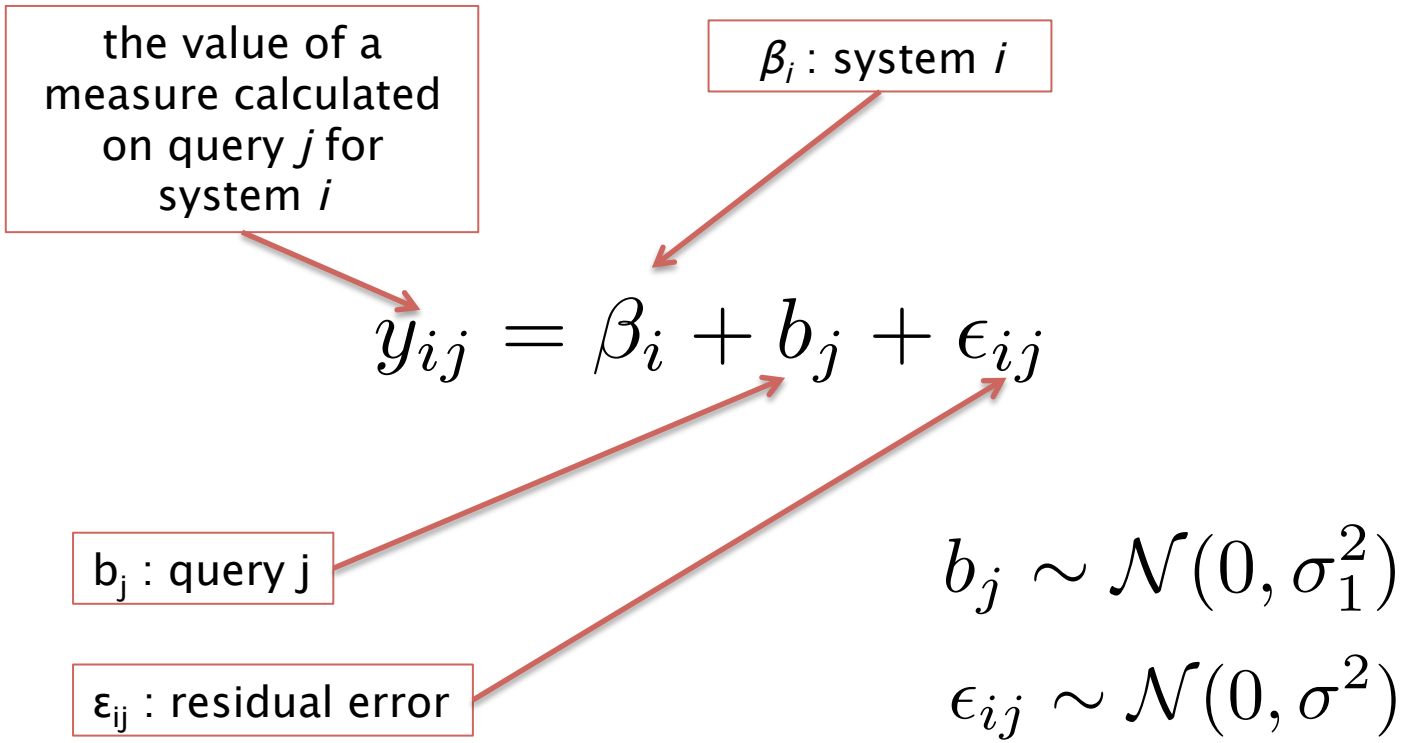
Mixed Effects Models

- In the statistical programming environment **R**

```
lme(effectiveness ~ system, data=data, random=~1|query)
```

Mixed Effects Models

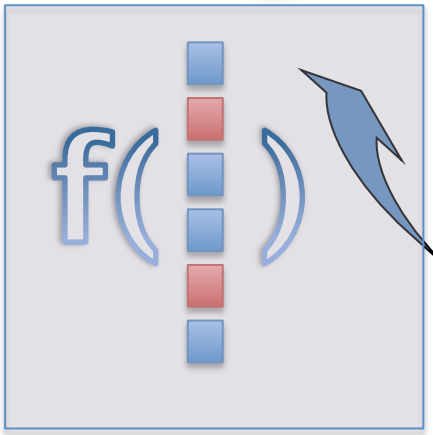
- T-test



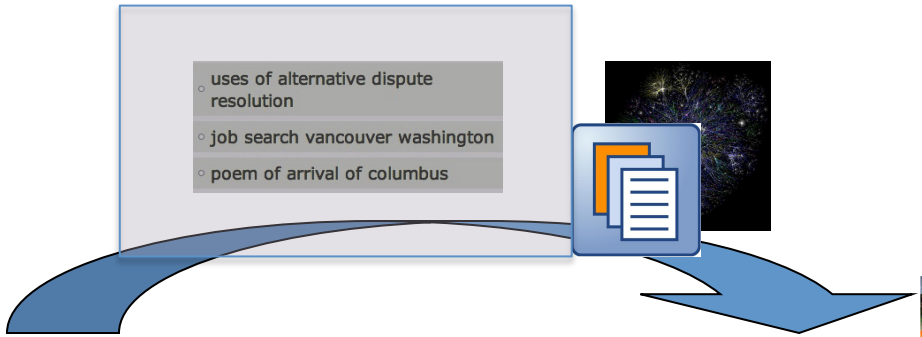
Sources of Varia

- uses of alternative dispute resolution
- job search vancouver washin
- poem of arrival of columbus

