Large Scale IR Evaluation

Virgil Pavlu

LARGE SCALE INFORMATION RETRIEVAL EVALUATION

Large Scale IR Evaluation

Large Scale IR Evaluation

Large Scale IR Evaluation

- People are not well organized
  - nor consistent with each other

Large Scale IR Evaluation

- People are not well organized
  - nor consistent with each other
- People are publishing everything

Large Scale IR Evaluation

- People are not well organized
  - nor consistent with each other
- People are publishing everything
- Today, Culture, Commerce, Science and Military depend on ability to store and make use of information

Large Scale IR Evaluation

- People are not well organized
  - nor consistent with each other
- People are publishing everything
- Today, Culture, Commerce, Science and Military depend on ability to store and make use of information
- IR field is dedicated to organization of information
  - search is the principal component

Large Scale IR Evaluation

- IR vs Databases

	Databases	IR
Data	**Structured**	**Unstructured**
Fields	**Clear semantics** (SSN, age)	**No fields** (other than text)
Queries	**Defined** (relational algebra, SQL)	**Free text** ("natural language"), Boolean
Recoverability	**Critical** (concurrency control, recovery, atomic operations)	**Downplayed**, though still an issue
Matching	**Exact** (results are *always* "correct")	**Imprecise** (need to measure effectiveness)

CMPSCI 646

Large Scale IR Evaluation

Large Scale IR Evaluation

- Critical for research
- Commercially used for optimization
- Can measure many aspects of returned results
- Text REtrieval Conference
  - queries, collections, search engines, performance
  - many tracks every year
  - judges several 100K documents

Large Scale IR Evaluation

INTRODUCE

2 METHODS
FOR EVALUATION

THAT DEALS WITH LARGE

Large Scale IR Evaluation

- Current size of datasets make traditional methods less applicable
  - Terabyte GOV2 collection has 25M documents
  - On the web, focus is on top of the list

INTRODUCE

2 METHODS
FOR EVALUATION

THAT DEALS WITH LARGE

Large Scale IR Evaluation

- Current size of datasets make traditional methods less applicable
  - Terabyte GOV2 collection has 25M documents
  - On the web, focus is on top of the list
- Assuming a list of results is summary-able, we should be able to interpret measurements as statistics
  - what is the point of retrieving 30000 d

INTRODUCE

2 METHODS
FOR EVALUATION

THAT DEALS WITH LARGE

overview

- Introduction
- Relevance Prior
- Hedge
- Sampling
- Future work

my contribution

Precision and Recall

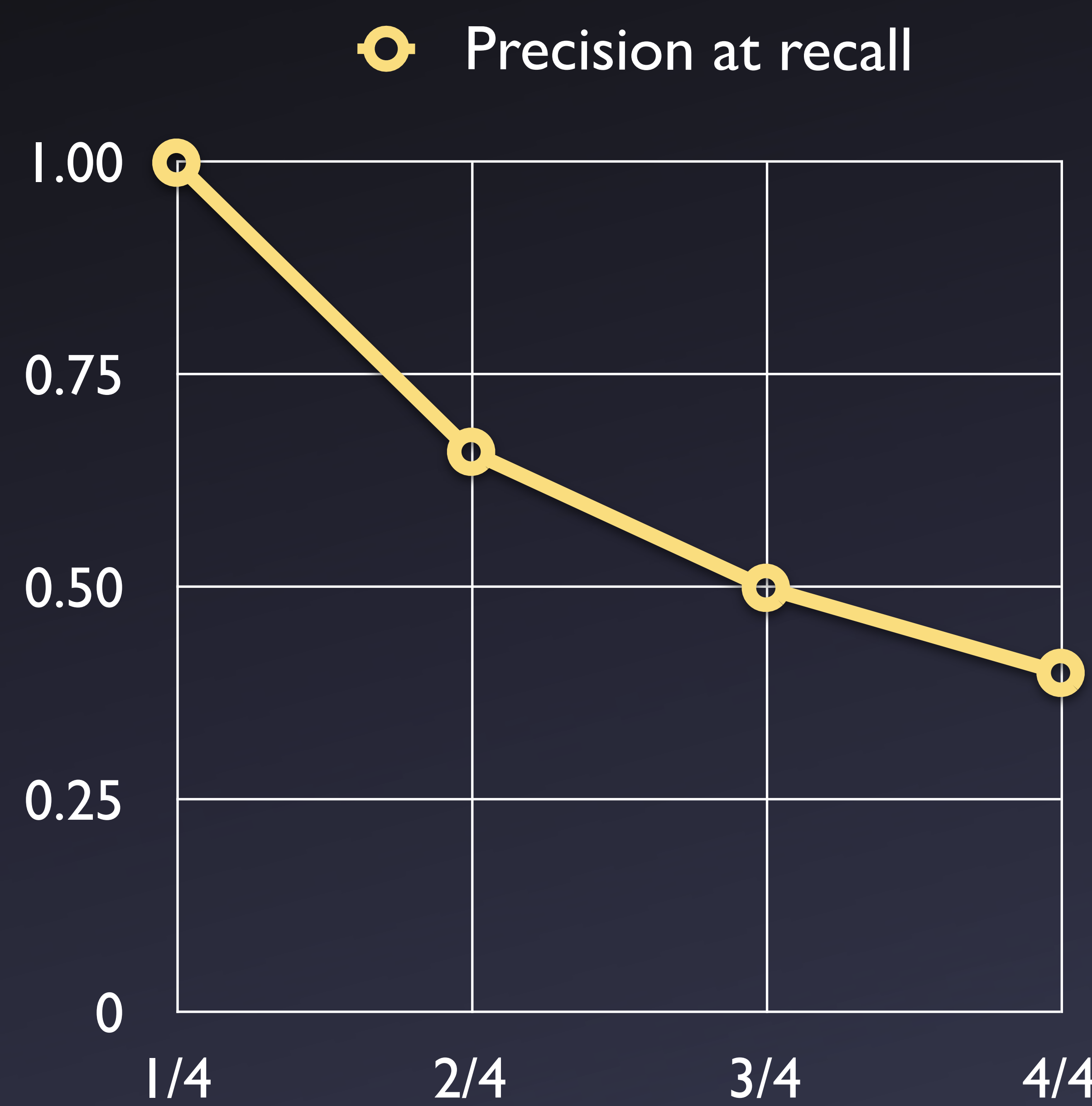
List:

	precision	recall
R	1/1	1/4
N	1/2	1/4
R	2/3	2/4
N	2/4	2/4
N	2/5	2/4
R	3/6	3/4
N	3/7	3/4
N	3/8	3/4
N	3/9	3/4
R	4/10	4/4

Precision and Recall

List:

	precision	recall
R	1/1	1/4
N	1/2	1/4
R	2/3	2/4
N	2/4	2/4
N	2/5	2/4
R	3/6	3/4
N	3/7	3/4
N	3/8	3/4
N	3/9	3/4
R	4/10	4/4



7

Average Precision

List:

	precision
R	1/1
N	
R	2/3
N	
N	
R	3/6
N	
N	
N	
R	4/10

- AP = average of precisions at relevant ranks
  - use 0 for relevant documents not returned

change in AP for 1 rank

$$\mathrm{AP} = \frac{1 + 2 \ldots /10}{\ldots} \approx 0.6417$$

relevance prior

- Assess the relative importance of ranks
  - should be monotonic decreasing
  - can we use search engines scores ?
- Many developed
  - zipfian, logarithmic, logistic regression etc

AP relevance prior

- Sum-Precision is the numerator of AP
  - denominator is query-constant

rel = relevant (0,1)

$$SP = \sum_{i\,:\,rel(i)=1} Prec(i) = \sum_{i=1}^{Z} rel(i) \cdot Prec(i)$$

$$= \sum_{i=1}^{Z} rel(i) \sum_{j=1}^{i} rel(j)/i = \sum_{1 \le j \le i \le Z} \frac{1}{i} \cdot rel(i) \cdot rel(j)$$

AP relevance prior

- Sum-Precision is the numerator of AP
  - denominator is query-constant

rel = relevant (0,1)

$$SP = \sum_{i\,:\,rel(i)=1} Prec(i) = \sum_{i=1}^{Z} rel(i) \cdot Prec(i)$$

$$= \sum_{i=1}^{Z} rel(i) \sum_{j=1}^{i} rel(j)/i = \sum_{1 \le j \le i \le Z} \frac{1}{i} \cdot rel(i) \cdot rel(j)$$

- Infer a weighting scheme for ranks

	1	2	3	$\dots$	Z
1	1				
2	$\frac{1}{2}$	$\frac{1}{2}$			
3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$		
$\vdots$					
Z	$\frac{1}{Z}$	$\frac{1}{Z}$	$\frac{1}{Z}$	$\dots$	$\frac{1}{Z}$

	1	2	3	$\dots$	Z
1	2	$\frac{1}{2}$	$\frac{1}{3}$	$\dots$	$\frac{1}{Z}$
2	$\frac{1}{2}$	1	$\frac{1}{3}$	$\dots$	$\frac{1}{Z}$
3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\dots$	$\frac{1}{Z}$
$\vdots$					
Z	$\frac{1}{Z}$	$\frac{1}{Z}$	$\frac{1}{Z}$	$\dots$	$\frac{2}{Z}$

10

AP relevance prior

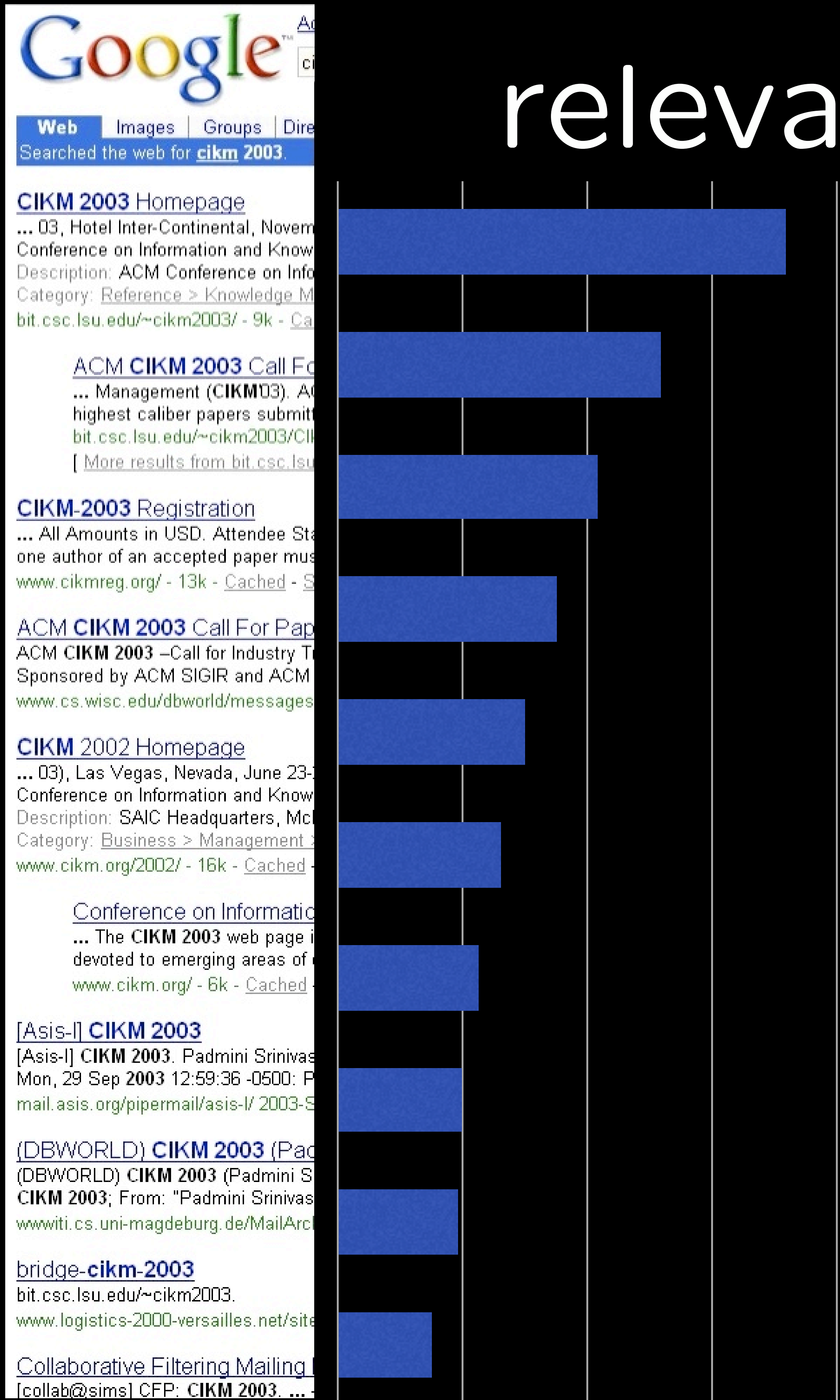
- After marginalization

$$\mathrm{W}(r) = \frac{1}{2Z}\left(1 + \frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{Z}\right)$$

  - future work : raised at 3/2 for variance reduction

- If we have many systems we can derive an average prior over documents

$$M(i) = \frac{1}{N}\sum_{s} W_s(rank(i,s))$$



relevance prior

$$W(r) = \frac{1}{2Z}\left(1+\frac{1}{r}+\frac{1}{r+1}+\cdots+\frac{1}{Z}\right)$$

12

overview

- Introduction
- Relevance Prior
- Hedge
- Sampling
- Future work

Hedge online learning

Hedge online learning

- Say I want to invest in stocks
  - quick return : sell after a week or less
  - repeat over a long period of time
  - invest a fixed amount every day

14

Hedge online learning

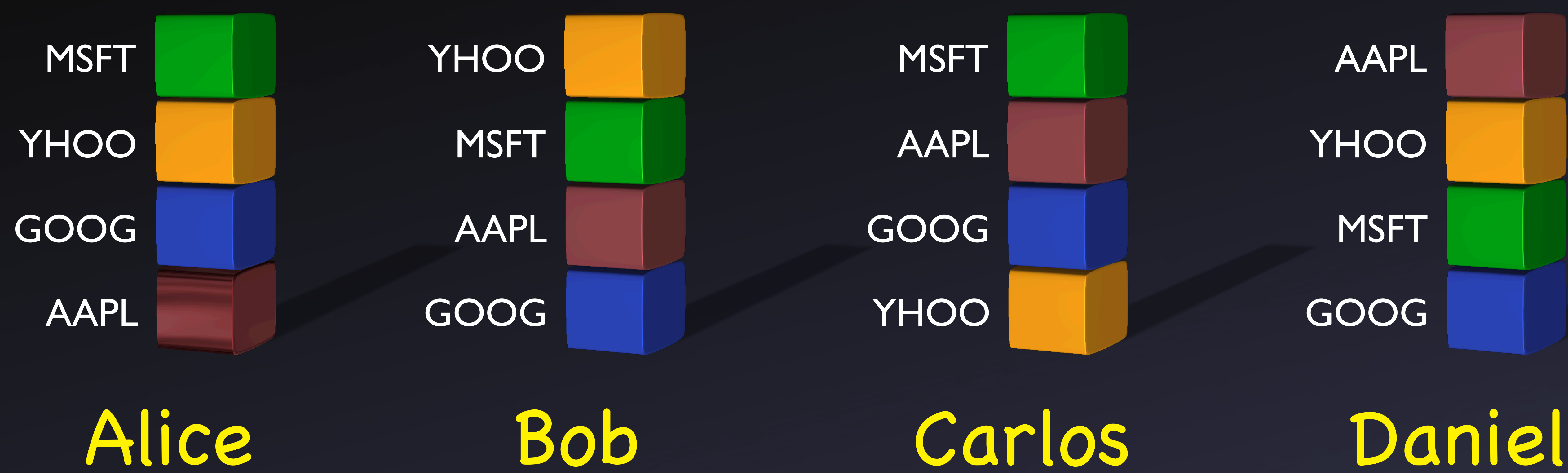
- Say I want to invest in stocks
  - quick return : sell after a week or less
  - repeat over a long period of time
  - invest a fixed amount every day
- I have 4 friends brokers
  - Alice, Bob, Carlos and Daniel
  - each recommends a ranked list
  - I decide to ask all 4 for advice every time

combining experts

- Day 1 : which stock should I buy ?

combining experts

- Day 1 : which stock should I buy ?



15

combining experts

- Day 1 : which stock should I buy ?