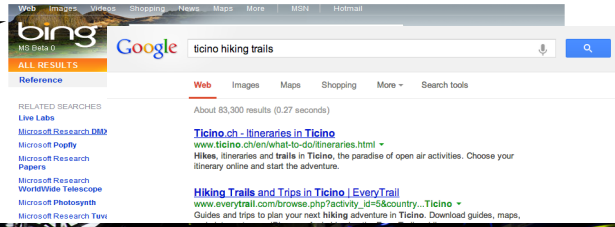
Search Algorithms



Sources of Variance in IR

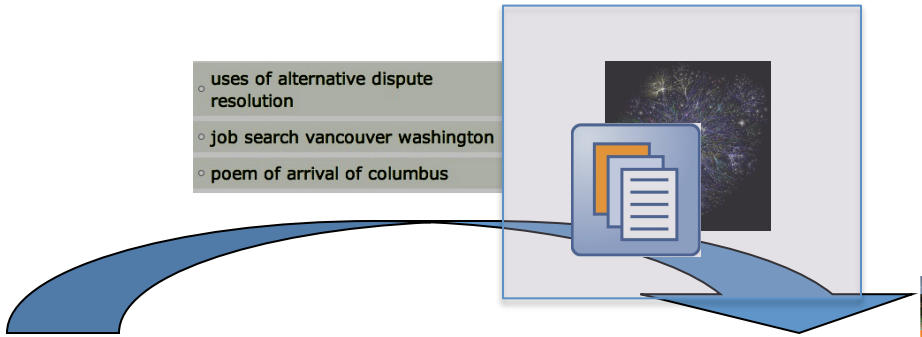


Results

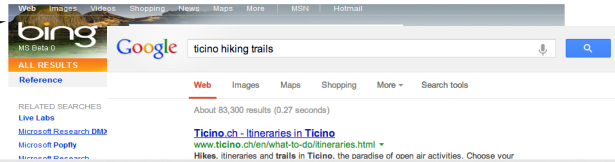


Search Algorithms

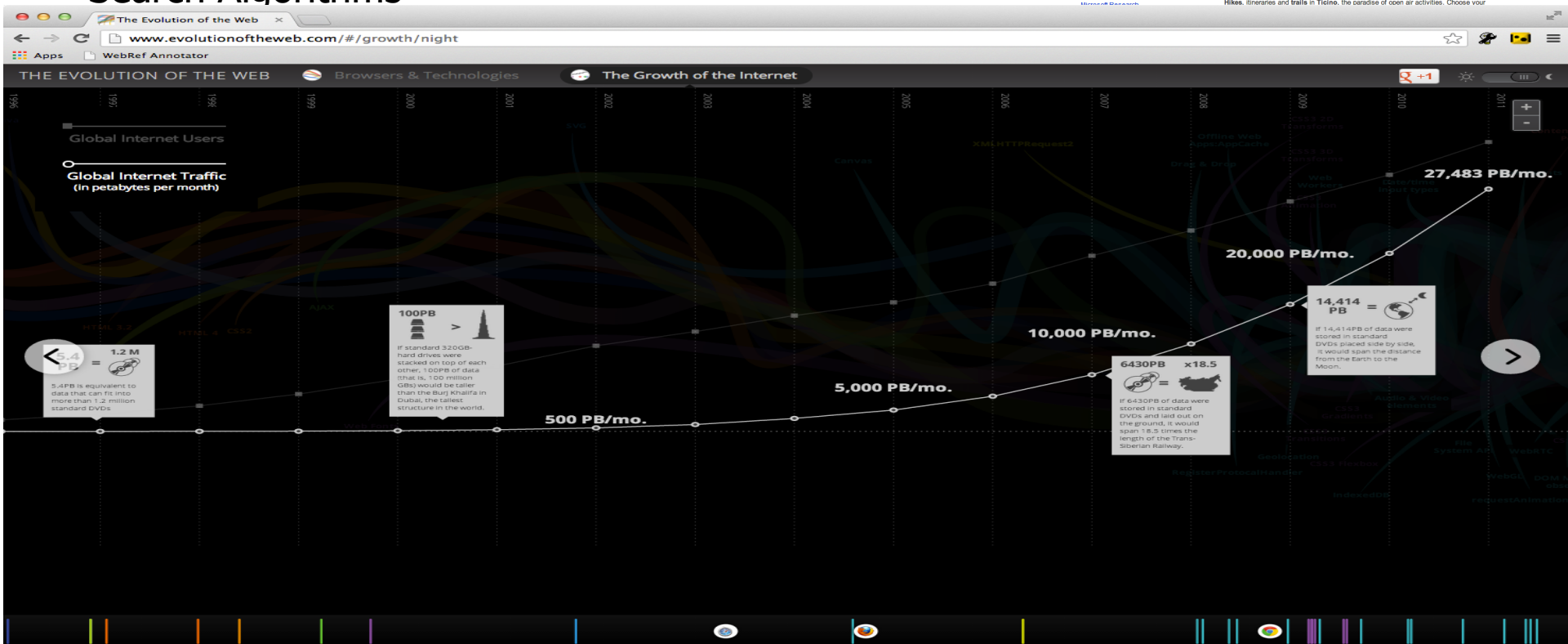
Sources of Variance in IR



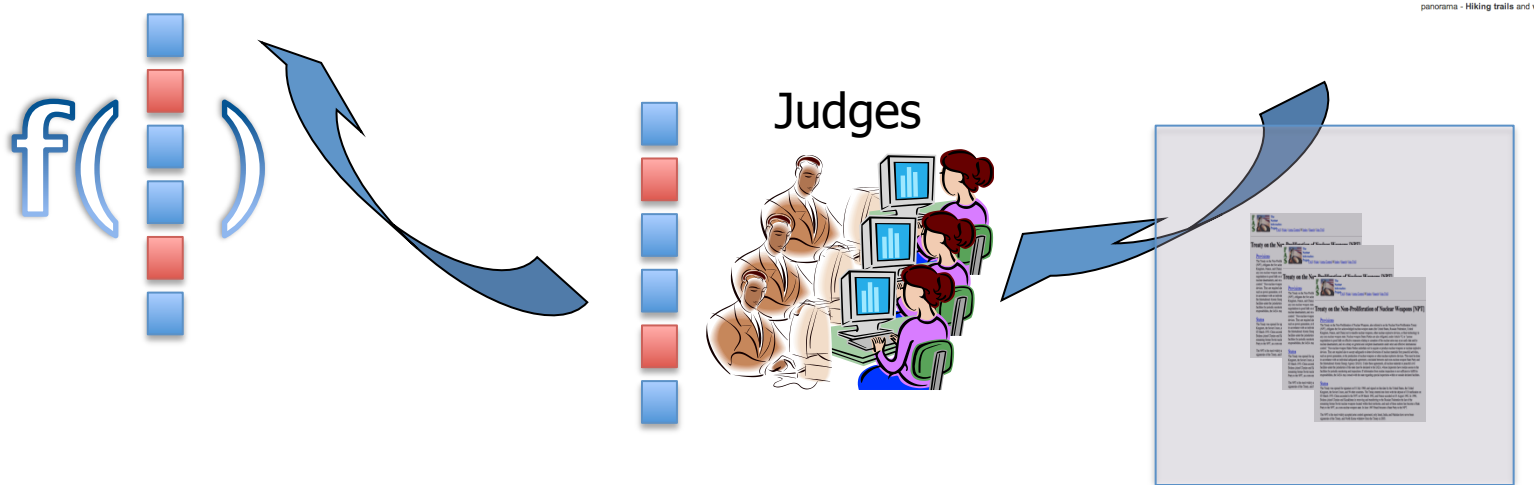
Results



Search Algorithms



panorama - **Hiking trails** and **walking** holidays .... **Hiking Trails Ticino** ...

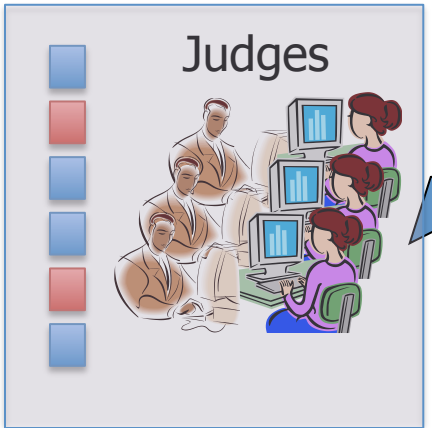


Clear all | Turn off

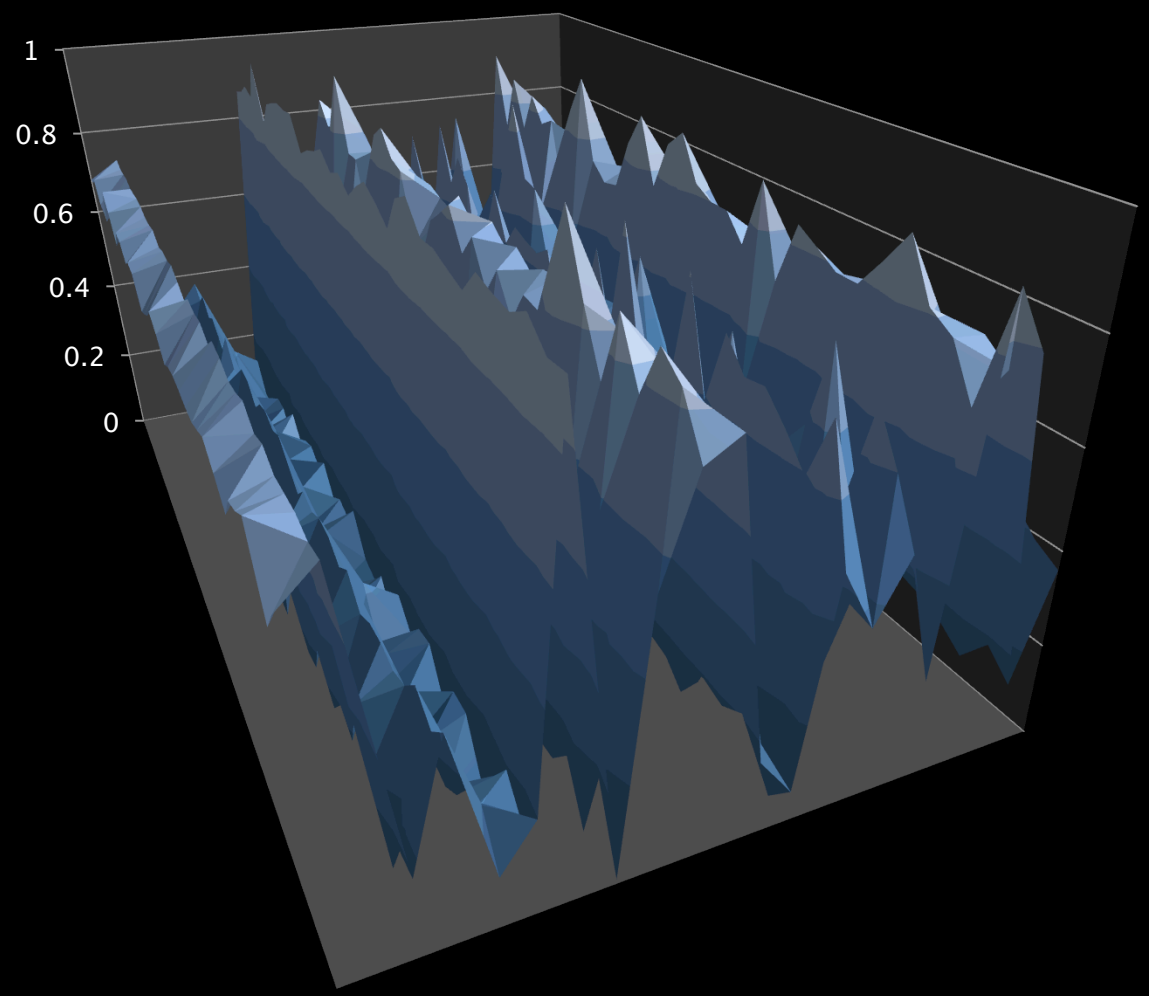
Ticino. Valais. Zürich Region. According to grade of difficulty: Conditions: .... The map also shows the entire **hiking trail** network (without additional information) ...

Ticino / Tessin : Climbing, **Hiking** & Mountaineering : SummitPost
www.summitpost.org › Alps - Western Part › Lepontine Alps
Nov 9, 2007 – **Ticino** / Tessin : SummitPost.org : Climbing, **hiking**, mountaineering. ... real beautiful **hiking trails** and challenging peaks waiting to be climbed.

Hiking Switzerland, **hiking trails** for **hikers**, holiday resorts in ...
www.switzerland-**hiking**.ch/
The website for **hikers** and nature-lovers with information on **hiking trails** and ... Swiss panorama - **Hiking trails** and **walking** holidays .... **Hiking Trails Ticino** ...



Multiple Sources of Variance

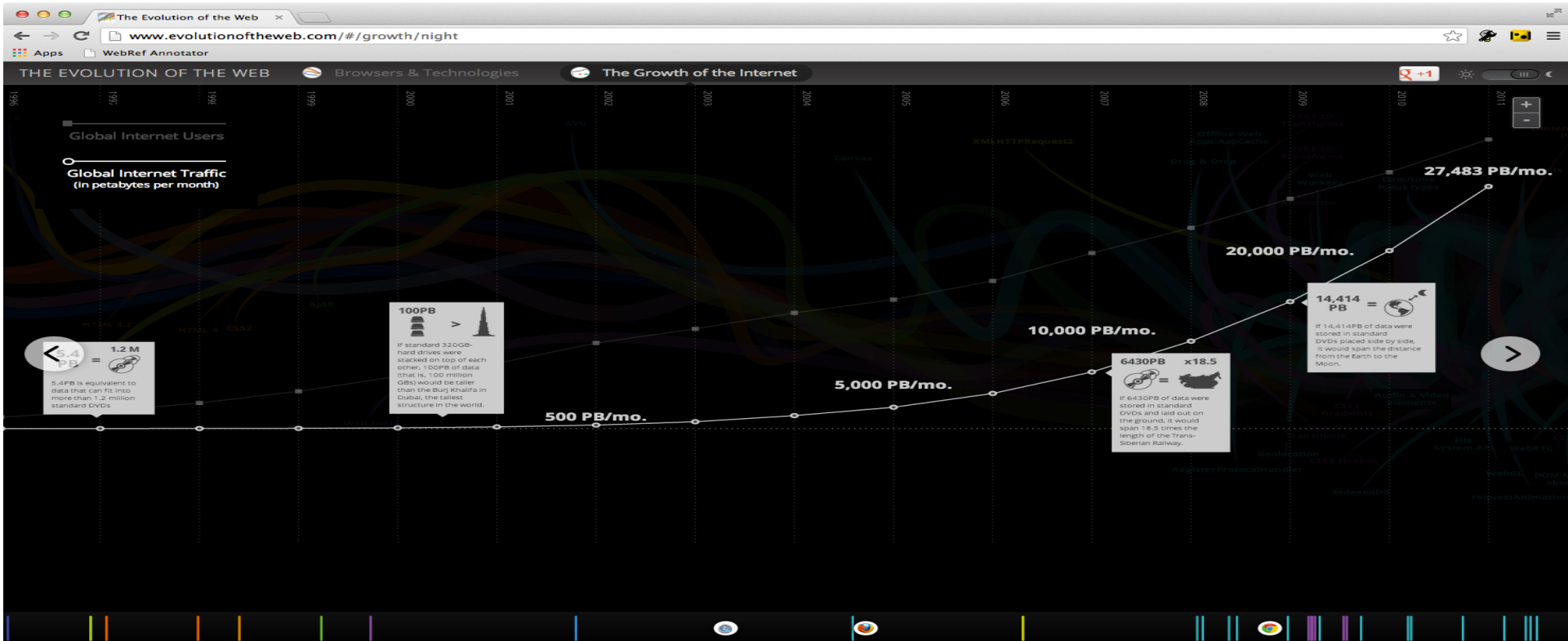


Variance due to Document Collection

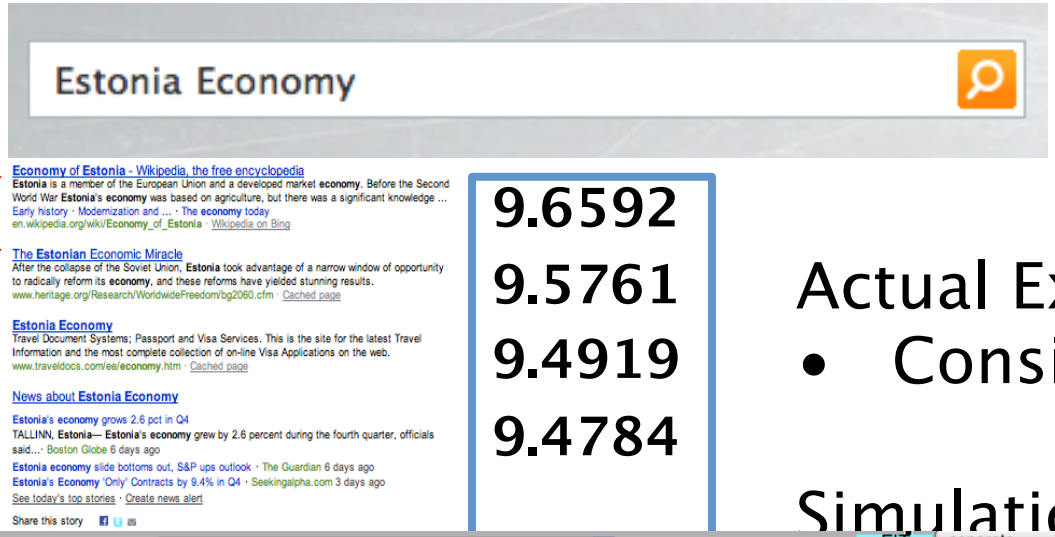
On Per-topic Variance in IR Evaluation

Stephen E. Robertson
Microsoft Research
7 JJ Thomson Avenue
Cambridge CB3 0FB, UK
stephenerobertson@hotmail.co.uk