- Maybe I will buy MSFT

episodic loss

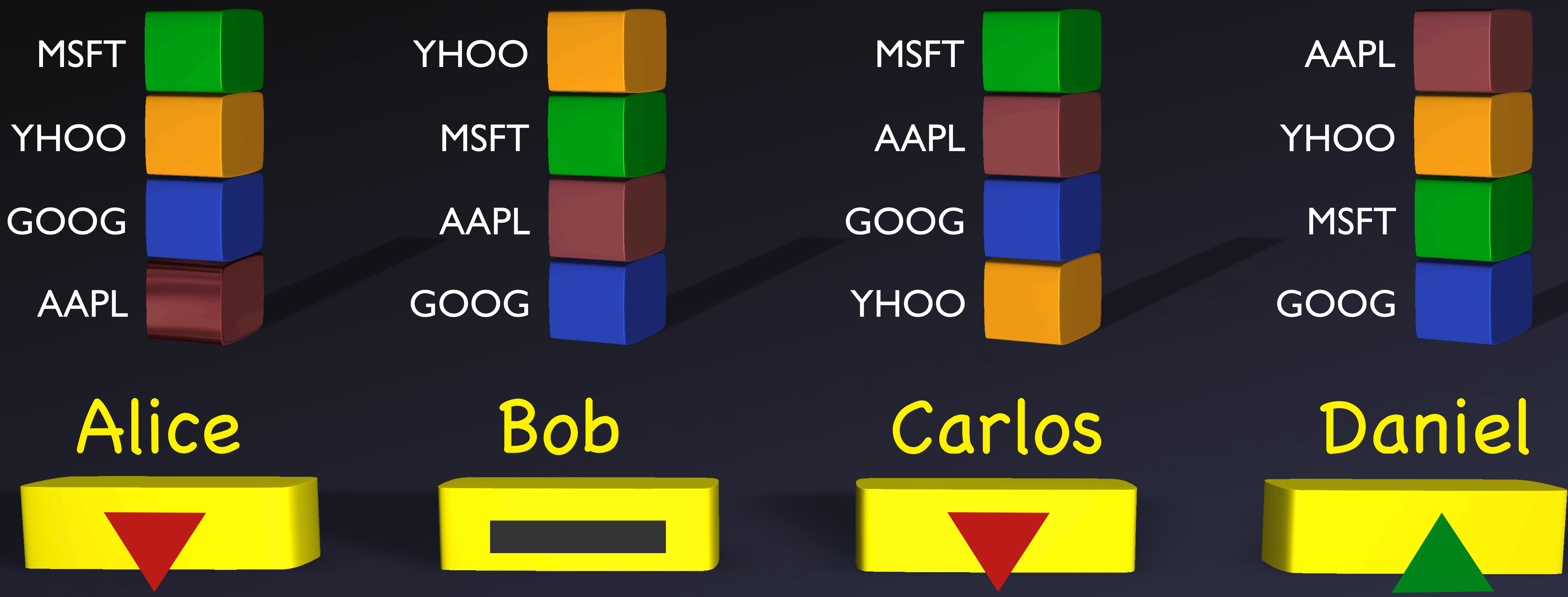
- Next day : MSFT announces buying YHOO
  - on expectation that they will ruin YMail, Ymessenger, Flickr etc MSFT stock falls
- I will reflect the loss into the confidence I have in each broker, based on the recommended list

episodic loss

- day 1 recommendations:

episodic loss

- day 1 recommendations:



update confidence

- after day 1:

18

update confidence

- after day 1:

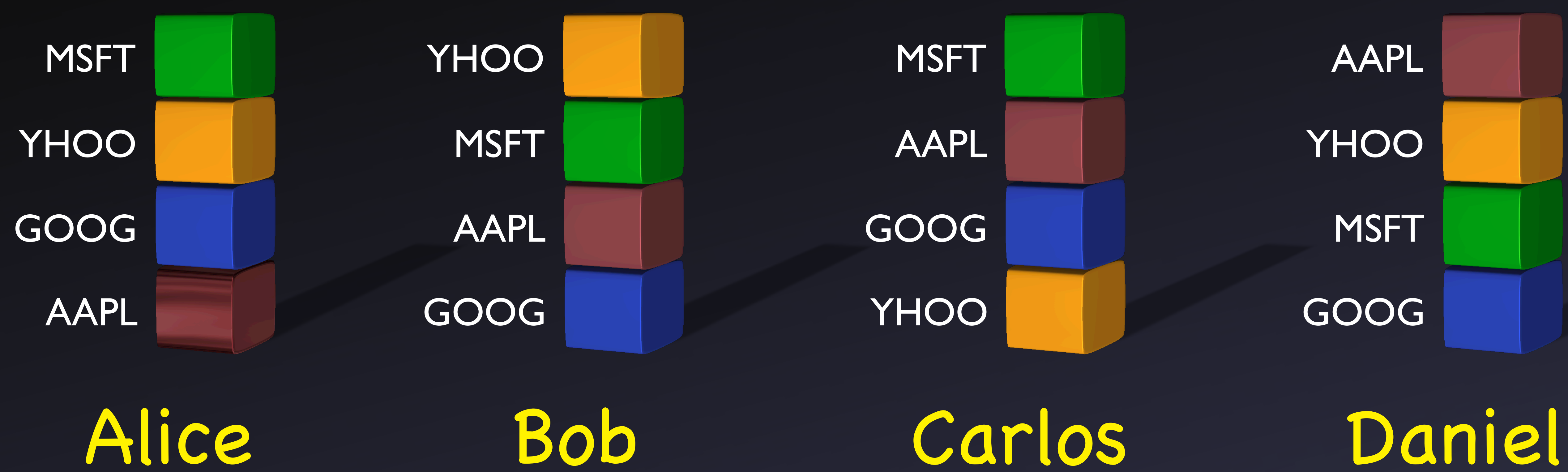
combining experts

- Day 2 : which stock should I buy ?

19

combining experts

- Day 2 : which stock should I buy ?



19

combining experts

- Day 2 : which stock should I buy ?

- Maybe I will buy AAPL

19

cumulative performance

- Next : AAPL most-expected IPhone gets delayed
  - I lose money, update confidence etc

20

cumulative performance

- Next : AAPL most-expected IPhone gets delayed
  - I lose money, update confidence etc
- But if at least one broker is any good (and consistent) I will make money overall

20

cumulative performance

- Next : AAPL most-expected IPhone gets delayed
  - I lose money, update confidence etc
- But if at least one broker is any good (and consistent) I will make money overall
- At the end of the year I compare my cumulative loss/gain with the overall best-broker

$$\text{Loss}_{Hedge} \leq \frac{\min_b\{L_b\} \cdot \ln(1/\beta) + \ln(N)}{1-\beta}$$

  - future work on β

adaptation to IR setup

HEDGE

- given : experts
- incoming : loses
- reweight: experts

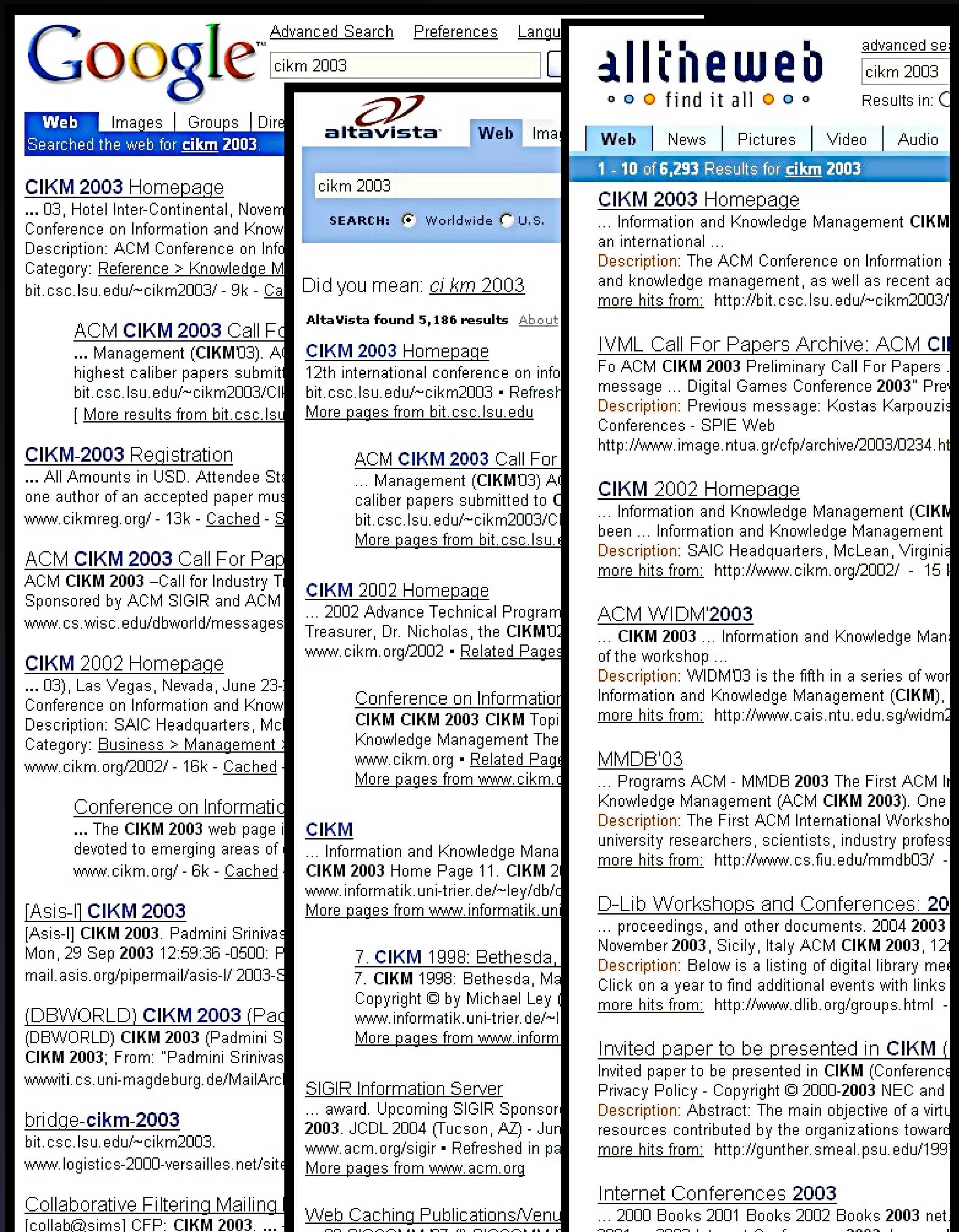
IR

- given : search engines
- incoming: documents
  - compute losses
- reweight : search engines

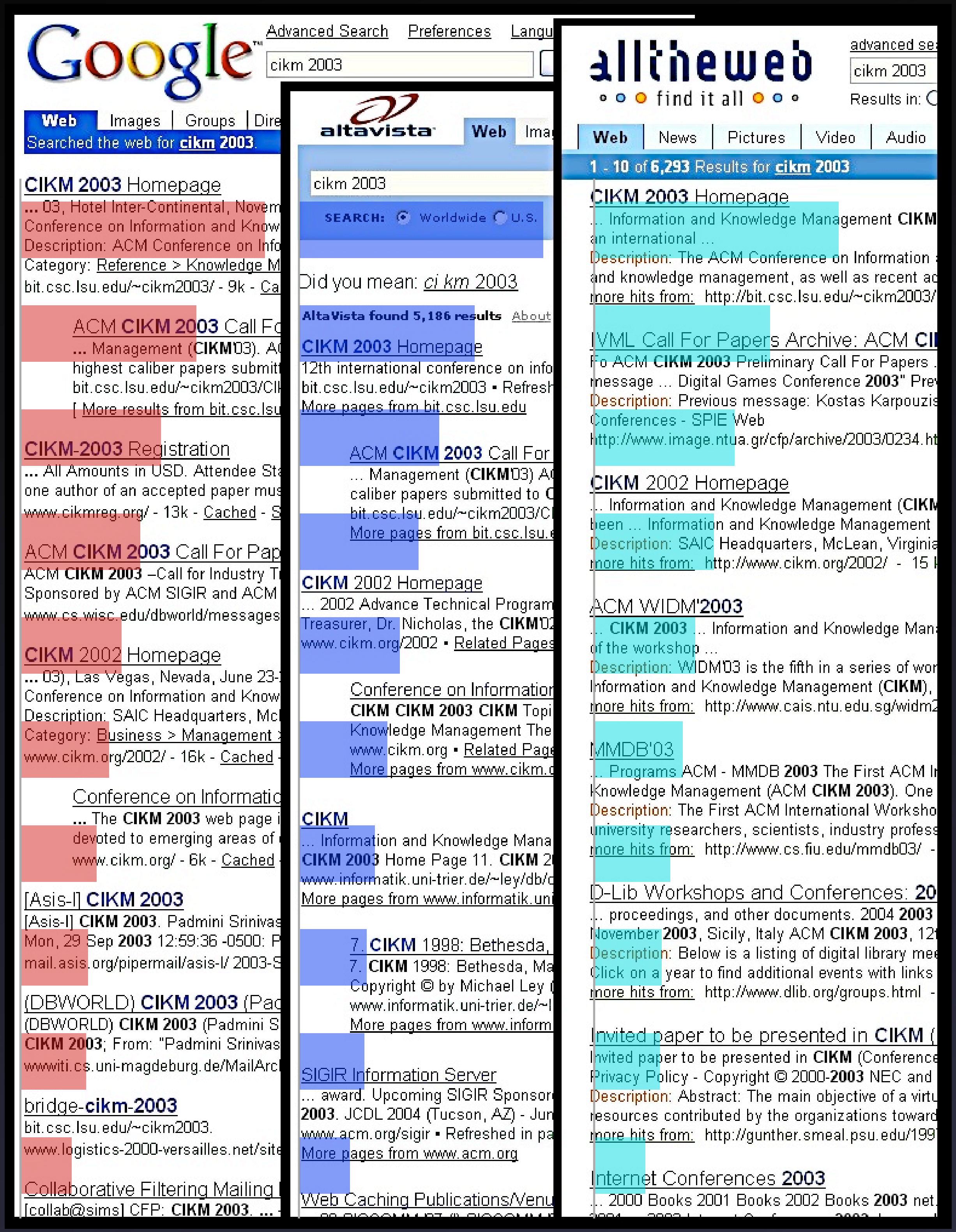
metasearch

- Given several lists returned on the same query, combine them in a single list
- Goal: perform as good as possible
  - usually compares with best system
- Rank-based or Score-based
- Several known techniques
  - Comb MNZ
  - Borda
  - Condorcet