Evangelos Kanoulas
Information School
University of Sheffield
Sheffield, UK
ekanoulas@gmail.com



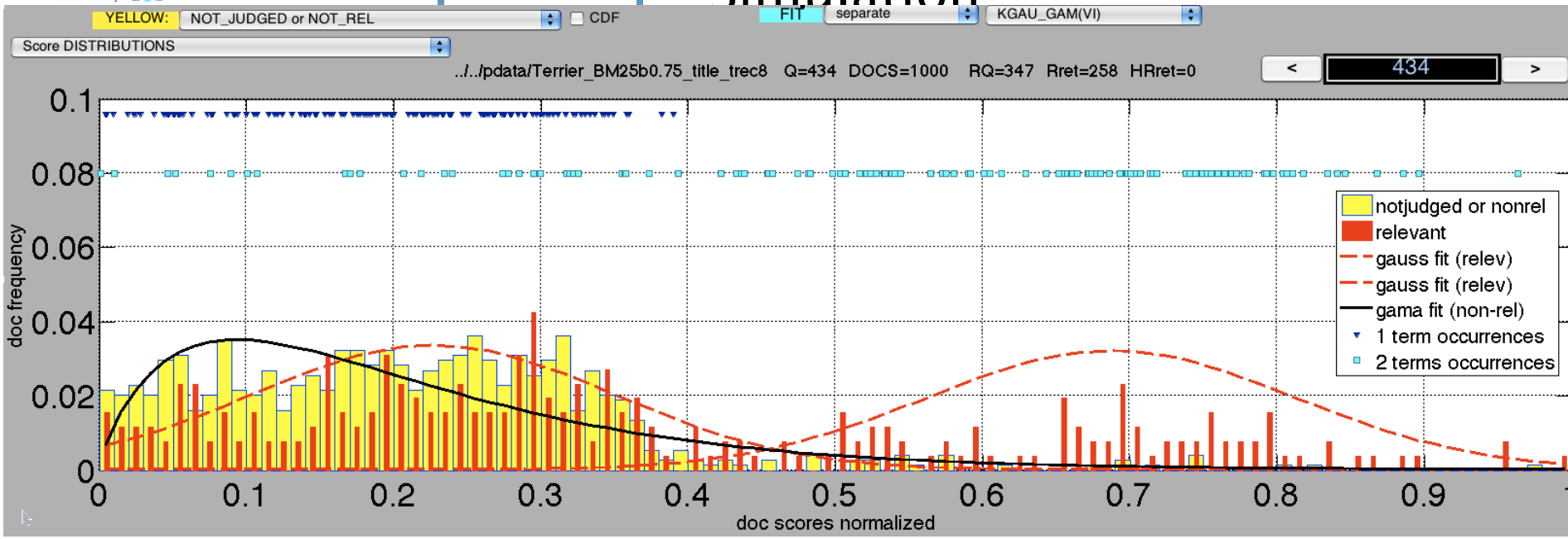
Simulation of multiple collections

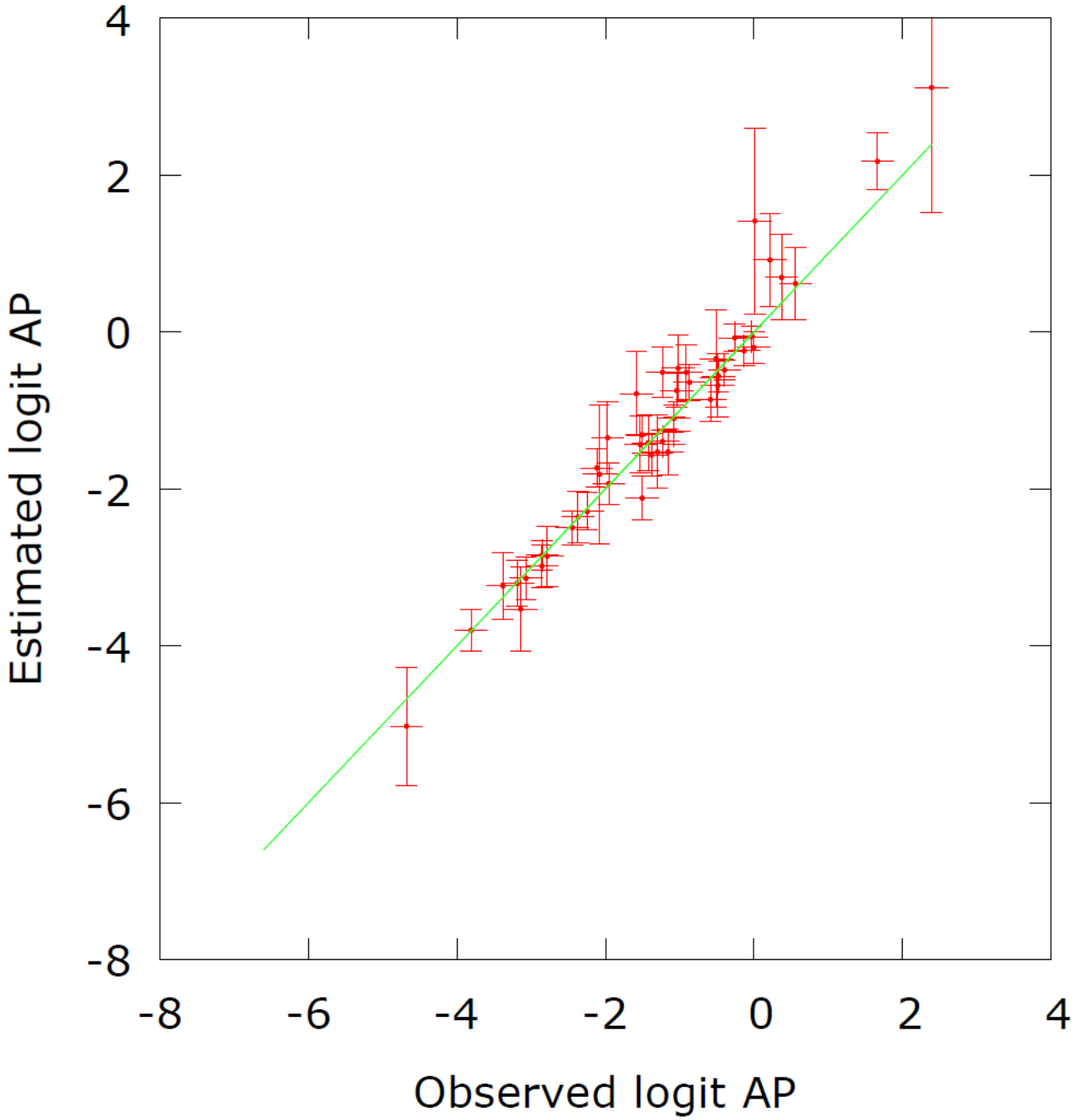


Actual Experiment:

- Consider multiple corpora

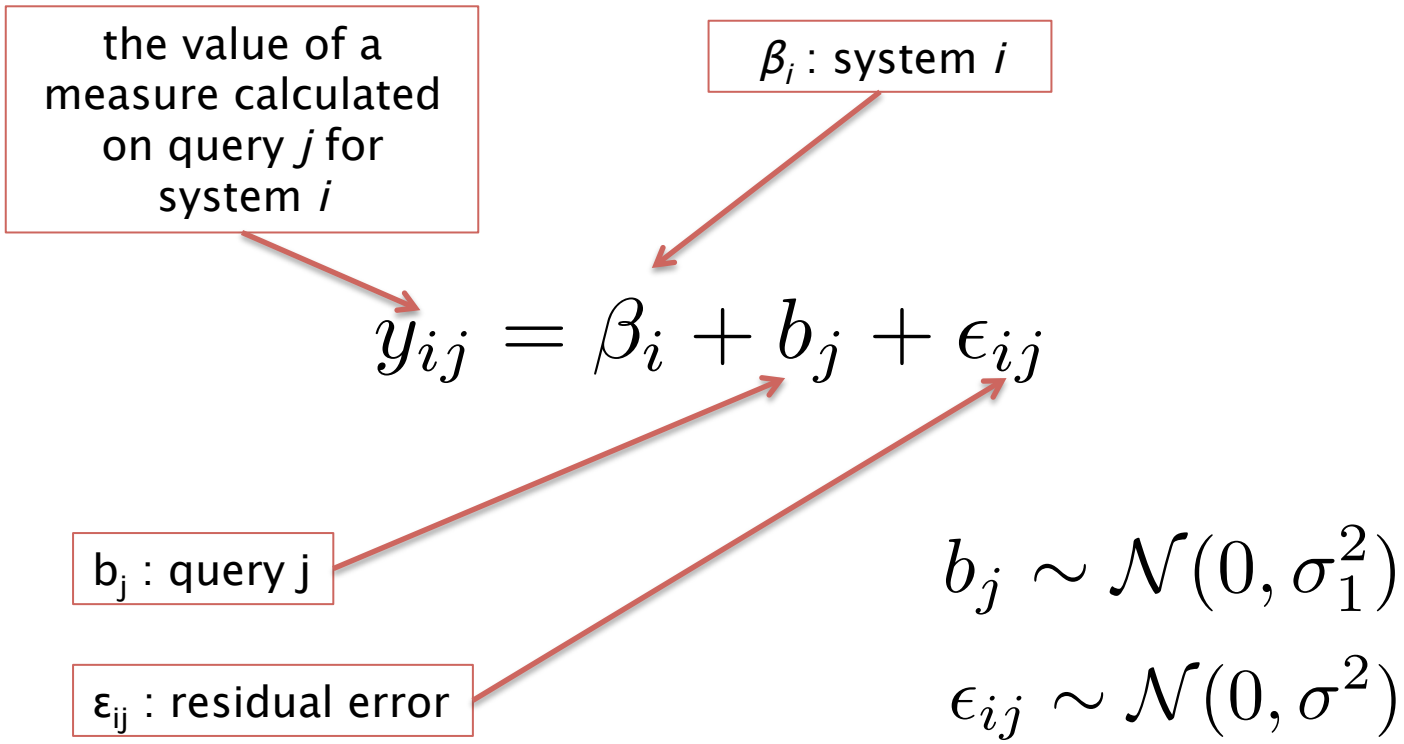
Simulation :




Estimated logit AP
Observed logit AP
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4

Mixed Effects Models

- Single source of variance
  - Query effect



Mixed Effects Models

- Two sources of variance
  - Query effect
  - Collection effect (within query variance)

$$y_{ijk} = \beta_i + b_j + c_{ij} + \epsilon_{ijk}$$

$$b_j \sim \mathcal{N}(0, \sigma_1^2) \qquad c_{ij} \sim \mathcal{N}(0, \sigma_2^2) \qquad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

Mixed Effects Models

- In the statistical programming environment **R**

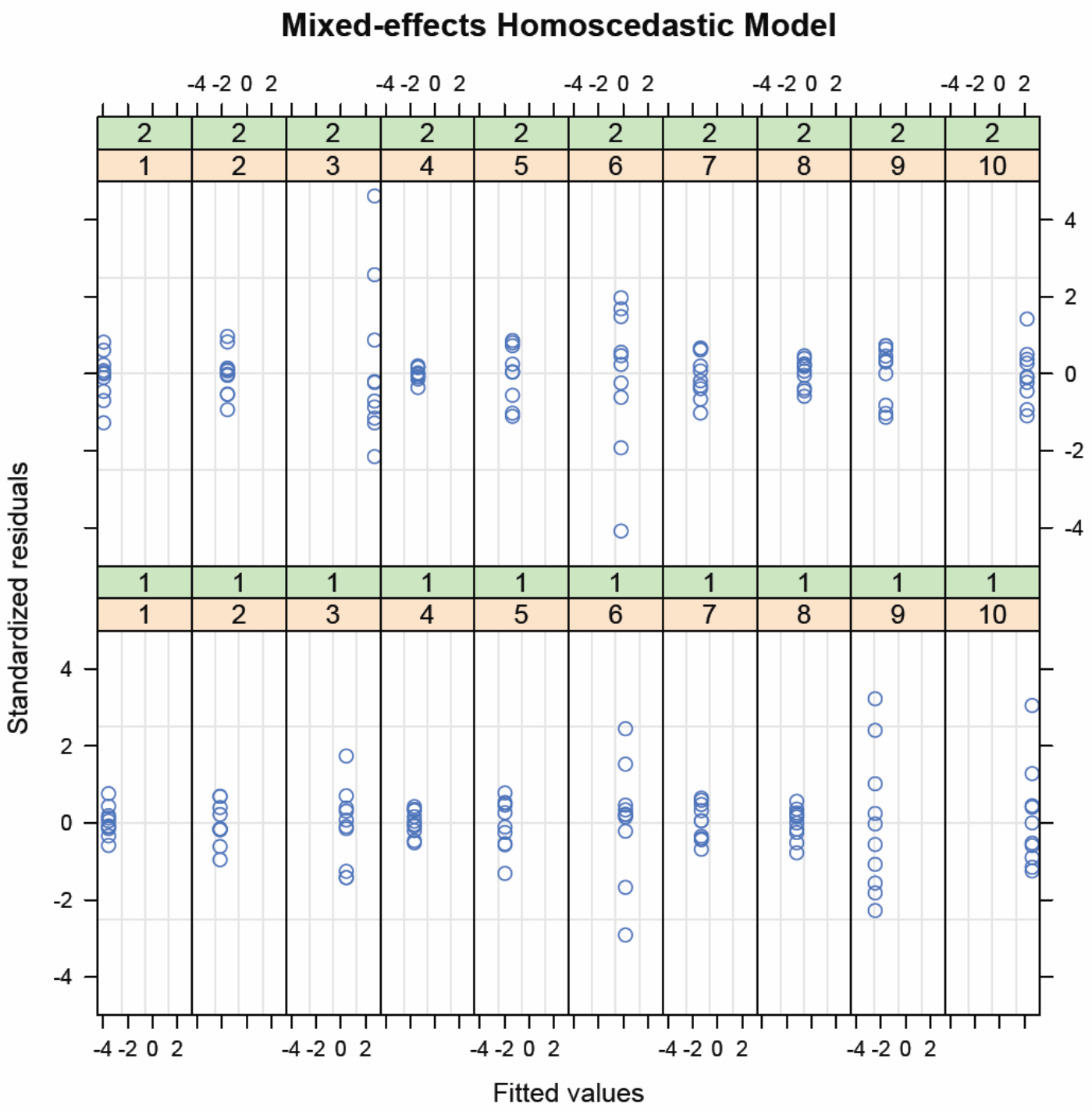
```
lme1 <- lme(effectiveness~system, data=df, random=~1|query/system)

summary(lme1)

Random effects:
   Formula: ~1 | query
            (Intercept)
StdDev:     1.539644

   Formula: ~1 | system %in% query
            (Intercept)  Residual
StdDev:
            0.6191864       0.6386645

Fixed effects: y ~ system
                   Value Std.Error    DF    t-value      p-value
(Intercept)  -1.3445 0.2438470 846     -5.514077 0.0000
system2           0.0999 0.1343512 46       0.744112 0.4606
```


Mixed-effects Homoscedastic Model
Standardized residuals
Fitted values
2 2 2 2 2 2 2 2 2 2
1 2 3 4 5 6 7 8 9 10
1 1 1 1 1 1 1 1 1 1
1 2 3 4 5 6 7 8 9 10
-4 -2 0 2
4
2
0
-2
-4