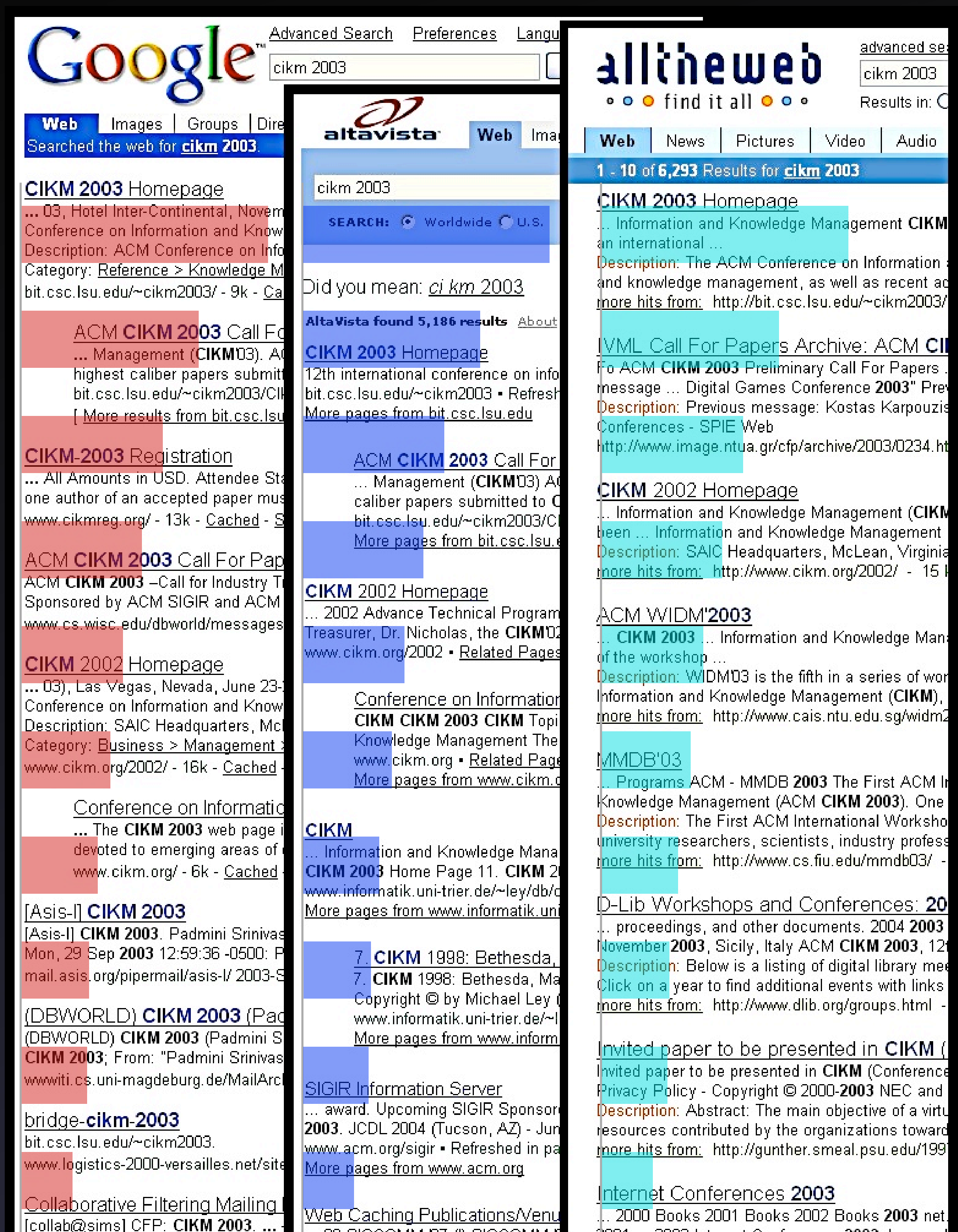
Hedge adaptation to IR setup



Hedge adaptation to IR setup



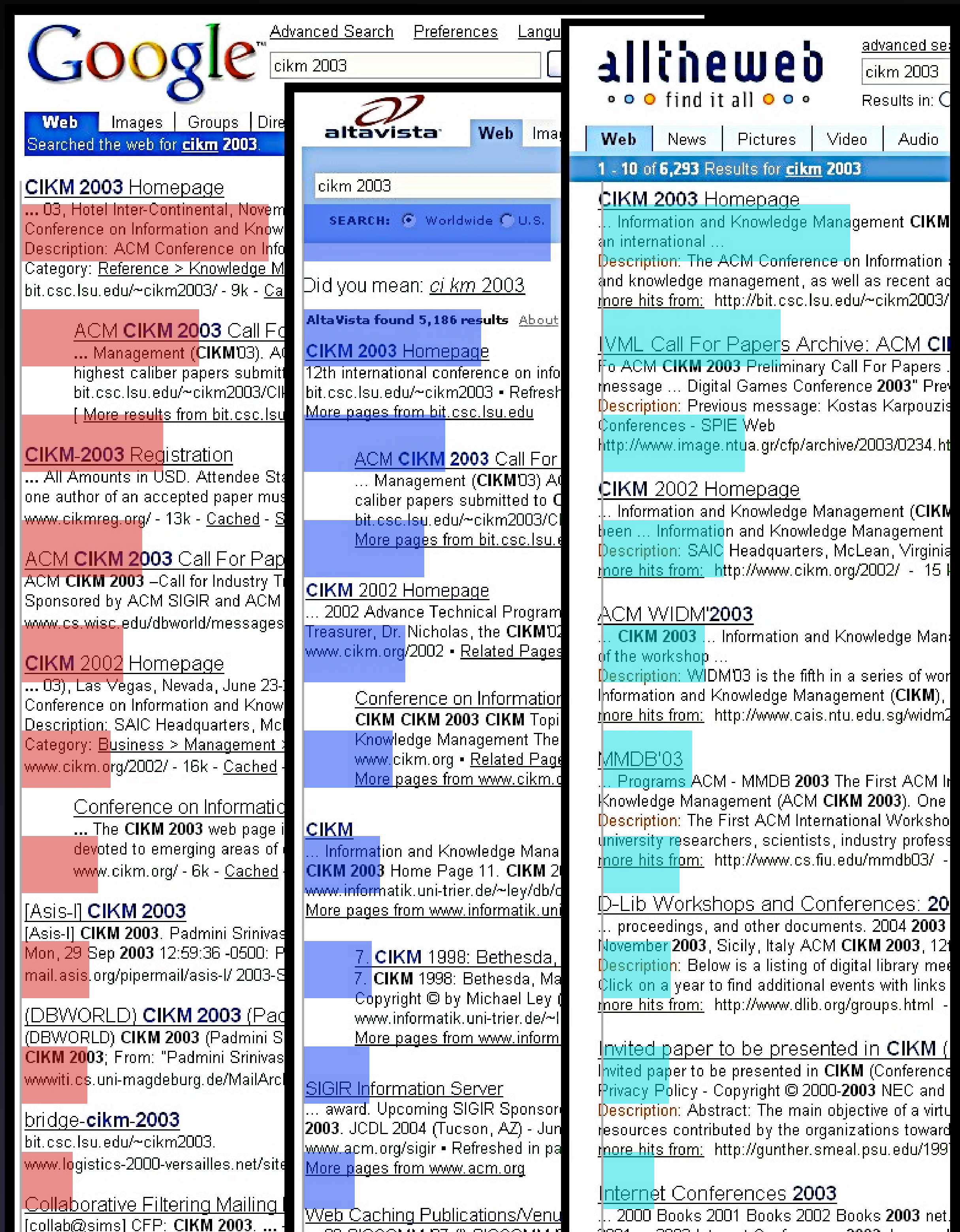
Hedge adaptation to IR setup



metasearch

CIKM 2003 Homepage
ACM CIKM 2003 Call For Pape
CIKM–2003 Registration
ACM CIKM 2003 Call For Pape
CIKM 2002 Homepage
Conference on Information an
[Asis-l] CIKM 2003
(DBWORLD) CIKM 2003 (Padm
bridge–cikm–2003
Collaborative Filtering Mailing
CIKM Home Page ACM DL: CIK
Yahoo! Groups : webir Messag
dBforums – Cfp: Cikm 03
Mailing List ARL–ERESERVE@a
Received: from cni.org by b.cr
ACM WIDM 2003
ACM – MMDB 2003
ACM CIKM 2003 PRELIMINARY
Selected Publications
Web Caching Publications/Ver
cikm '03
Call for papers
Mario A. Nascimento – Person
Iceved
Conferences and Journals on
Conferences On Information V
Calendrier des manifestations
(DBWORLD) Final Call for ACM
Collaborative Filtering Mailing

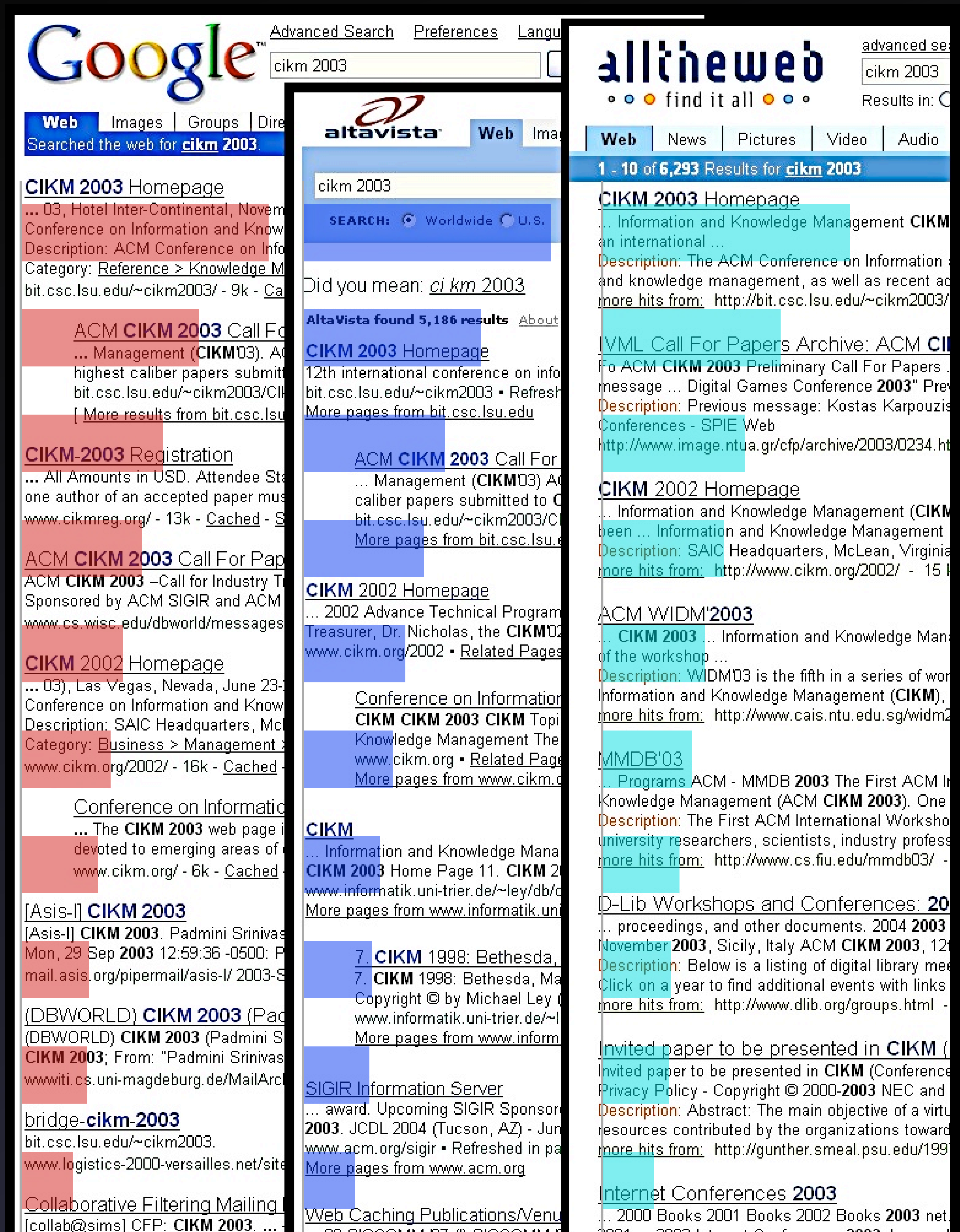
Hedge adaptation to IR setup



metasearch

- CIKM 2003 Homepage
- ACM CIKM 2003 Call For Pape
- CIKM–2003 Registration
- ACM CIKM 2003 Call For Pape
- CIKM 2002 Homepage
- Conference on Information an
- [Asis-l] CIKM 2003
- (DBWORLD) CIKM 2003 (Padm
- bridge–cikm–2003
- Collaborative Filtering Mailing
- CIKM Home Page ACM DL: CIK
- Yahoo! Groups : webir Messag
- dBforums – Cfp: Cikm 03
- Mailing List ARL–ERESERVE@a
- Received: from cni.org by b.cr
- ACM WIDM 2003
- ACM – MMDB 2003
- ACM CIKM 2003 PRELIMINARY
- Selected Publications
- Web Caching Publications/Ver
- cikm '03
- Call for papers
- Mario A. Nascimento – Person
- Iceved
- Conferences and Journals on
- Conferences On Information
- Calendrier des manifestations
- (DBWORLD) Final Call for ACM
- Collaborative Filtering Mailing

Hedge adaptation to IR setup



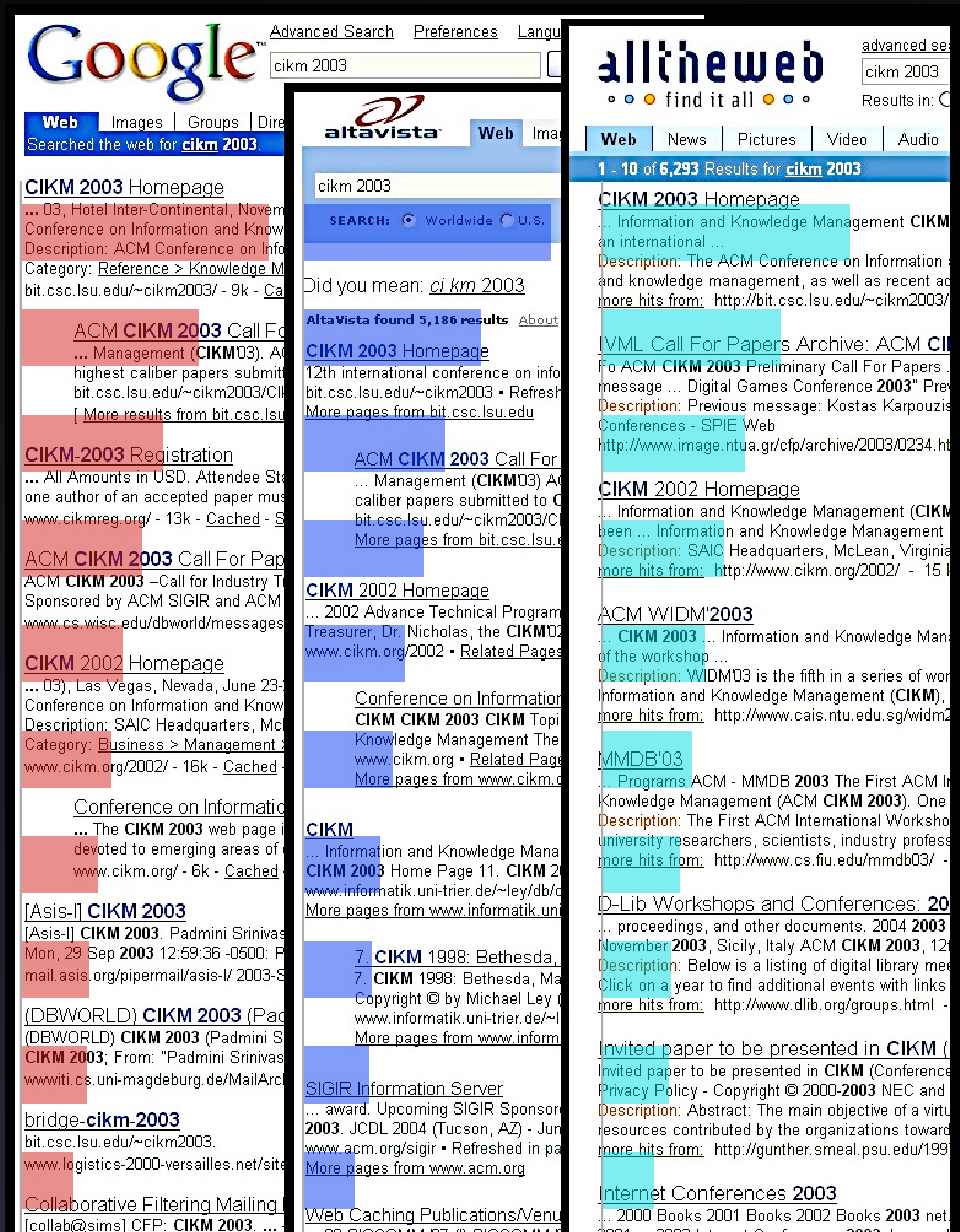
metasearch

- CIKM 2003 Homepage
- ACM CIKM 2003 Call For Pape
- CIKM–2003 Registration
- ACM CIKM 2003 Call For Pape
- CIKM 2002 Homepage
- Conference on Information an
- [Asis–l] CIKM 2003
- (DBWORLD) CIKM 2003 (Padm
- bridge–cikm–2003
- Collaborative Filtering Mailing
- CIKM Home Page ACM DL: CIK
- Yahoo! Groups : webir Messag
- dBforums – Cfp: Cikm 03
- Mailing List ARL–ERESERVE@ar
- Received: from cni.org by b.cr
- ACM WIDM 2003
- ACM – MMDB 2003
- ACM CIKM 2003 PRELIMINARY
- Selected Publications
- Web Caching Publications/Ver
- cikm '03
- Call for papers
- Mario A. Nascimento – Person
- Iceved
- Conferences and Journals on T
- Conferences On Information V
- Calendrier des manifestations
- (DBWORLD) Final Call for ACM
- Collaborative Filtering Mailing

metasearch

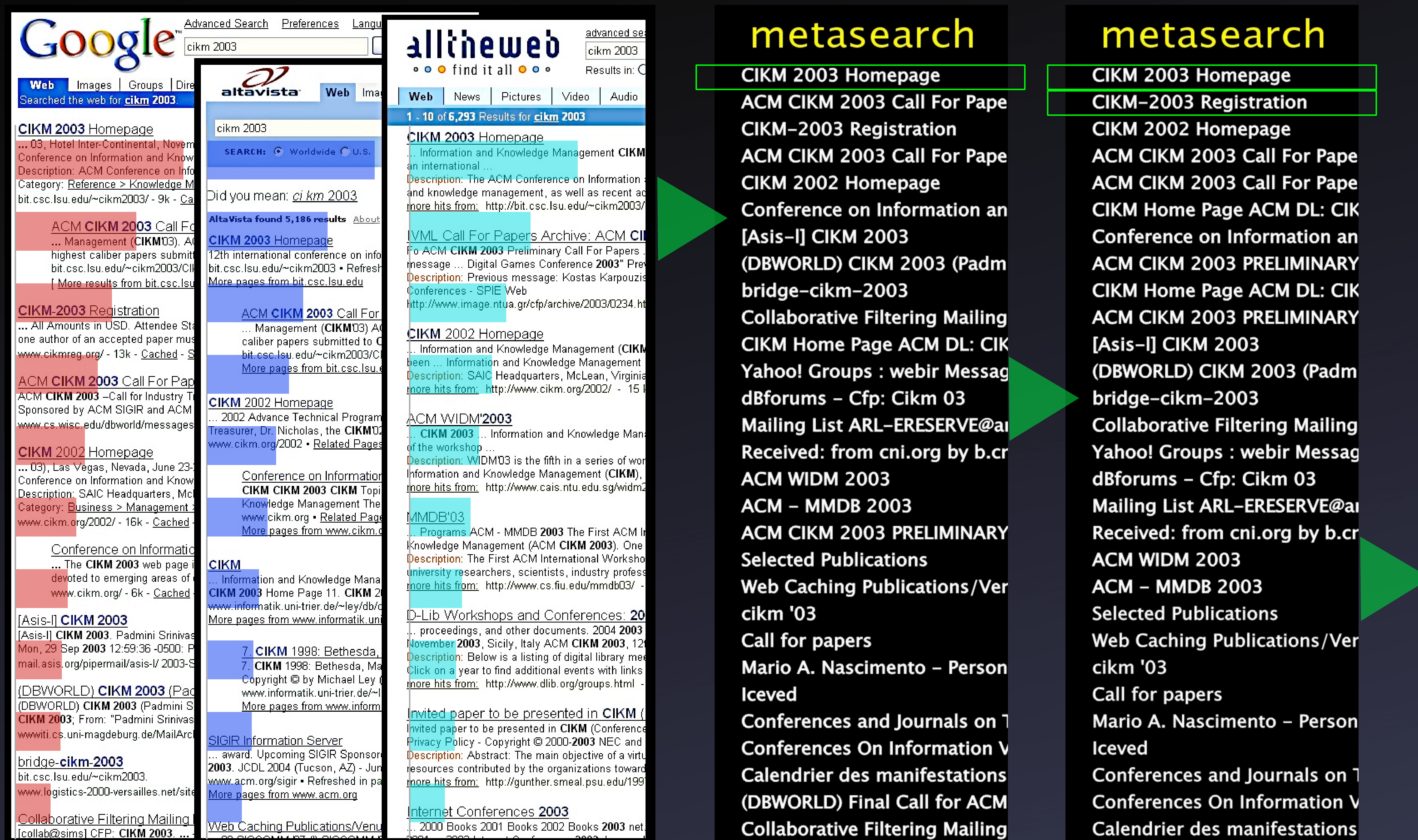
- CIKM 2003 Homepage
- CIKM–2003 Registration
- CIKM 2002 Homepage
- ACM CIKM 2003 Call For Pape
- ACM CIKM 2003 Call For Pape
- CIKM Home Page ACM DL: CIK
- Conference on Information an
- ACM CIKM 2003 PRELIMINARY
- CIKM Home Page ACM DL: CIK
- ACM CIKM 2003 PRELIMINARY
- [Asis–l] CIKM 2003
- (DBWORLD) CIKM 2003 (Padm
- bridge–cikm–2003
- Collaborative Filtering Mailing
- Yahoo! Groups : webir Messag
- dBforums – Cfp: Cikm 03
- Mailing List ARL–ERESERVE@ar
- Received: from cni.org by b.cr
- ACM WIDM 2003
- ACM – MMDB 2003
- Selected Publications
- Web Caching Publications/Ver
- cikm '03
- Call for papers
- Mario A. Nascimento – Person
- Iceved
- Conferences and Journals on T
- Conferences On Information V
- Calendrier des manifestations

Hedge adaptation to IR setup

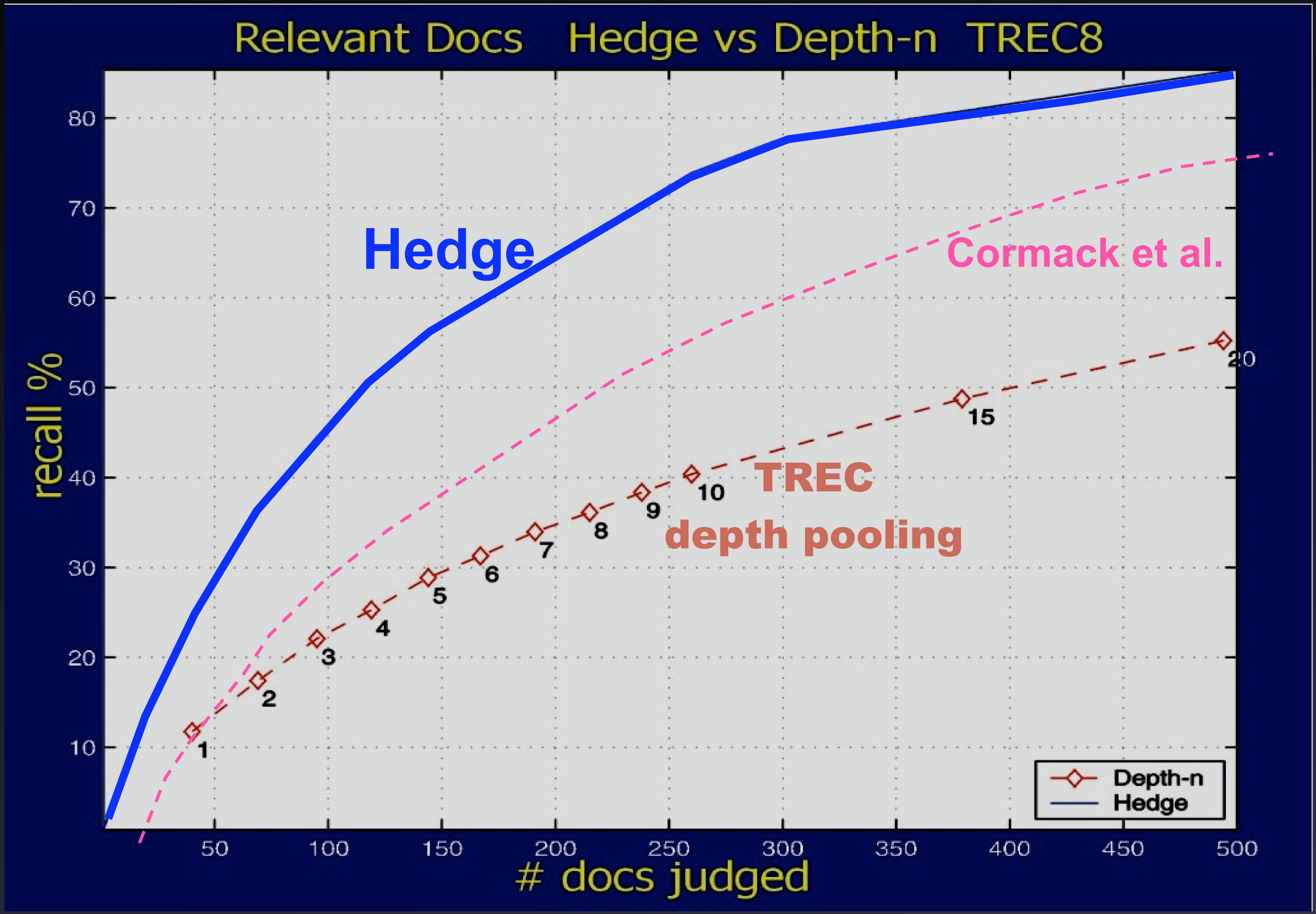


metasearch	metasearch
CIKM 2003 Homepage	CIKM 2003 Homepage
ACM CIKM 2003 Call For Pape	CIKM–2003 Registration
CIKM–2003 Registration	CIKM 2002 Homepage
ACM CIKM 2003 Call For Pape	ACM CIKM 2003 Call For Pape
CIKM 2002 Homepage	ACM CIKM 2003 Call For Pape
Conference on Information an	CIKM Home Page ACM DL: CIK
[Asis–l] CIKM 2003	Conference on Information an
(DBWORLD) CIKM 2003 (Padm	ACM CIKM 2003 PRELIMINARY
bridge–cikm–2003	CIKM Home Page ACM DL: CIK
Collaborative Filtering Mailing	ACM CIKM 2003 PRELIMINARY
CIKM Home Page ACM DL: CIK	[Asis–l] CIKM 2003
Yahoo! Groups : webir Messag	(DBWORLD) CIKM 2003 (Padm
dBforums – Cfp: Cikm 03	bridge–cikm–2003
Mailing List ARL–ERESERVE@ai	Collaborative Filtering Mailing
Received: from cni.org by b.cr	Yahoo! Groups : webir Messag
ACM WIDM 2003	dBforums – Cfp: Cikm 03
ACM – MMDB 2003	Mailing List ARL–ERESERVE@ai
ACM CIKM 2003 PRELIMINARY	Received: from cni.org by b.cr
Selected Publications	ACM WIDM 2003
Web Caching Publications/Ver	ACM – MMDB 2003
cikm '03	Selected Publications
Call for papers	Web Caching Publications/Ver
Mario A. Nascimento – Person	cikm '03
Iceved	Call for papers
Conferences and Journals on T	Mario A. Nascimento – Person
Conferences On Information V	Iceved
Calendrier des manifestations	Conferences and Journals on T
(DBWORLD) Final Call for ACM	Conferences On Information V
Collaborative Filtering Mailing	Calendrier des manifestations