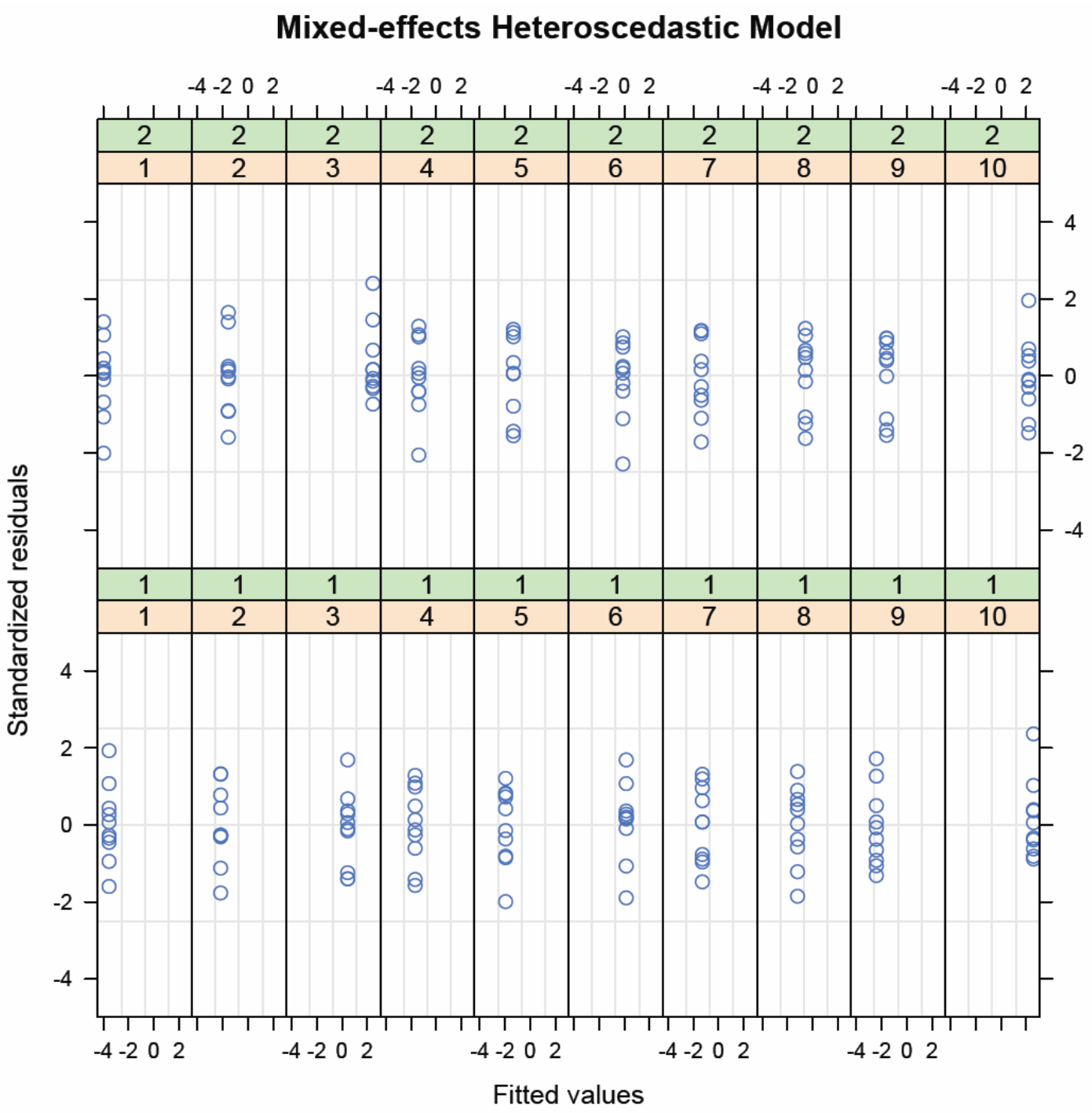
Mixed Effects Models

- Two sources of variance
  - Query effect
  - Collection effect (within query variance)

$$y_{ijk} = \beta_i + b_j + c_{ij} + \epsilon_{ijk}$$

$$b_j \sim \mathcal{N}(0, \sigma_1^2) \qquad c_{ij} \sim \mathcal{N}(0, \sigma_2^2) \qquad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma_{ij}^2)$$

Heteroscedastic Model


Mixed-effects Heteroscedastic Model
Standardized residuals
Fitted values
2 2 2 2 2 2 2 2 2 2
1 2 3 4 5 6 7 8 9 10
1 1 1 1 1 1 1 1 1 1
1 2 3 4 5 6 7 8 9 10
-4 -2 0 2
4
2
0
-2
-4

Mixed Effects Models

- In the statistical programming environment **R**

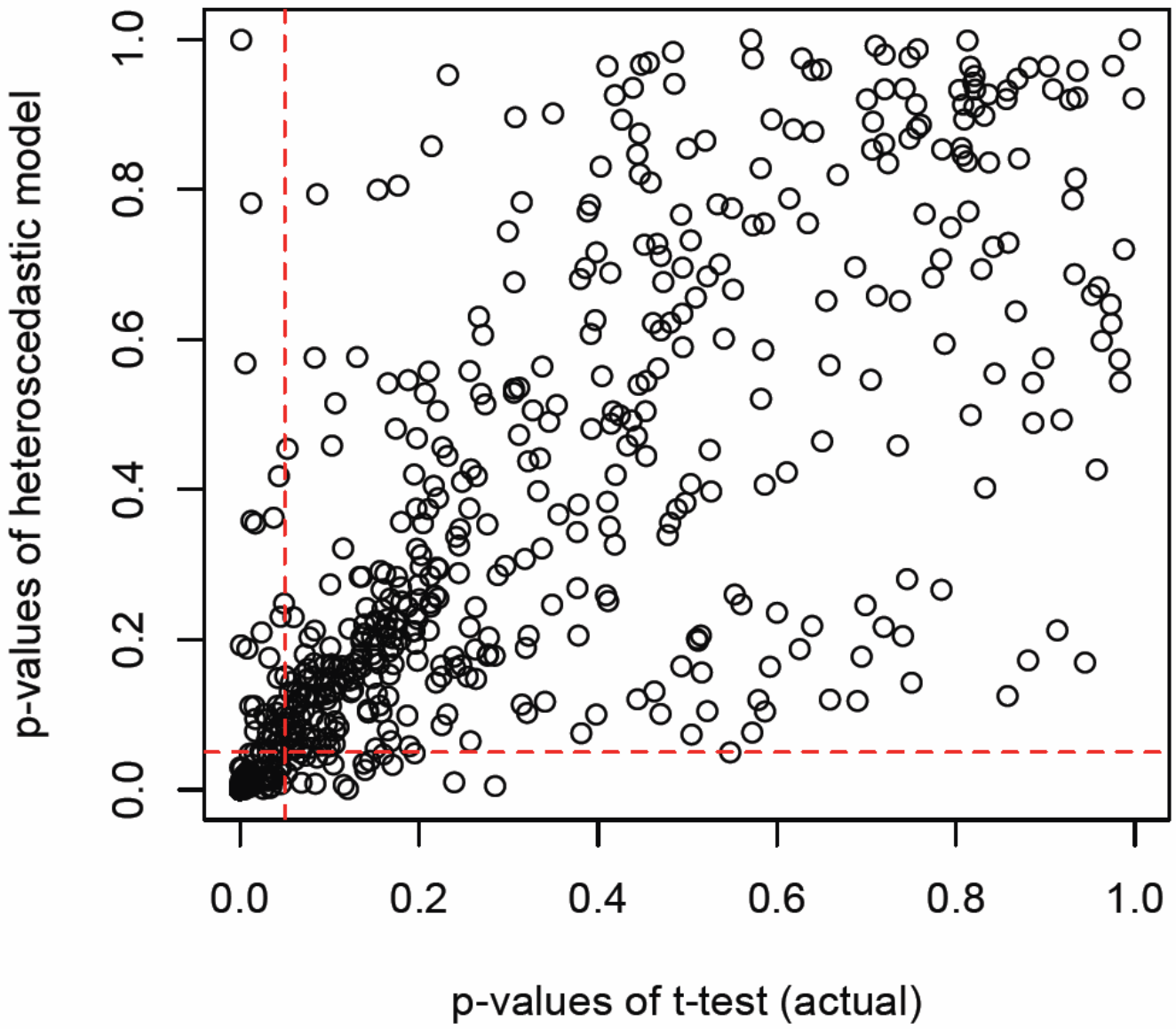
```
lme2 <- lme(effectiveness~system, data=df, random=~1|query/system,
      weights=varIdent(form=~1|query*system))
```

```
Random effects:
 Formula: ~1 | query
        (Intercept)
StdDev:    1.447164

 Formula: ~1 | system %in% query
        (Intercept)  Residual
StdDev:   0.4537618 0.186183
```

```
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | query* system
 Parameter estimates:
      1*1       1*2       2*1       2*2       ...
1.0000000 1.6108387 1.3969085 1.5405710 ...
```

```
Fixed effects: y ~ system
                Value  Std.Error  DF   t-value p-value
(Intercept)  -1.4385 0.22266286 846 -6.460817  0.0000
system2       0.1834 0.09844907  46  1.863342  0.0688
```


p-values of heteroscedastic model
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
p-values of t-test (actual)