Hedge adaptation to IR setup

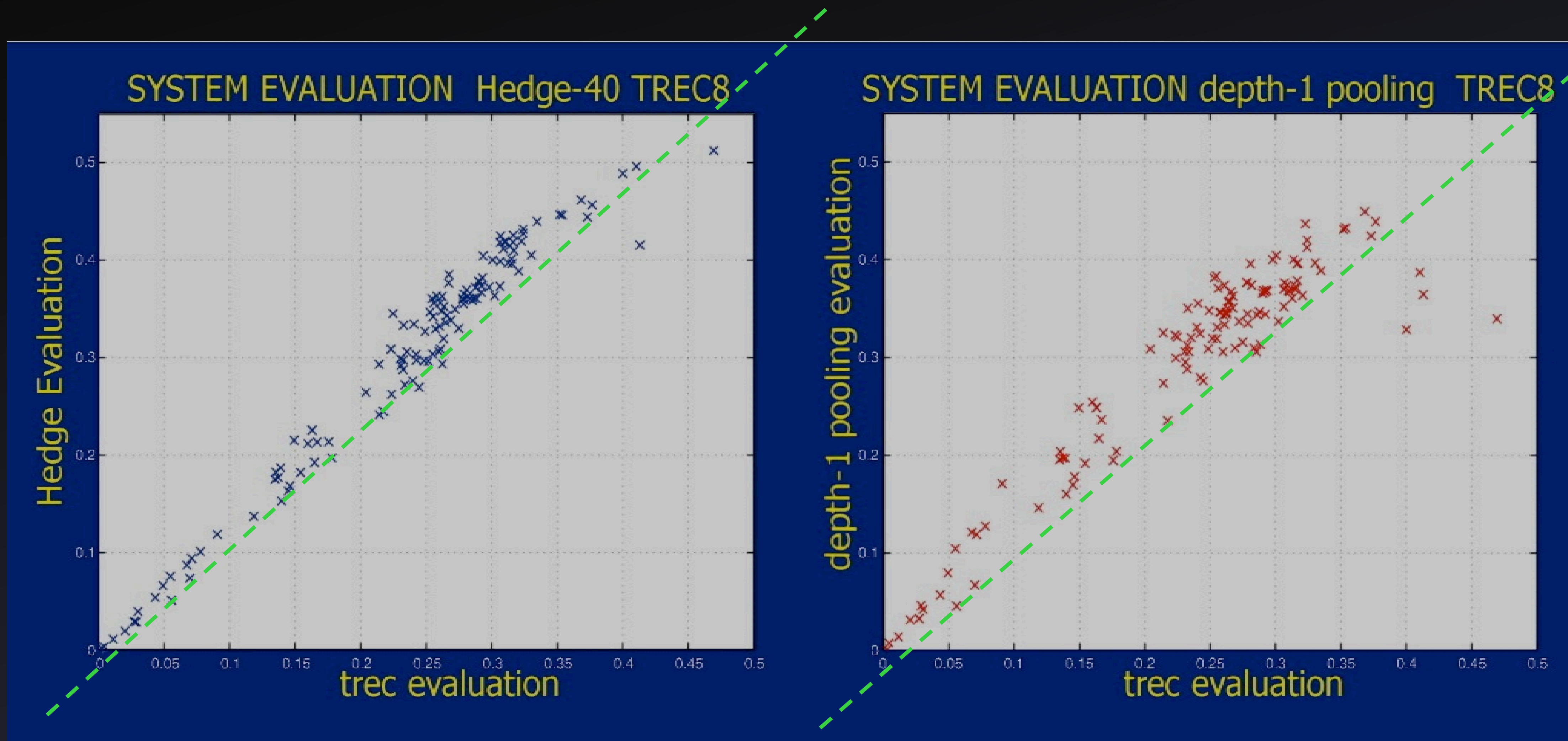


Hedge results - recall rate



24

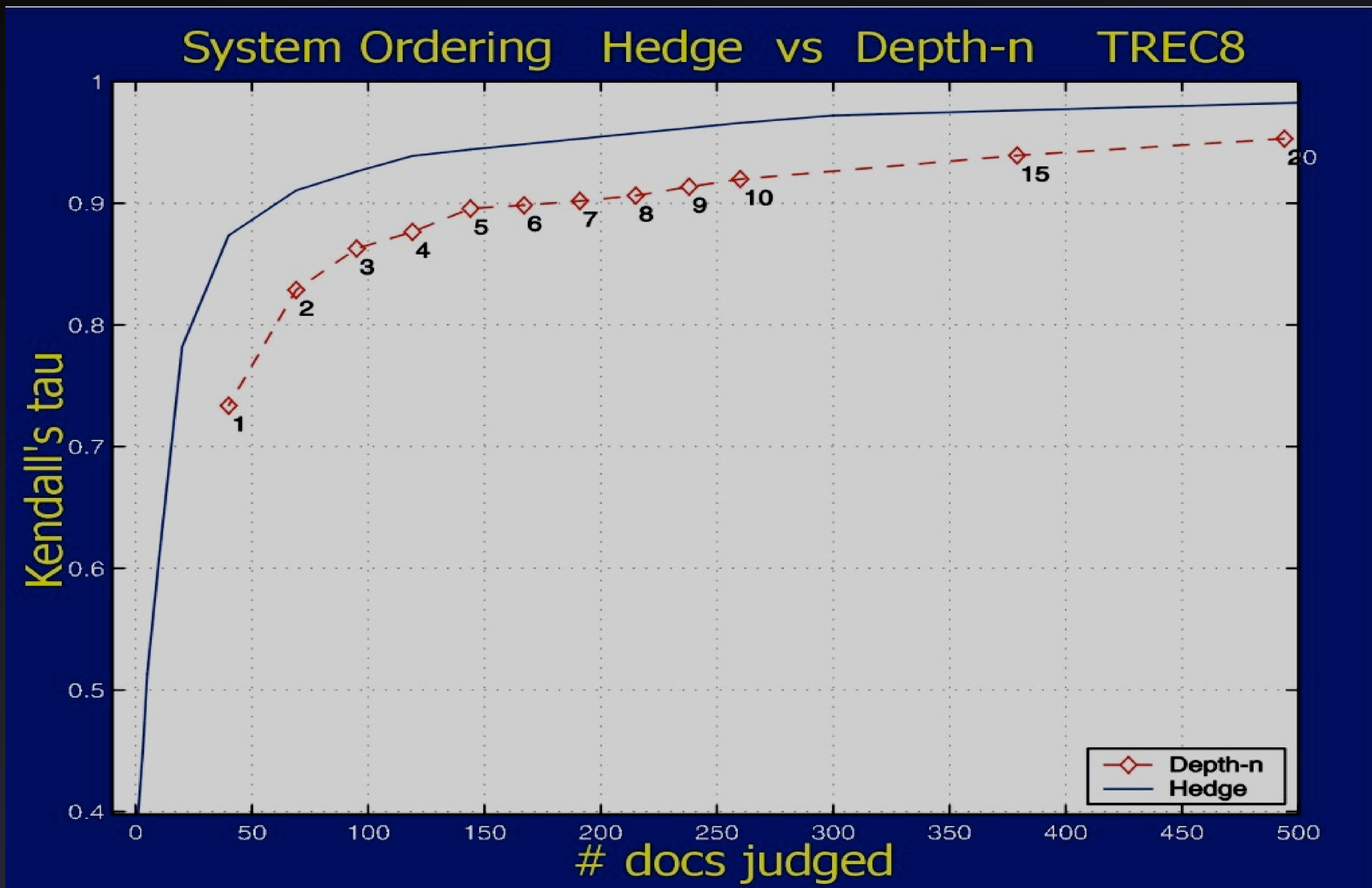
Hedge results - evaluation



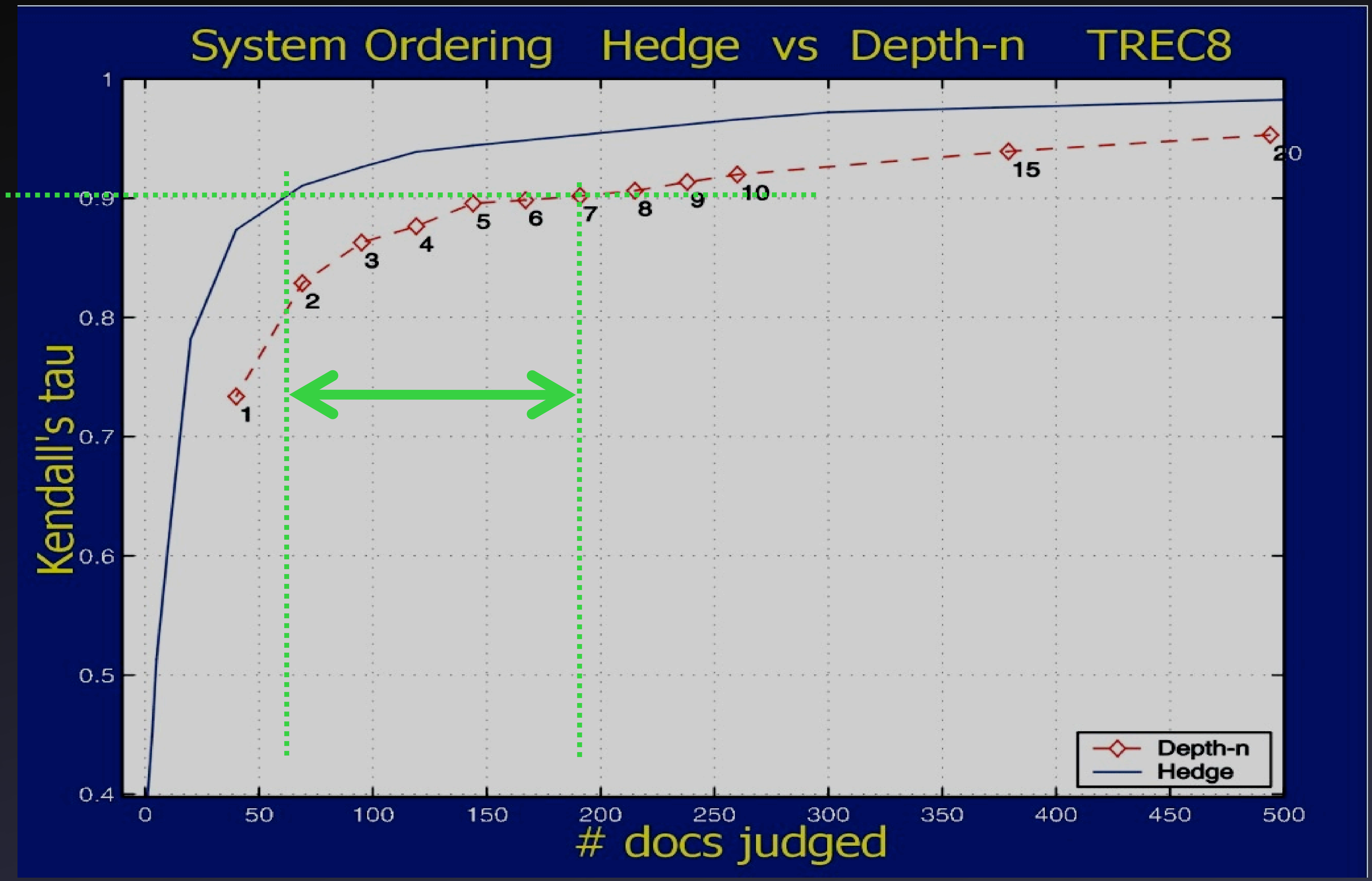
Hedge results - evaluation

26

Hedge Evaluation - Kendall's τ



Hedge Evaluation - Kendall's τ



Hedge results - metasearch

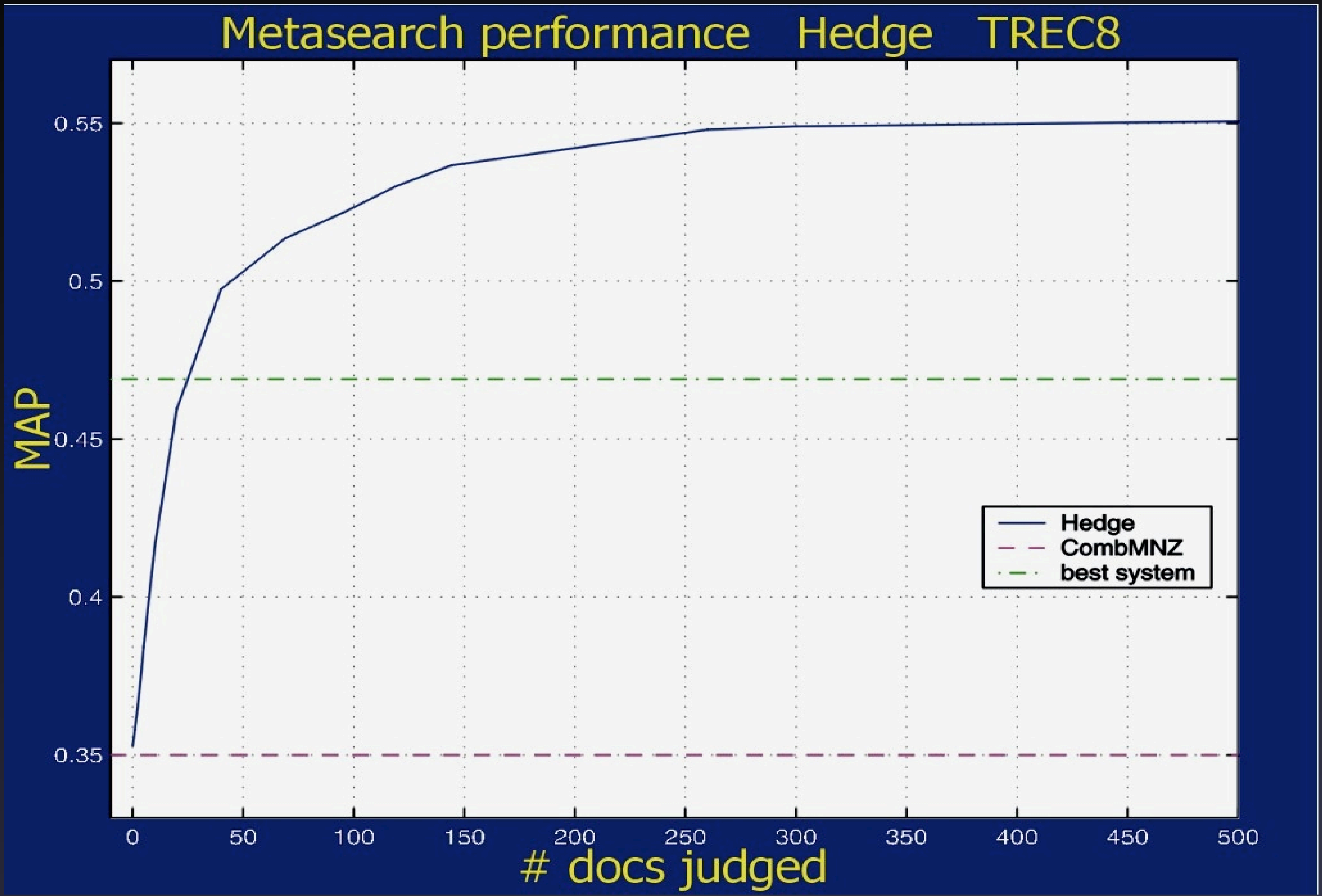
no relevant judgments
(uniform system weights)

TREC	MNZ	COND	Hedge-0	$\%MNZ$	$\%COND$
3	0.423	0.403	0.418	-1.2	$+3.7$
5	0.294	0.307	0.309	$+5.1$	$+0.6$
6	0.341	0.315	0.345	$+1.2$	$+9.5$
7	0.320	0.308	0.323	$+0.9$	$+4.9$
8	0.350	0.343	0.352	$+1.4$	$+2.6$

MNZ=CombMNZ(Fox,Shaw,Lee et al)
COND=Condorcet(Aslam,Montague)

28

Hedge results - metasearch



29

overview

- Introduction
- Relevance Prior
- Hedge
- Sampling
- Future work

SAMPLING TRICKS:

- non -uniform
- without replacement
- ratio estimator

30

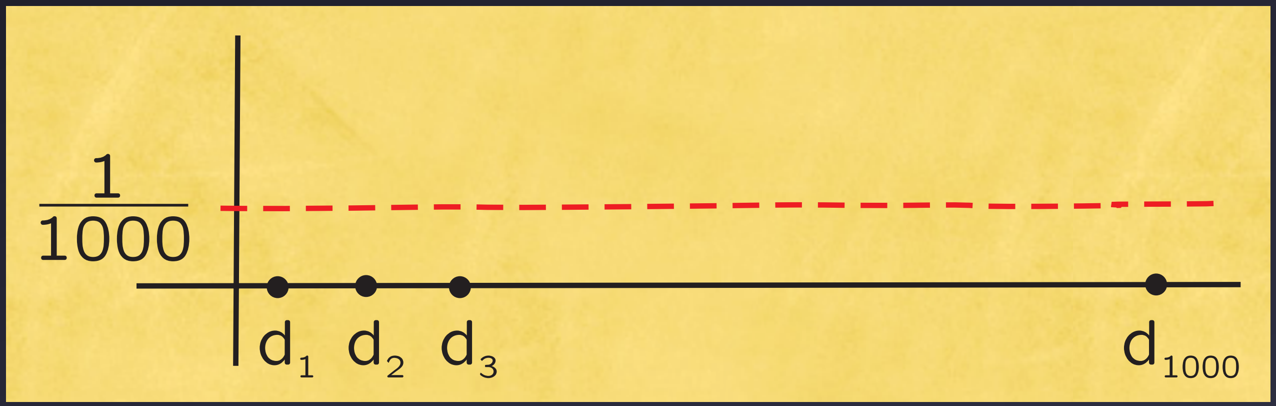
sampling example

- Say I have 1000 animals
- I want to find percentage of sick animals
- Obvious solution:
  - examine all 1000
  - return #sick/1000

31

sampling example

- Alternate solution:
  - uniformly sample animals
  - examine the sampled ones
  - return #sick-seen/#samples
- Distribution: uniform over 1000
  - $p_i = 1/1000$



- Random variable: X = sick
  - 1 if sick, 0 if not

theory

- LLN: Avg of random sample converges to mean

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \rightarrow E[X]$$

- CLT: But how fast ?
  - Average of i.i.d. r. v. rapidly becomes Gaussian
  - Mean is preserved
  - Variance decreases linearly in n
  - SD decreases by root n

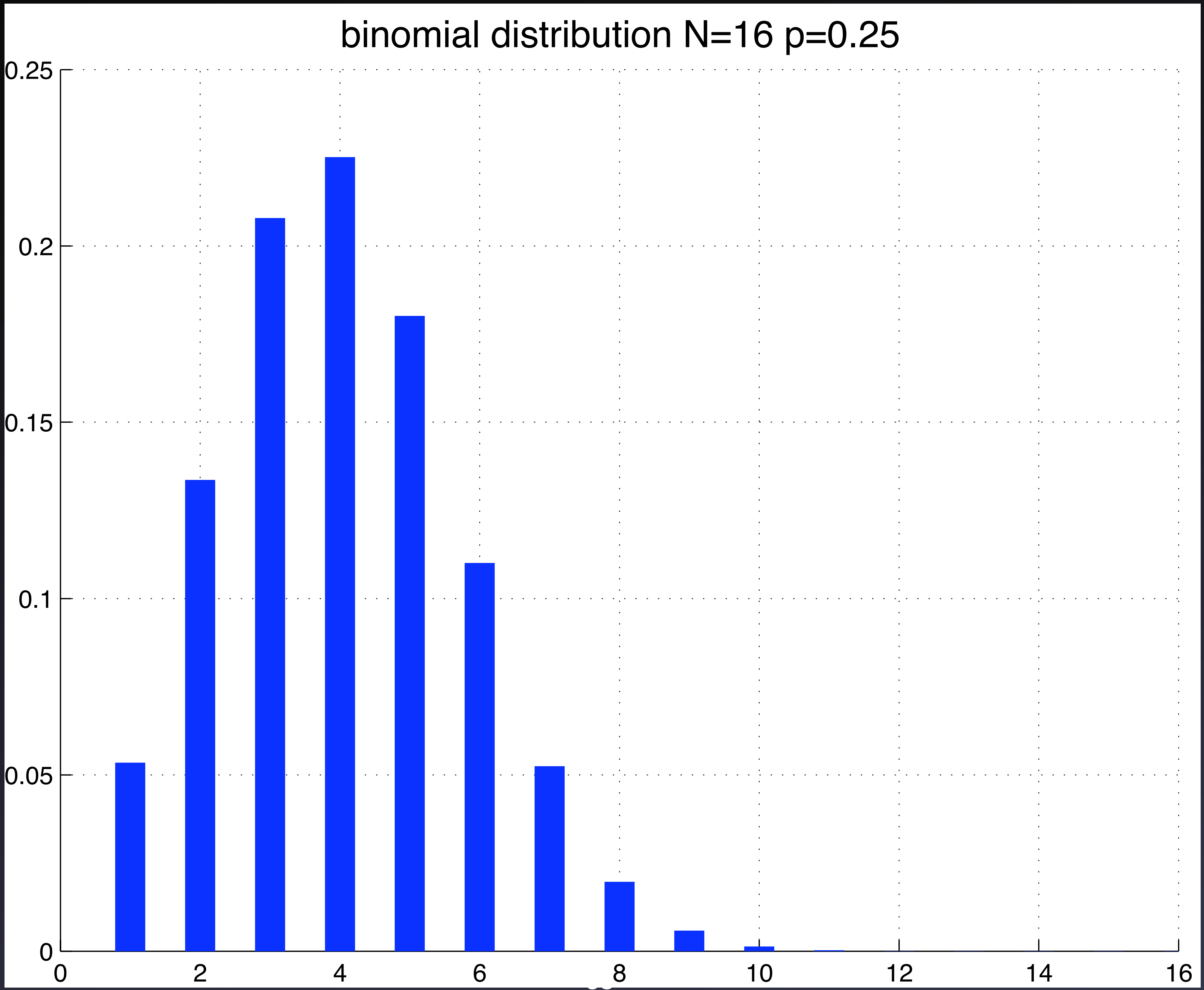
FAST

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \rightarrow \quad N(\mu, \sigma/\sqrt{n})$$

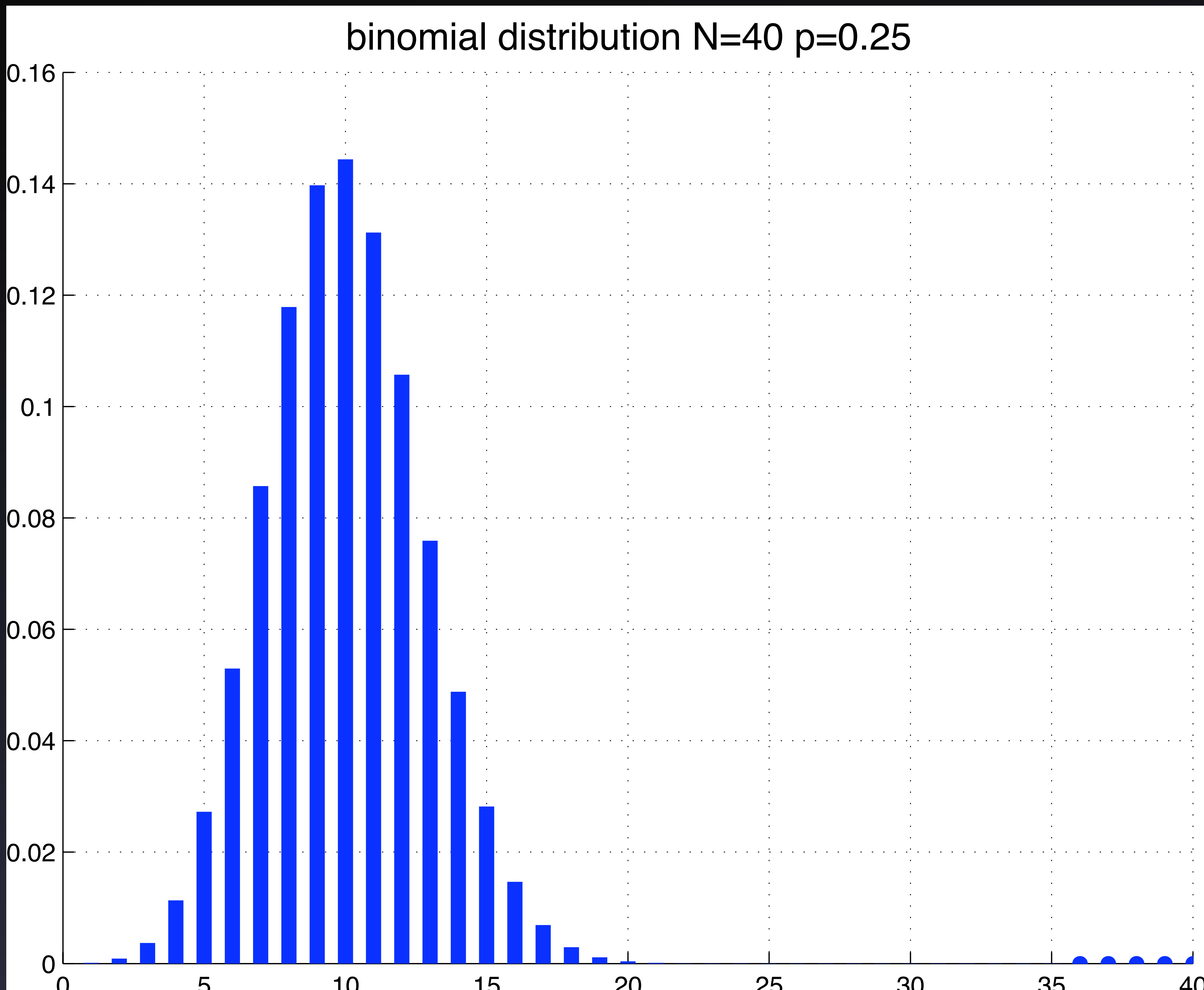
estimator accuracy

- Suppose that, unknown to us, Percentage_sick = p = 0.25
  - i.e., there were 250 sick animals
- How many samples until we could estimate that accurately, say +/- 0.03?
- LLN & CLT

LLN and CLT N=16

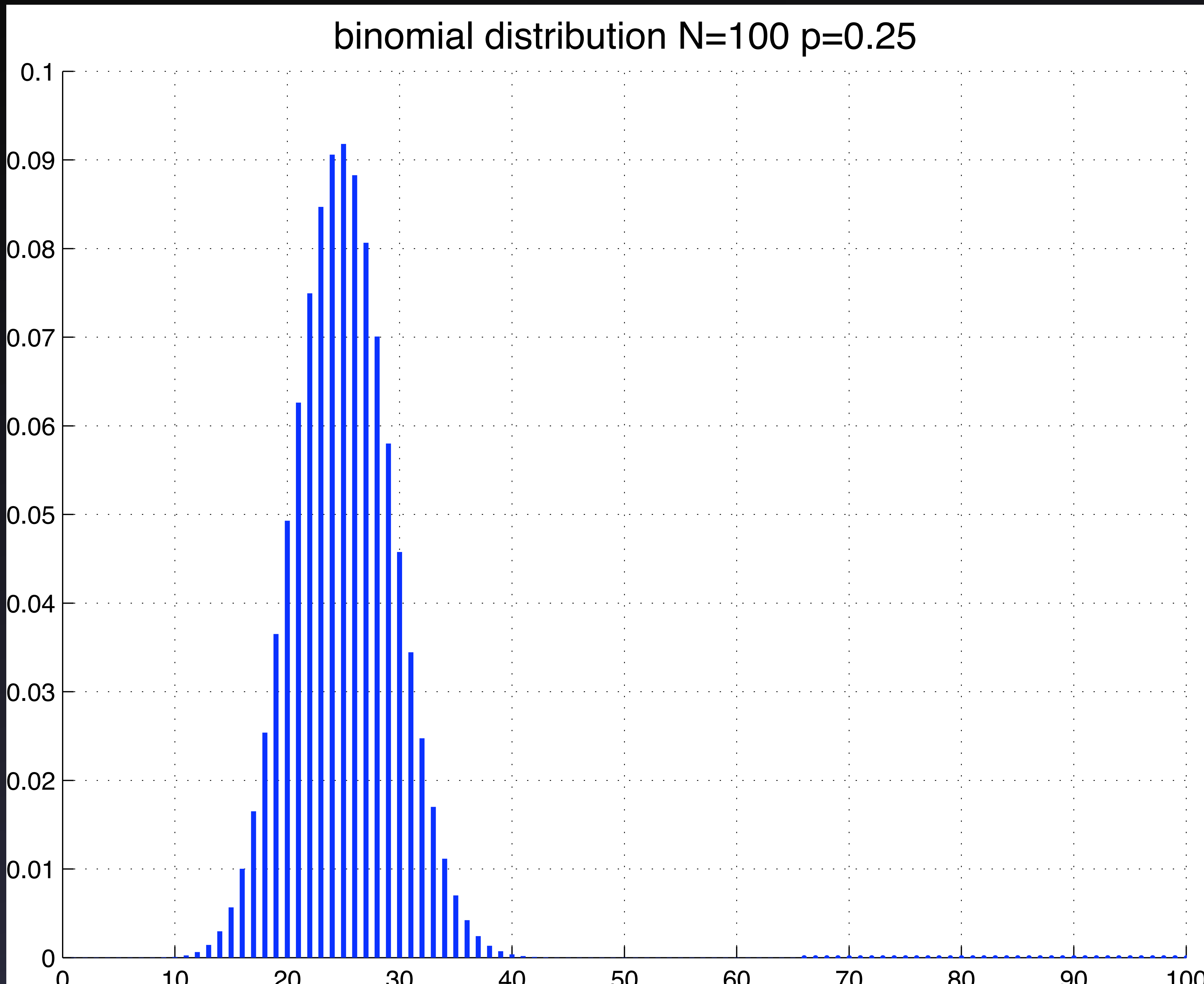


LLN and CLT N=40

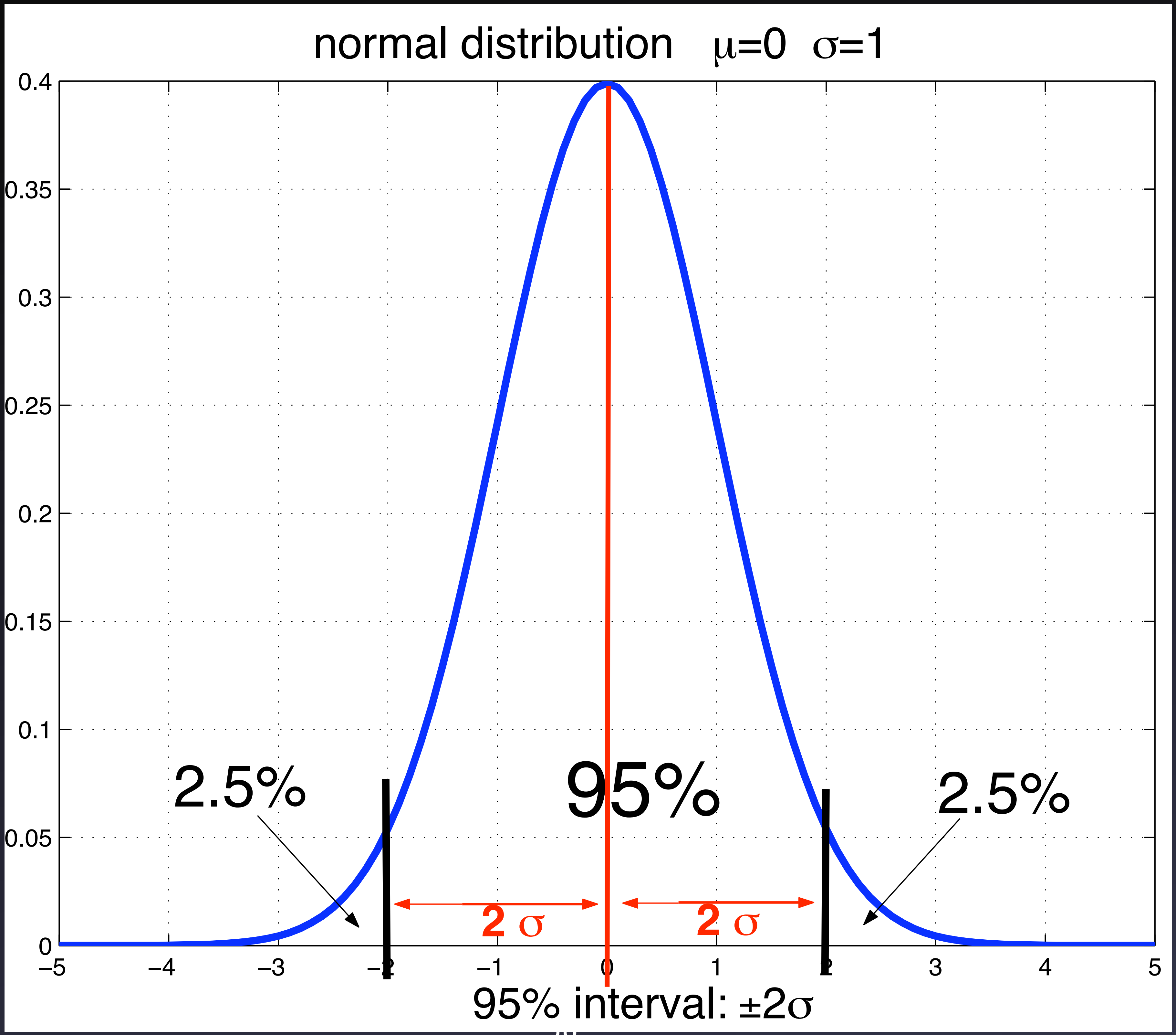


36

LLN and CLT N=100



Normal Distribution 95% CI



38

uniform sample size

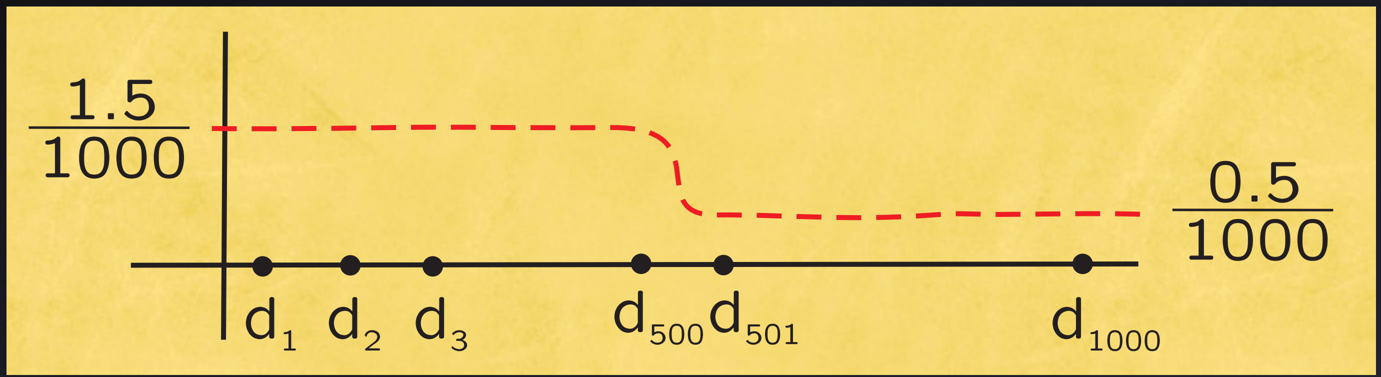
$$Var[X] = p(1-p) = 3/16 \qquad 2\sigma = 2\sqrt{p(1-p)/n} = 0.03$$

$$Var[\overline{X}] = p(1-p)/n \qquad n = 4p(1-p)/0.03^2 \approx 833$$

- That's a lot of samples...
- Is there a "smarter" sampling strategy?

importance sampling

- Sample "more" where sick animals are
  - for example categorize/order them by age:
    - 1-500 old;   501-1000 young



$$p_i = \begin{cases} 1.5/1000 & i \leq 500 \\ 0.5/1000 & i > 500 \end{cases}$$

- How to correct for estimated mean ?
  - scaling factors

$$x_i = \begin{cases} sick(i) \cdot 2/3 & i \leq 500 \\ sick(i) \cdot 2 & i > 500 \end{cases}$$

2 extreme cases

all sick in top half

$$
\begin{aligned}
Var[X] &= E[X^2] - E^2[X] \\
&= \sum_{i=1}^{1000} p_i \cdot {x_i}^2 - (1/4)^2 \\
&= \sum_{i=1}^{500} p_i \cdot {x_i}^2 + \sum_{i=501}^{1000} p_i \cdot {x_i}^2 - 1/16 \\
&= 250 \cdot \frac{3/2}{1000} \cdot (2/3)^2 + 0 - 1/16 \\
&= 1/6 - 1/16 \\
&= 5/48 \\
&\approx 0.1042
\end{aligned}
$$

sample size = 463

all sick in bottom half

$$
\begin{aligned}
Var[X] &= E[X^2] - E^2[X] \\
&= \sum_{i=1}^{1000} p_i \cdot {x_i}^2 - (1/4)^2 \\
&= \sum_{i=1}^{500} p_i \cdot {x_i}^2 + \sum_{i=501}^{1000} p_i \cdot {x_i}^2 - 1/16 \\
&= 0 + 250 \cdot \frac{1/2}{1000} \cdot 2^2 + 0 - 1/16 \\
&= 1/2 - 1/16 \\
&= 7/16 \\
&\approx 0.4375
\end{aligned}
$$

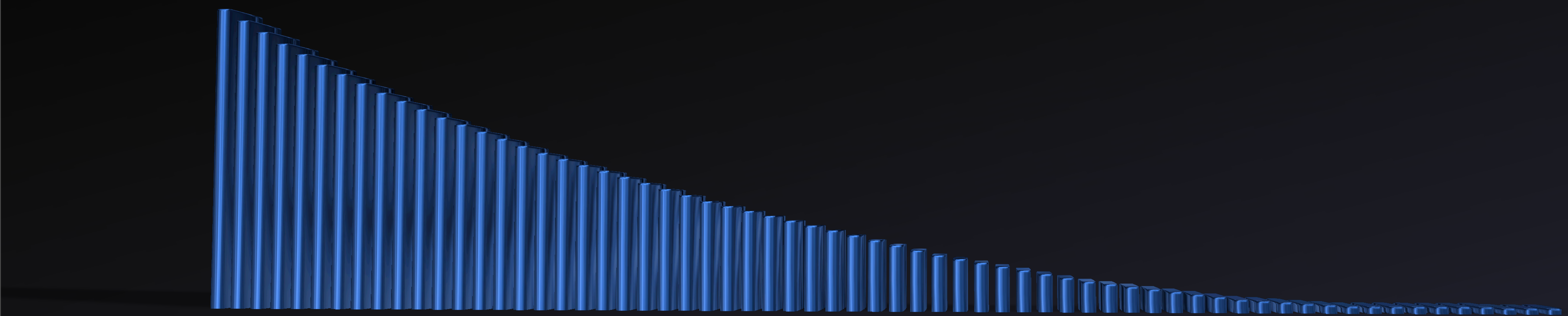
sample size = 1944

- Comparison with uniform sampling:
  - var = 1.87;   sample size required= 833

sampling and evaluation

- non-uniform
  - pps is ideal
  - we are going to use the prior (avg over systems)
- without replacement
  - π_k = inclusion probabilities must be computed
  - stratified sampling
- use a ratio estimator $$\hat{AP} = \frac{\sum_{k \in S} p_k / \pi_k}{\sum_{k \in S} 1/\pi_k}$$
- prior, sampling and estimation independent