Variance due to Measu

Simulating Simple User Behavior for System Effectiveness Evaluation

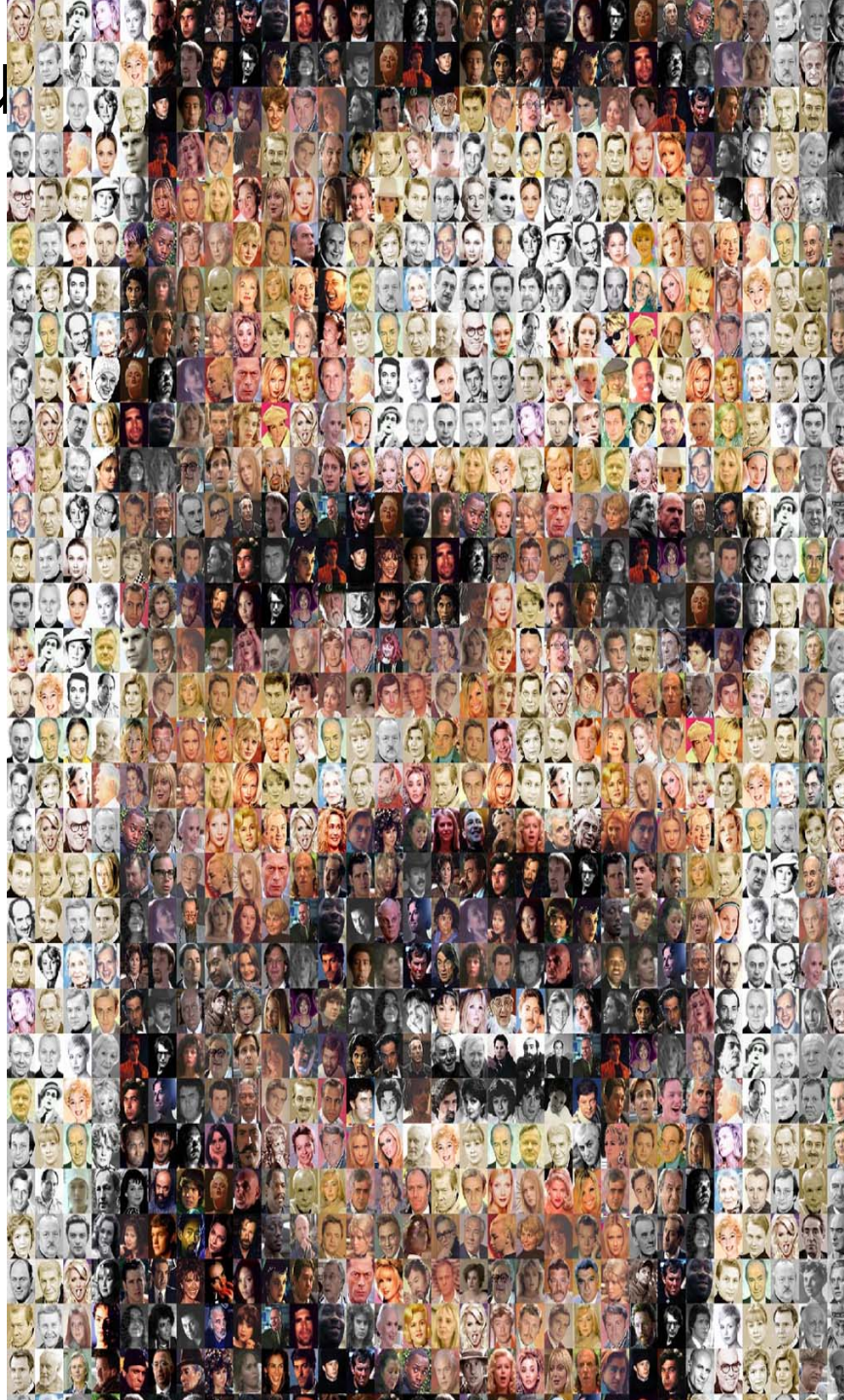
Ben Carterette*, Evangelos Kanoulas‡, Emine Yilmaz†
carteret@cis.udel.edu, e.kanoulas@shef.ac.uk, {eminey@microsoft.com, eyilmaz@ku.edu.tr}

* Department of Computer & Information Sciences, University of Delaware, Newark, DE
‡ Information School, University of Sheffield, Sheffield, UK
† Microsoft Research, Cambridge, UK & Koc University, Istanbul, Turkey

Incorporating Variability in User Behavior into Systems Based Evaluation

Ben Carterette*, Evangelos Kanoulas‡, Emine Yilmaz†
carteret@cis.udel.edu, ekanoulas@gmail.com, {eminey@microsoft.com, eyilmaz@ku.edu.tr}

* Department of Computer & Information Sciences, University of Delaware, Newark, DE
‡ Information School, University of Sheffield, Sheffield, UK
† Microsoft Research, Cambridge, UK & Koc University, Istanbul, Turkey



Mixed Effects Models

$$y_{ijk} = \alpha_i + (\beta_j + \phi_j p_k) + (\kappa_{ij} + \gamma_{ij} p_k) + \epsilon_{ijk}$$