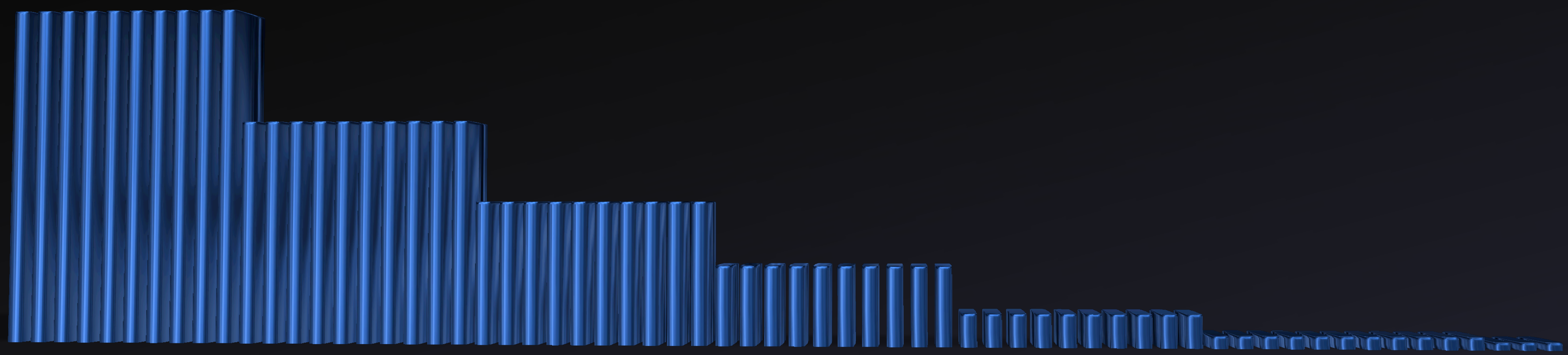
stratified sampling



- non-uniform distribution; goal sample size = m

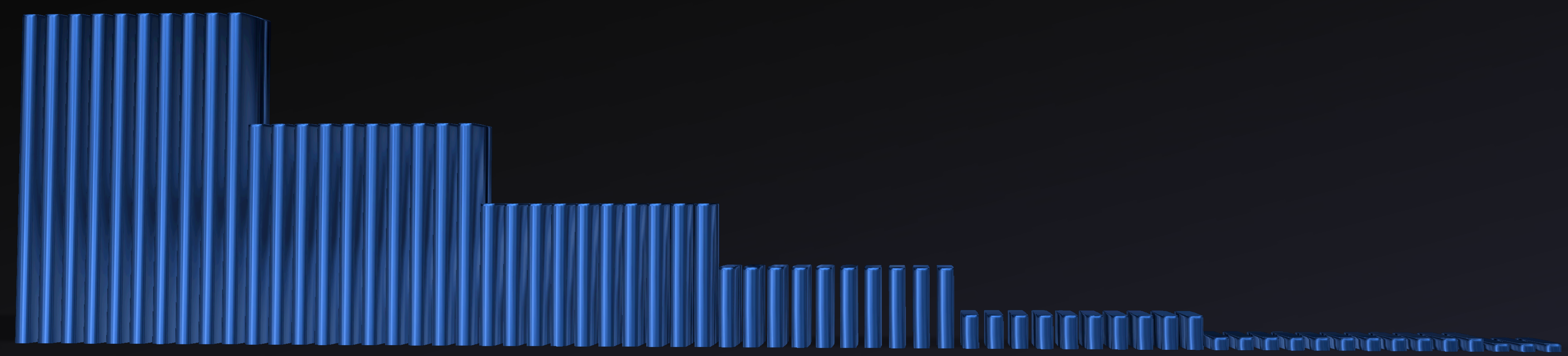
43

stratified sampling



- non-uniform distribution; goal sample size = m

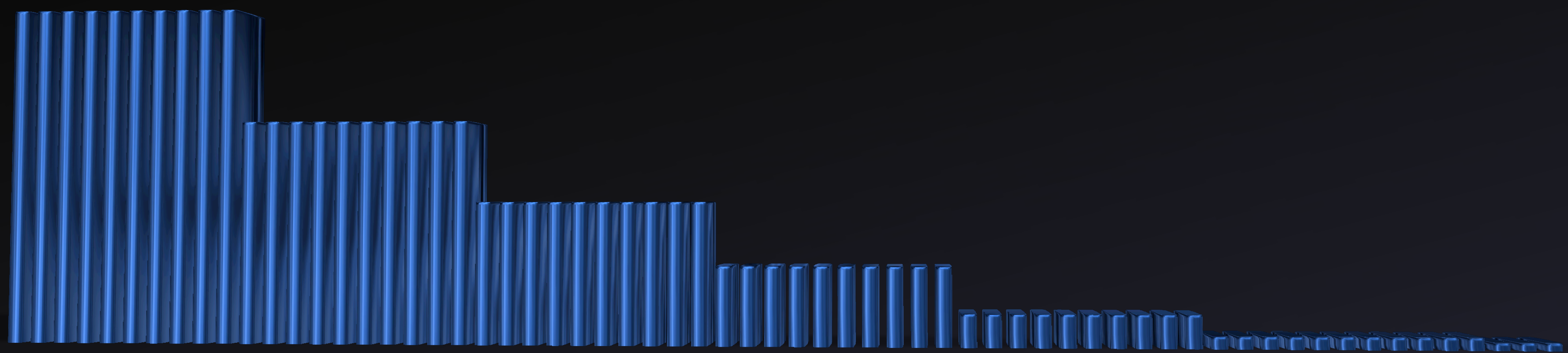
stratified sampling



- non-uniform distribution; goal sample size = m
- partition docs in buckets of size m each

43

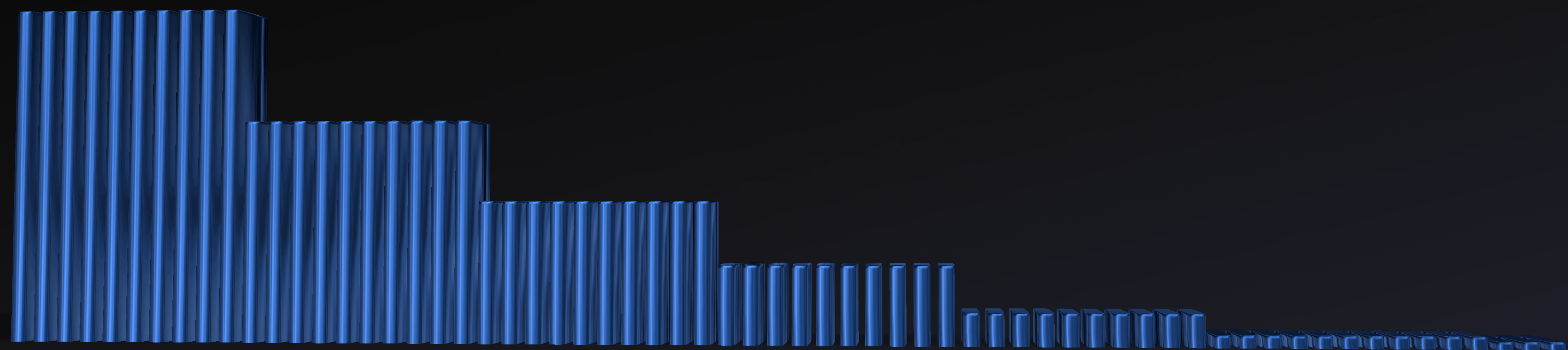
stratified sampling



- non-uniform distribution; goal sample size = m
- partition docs in buckets of size m each
- sample the buckets with replacement m times
  - based on the cumulative weight for each bucket

43

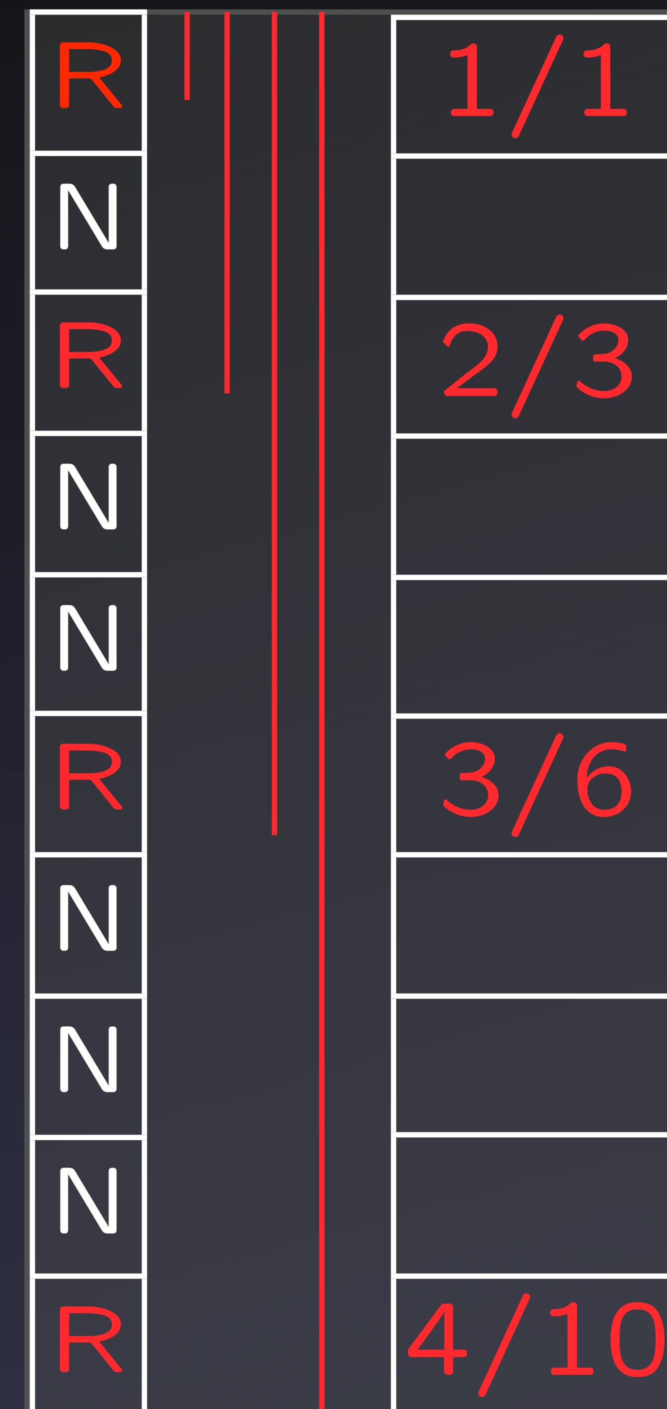
stratified sampling



- non-uniform distribution; goal sample size = m
- partition docs in buckets of size m each
- sample the buckets with replacement m times
  - based on the cumulative weight for each bucket
- for each bucket, if picked k times, sample uniform without replacement k docs in it

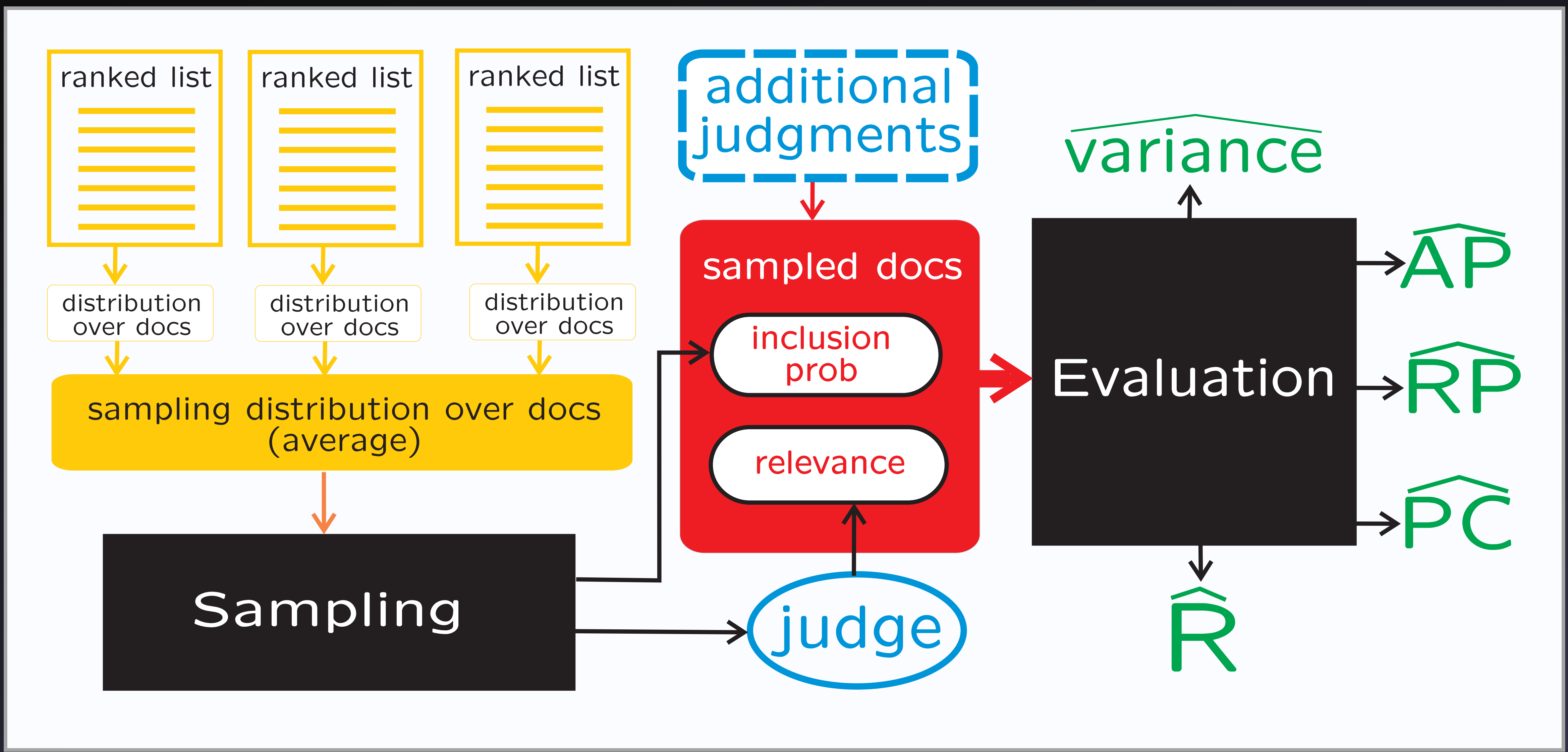
sampling for AP estimation

- AP is the average of precisions at relevant ranks
  - population is precision values at relevant ranks
  - those need to be estimated too
- Estimate performance on other systems
- We can also estimate other measures/quantities
- We can use additional judgments