y_{ijk}: value of a metric on topic j for system i with parameter p_k

α_i : effect of system i

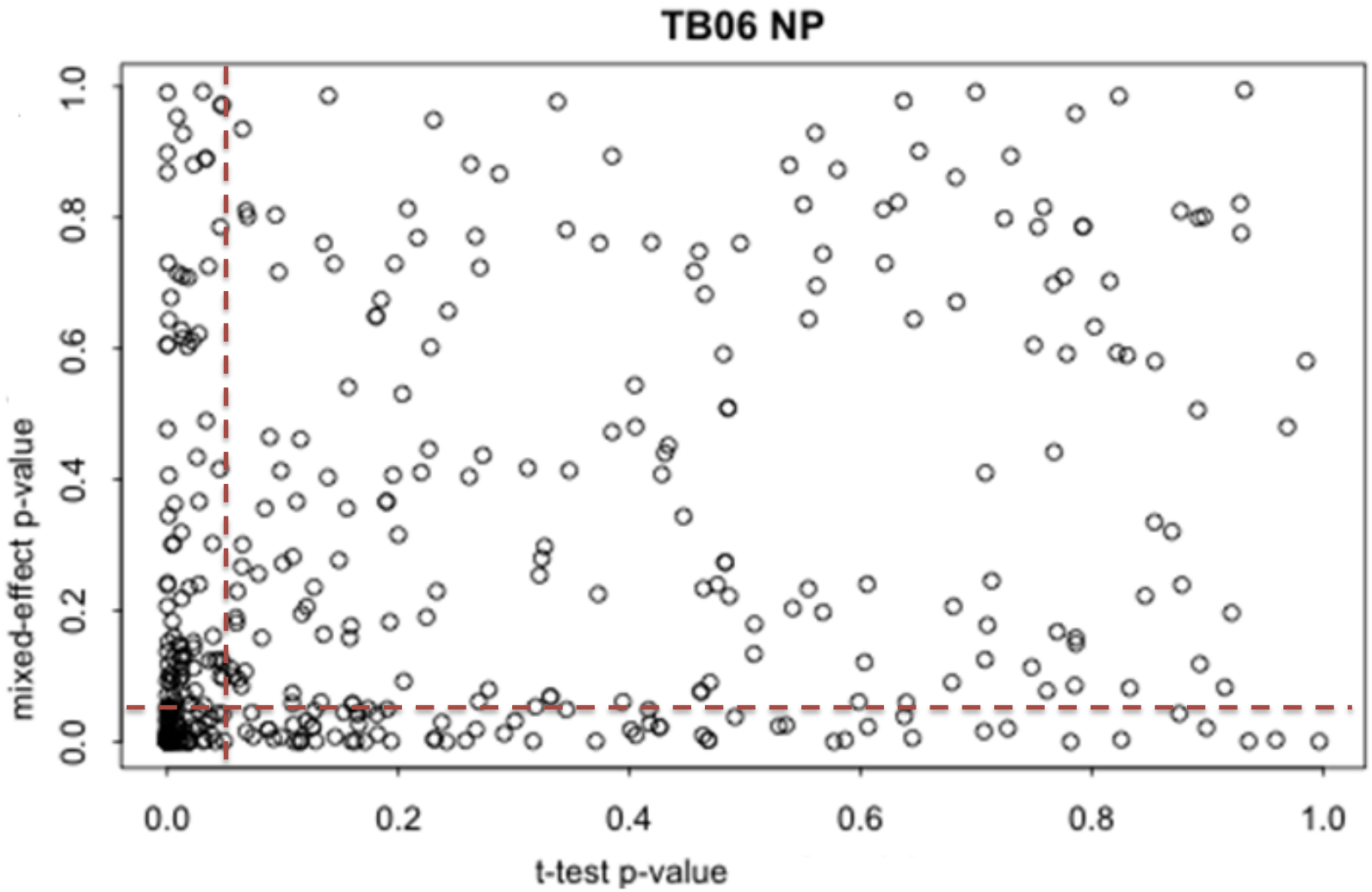
β_j : effect of topic j

$\varphi_j p_k$: interaction of topic with RBP parameter

κ_{ij} : system/topic interaction effect

$\gamma_{ij} p_k$: interaction of system/topic with RBP parameter

ε_{ijk} : system/topic/parameter interaction effect


TB06 NP
mixed-effect p-value
t-test p-value
0.0
0.2
0.4
0.6
0.8
1.0

Variance due to Query Intents

Effect of Intent Descriptions on Retrieval Evaluation

Emine Yilmaz
Dept. of Computer Science
University College London
e.yilmaz@cs.ucl.ac.uk

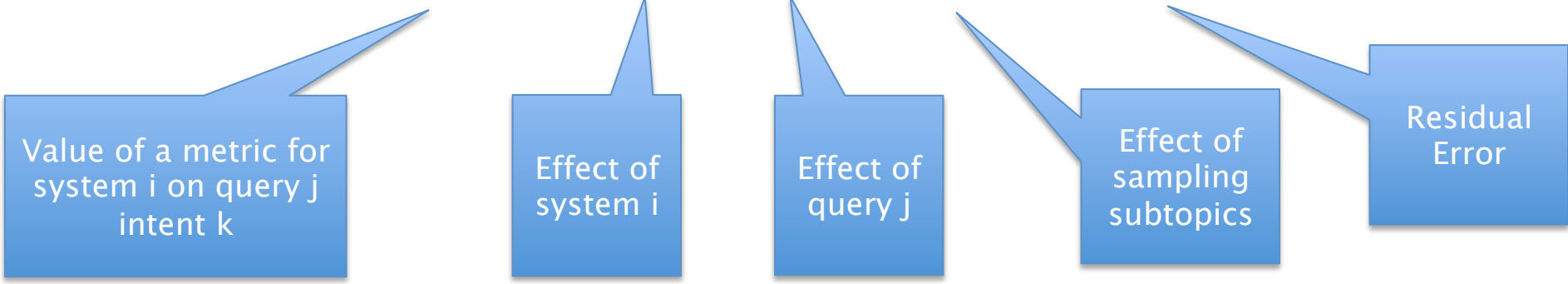
Evangelos Kanoulas
Google
ekanoulas@gmail.com

Nick Craswell
Microsoft
nickcr@microsoft.com

Mixed Effects Models

- Two sources of variance
  - Queries
  - Intents (within queries)

$$y_{ijk} = \beta_i + b_j + c_{ij} + \varepsilon_{ijk}$$



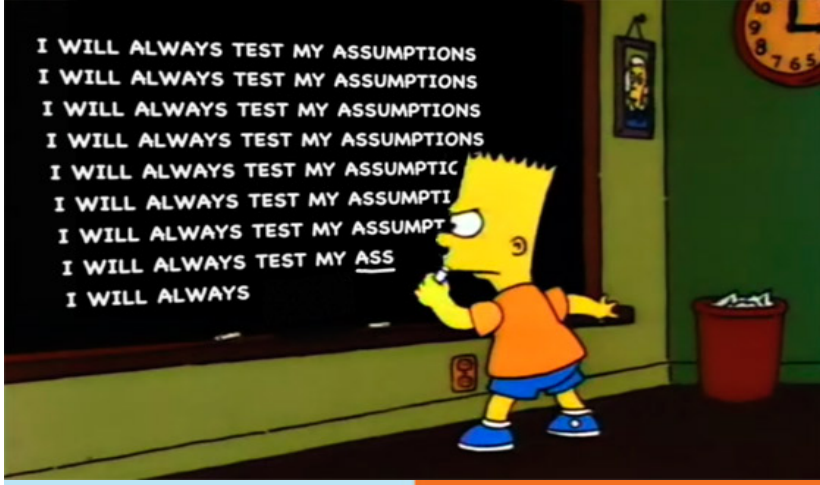
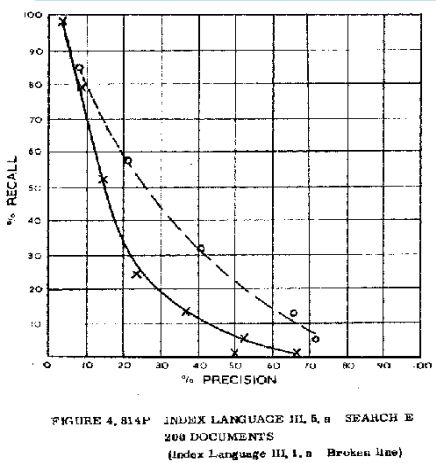
$$b_j \sim N(0,\sigma_1^2), \quad c_{ij} \sim N(0,\sigma_2^2), \quad \varepsilon_{ijk} \sim N(0,\sigma^2)$$

Summary of comparative evaluation

- Always do significance tests
  - Model all the effects
  - Check your assumptions
- Don't just report p-values or * to indicate significance
  - Always report confidence intervals
- Always take results of tests with a grain of salt
  - Especially when the effect size is low
  - Build your intuition and use it

Challenges in collection-based evaluation

- Do users and judges agree on relevance?
  - Particularly difficult for personalized search
  - Particularly difficult for specialized documents
- Do summary aggregate scores reflect users satisfaction?
  - Do real users agree with MAP@1000? NDCG@5?
- Ambiguous queries are particularly hard to judge realistically
  - Which intent is most popular? Which others are important?
  - Judges need to correctly appreciate uncertainty
- How do you identify when relevance changes?
  - Temporal changes: Document changes; Query intent changes
- It's expensive to collect new labels

Different approaches to evaluation

- User-studies
- Collection-based evaluation
- In-situ evaluation
  - A/B Testing
  - Interleaving