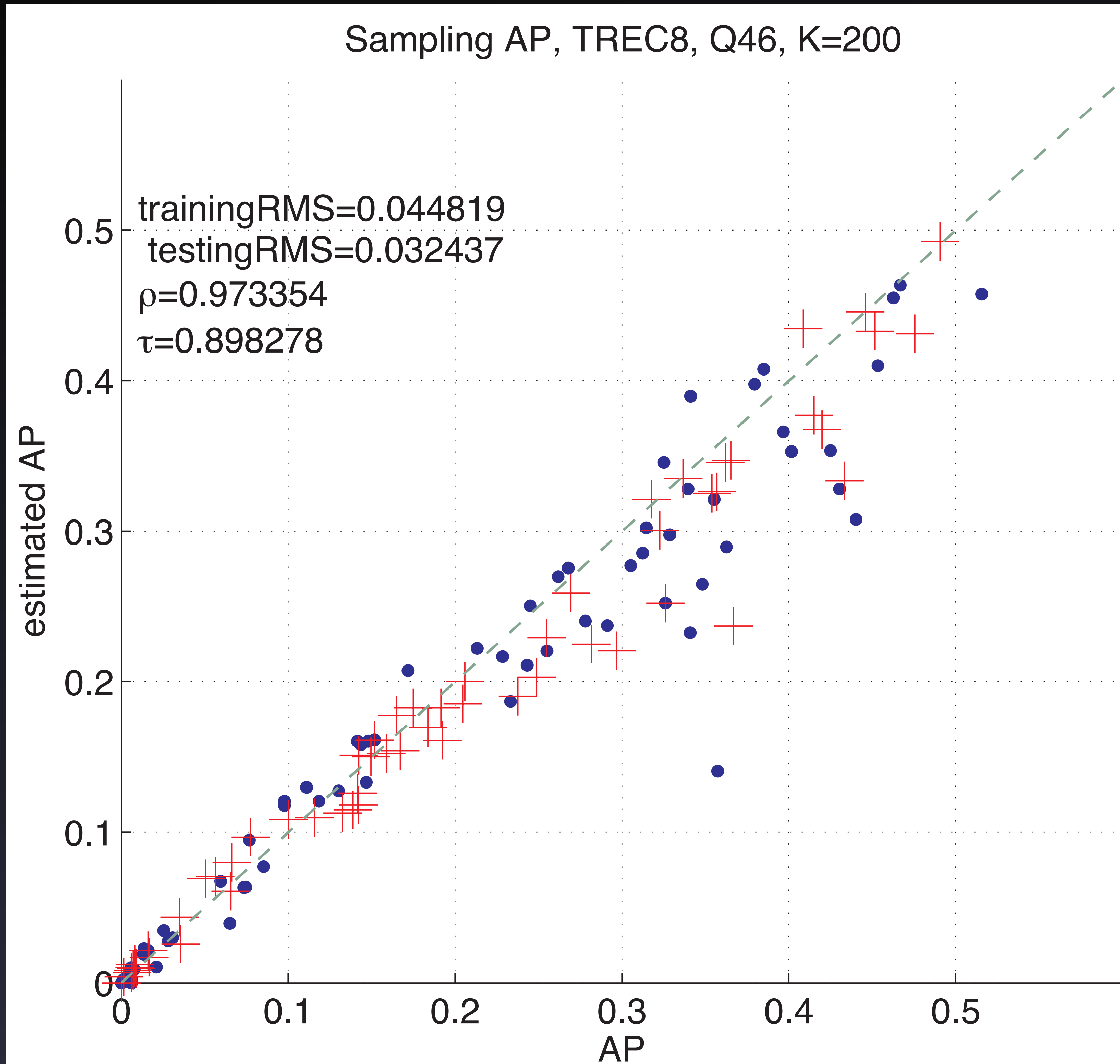


44

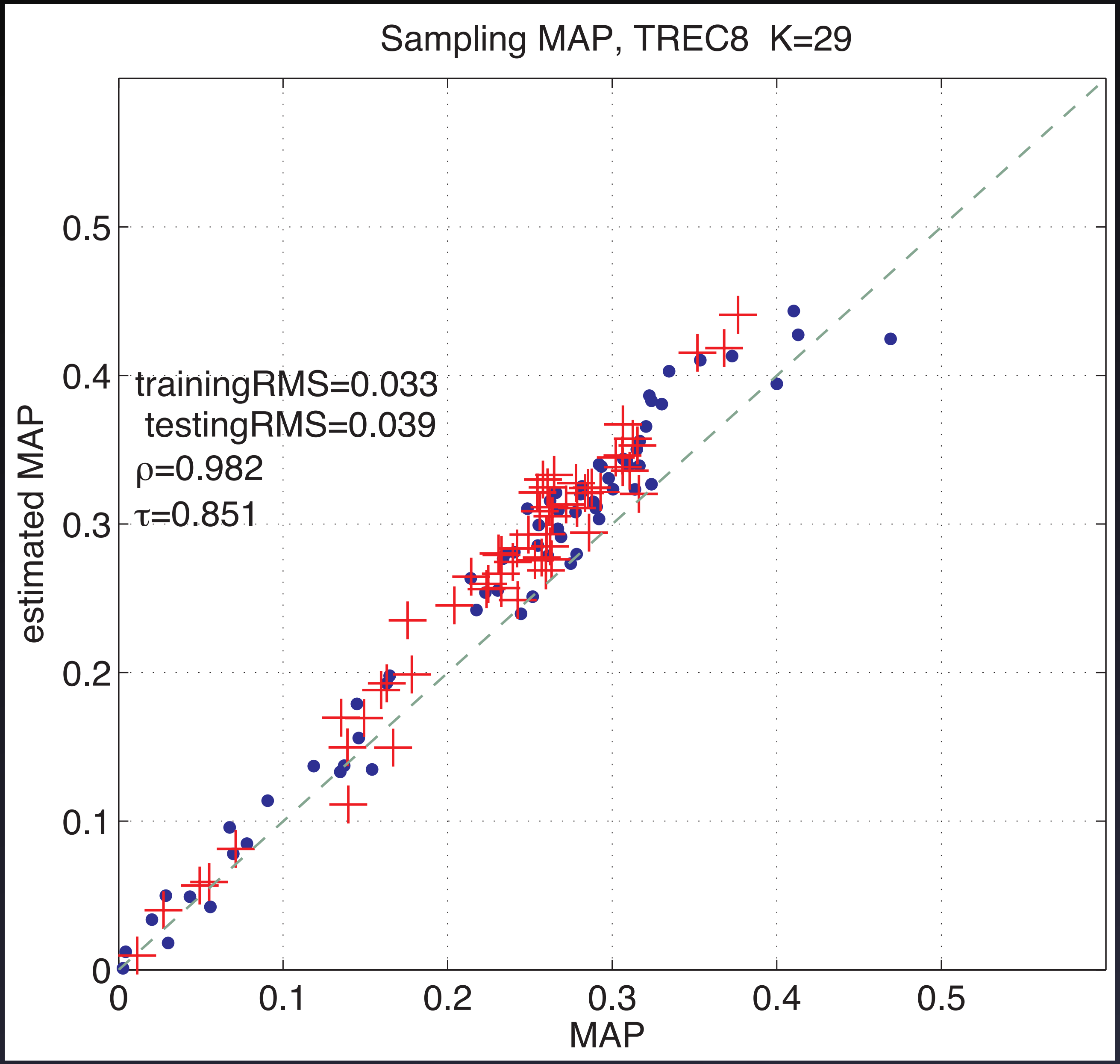
sampling for AP estimation

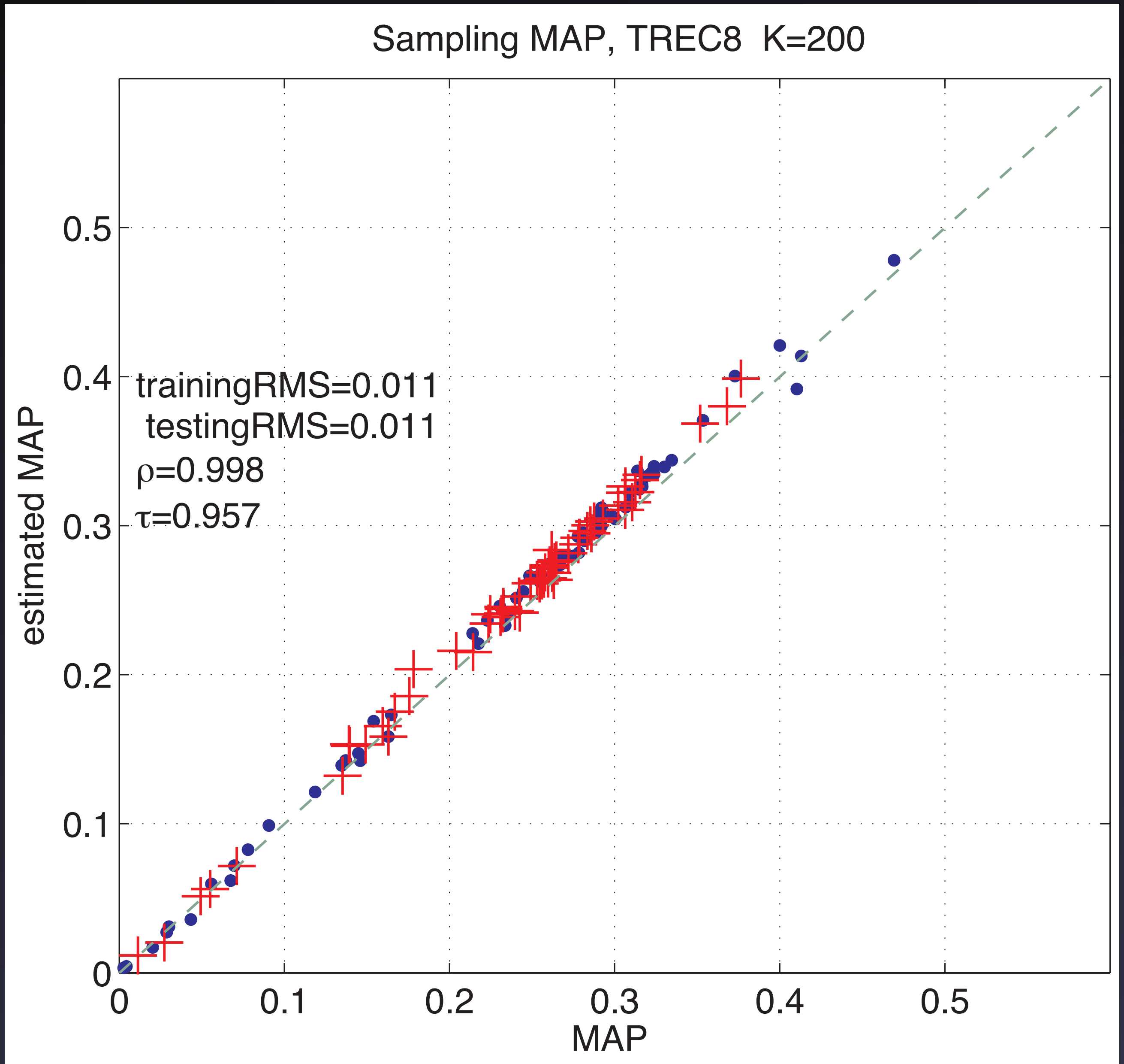


sampling results: one query

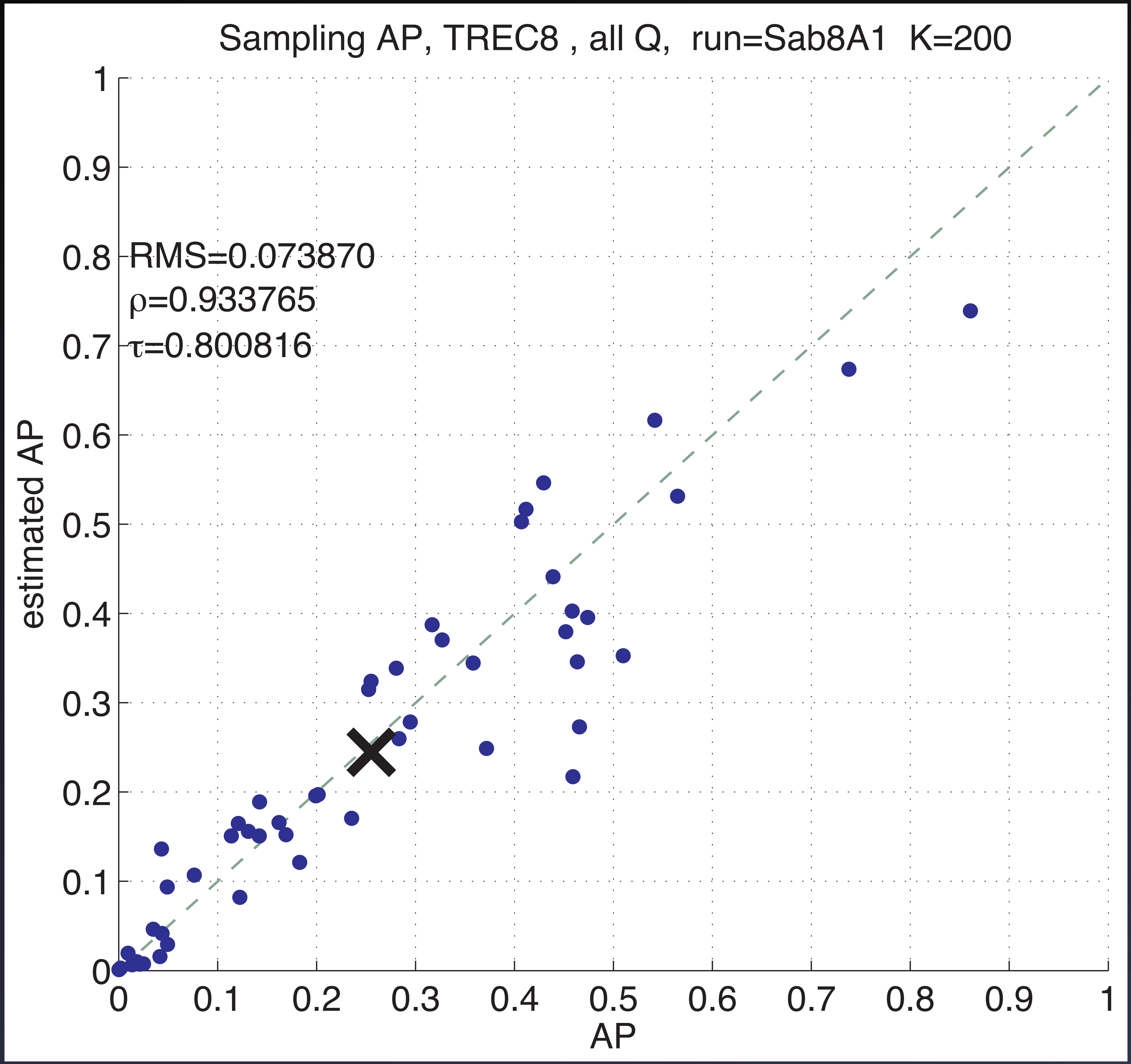


sampling results: all queries

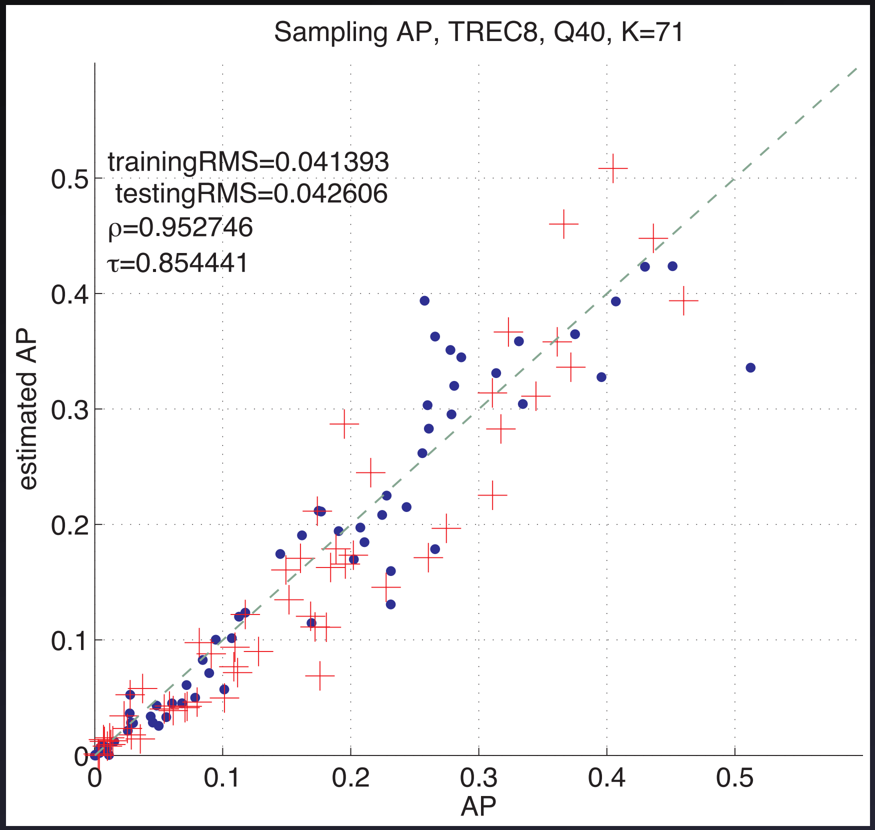


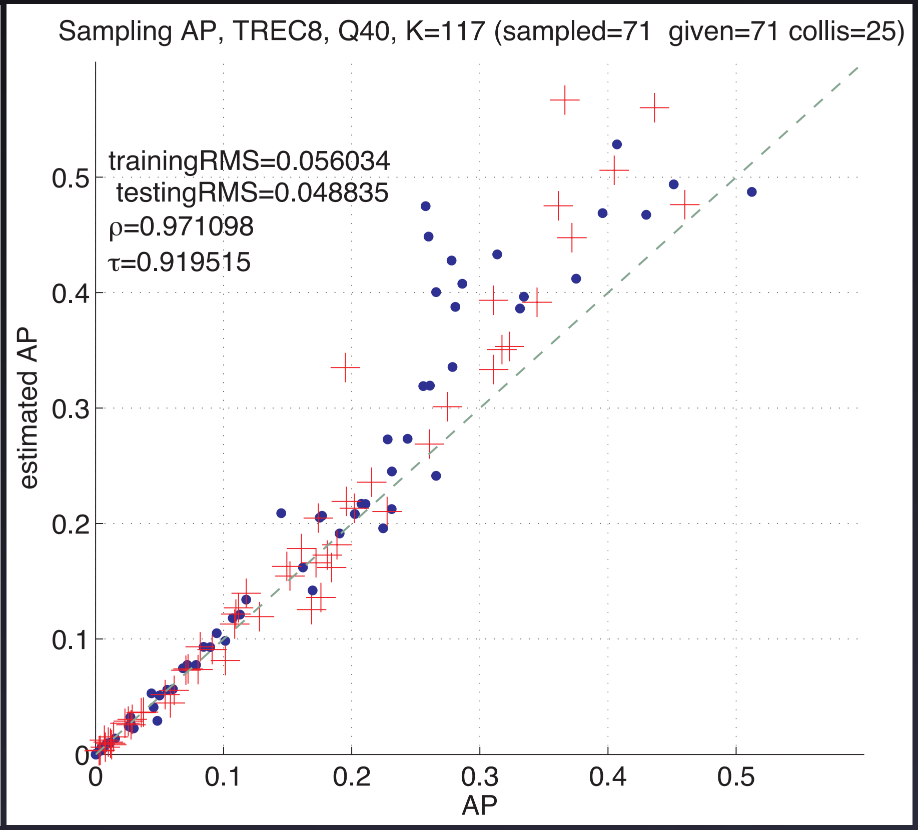


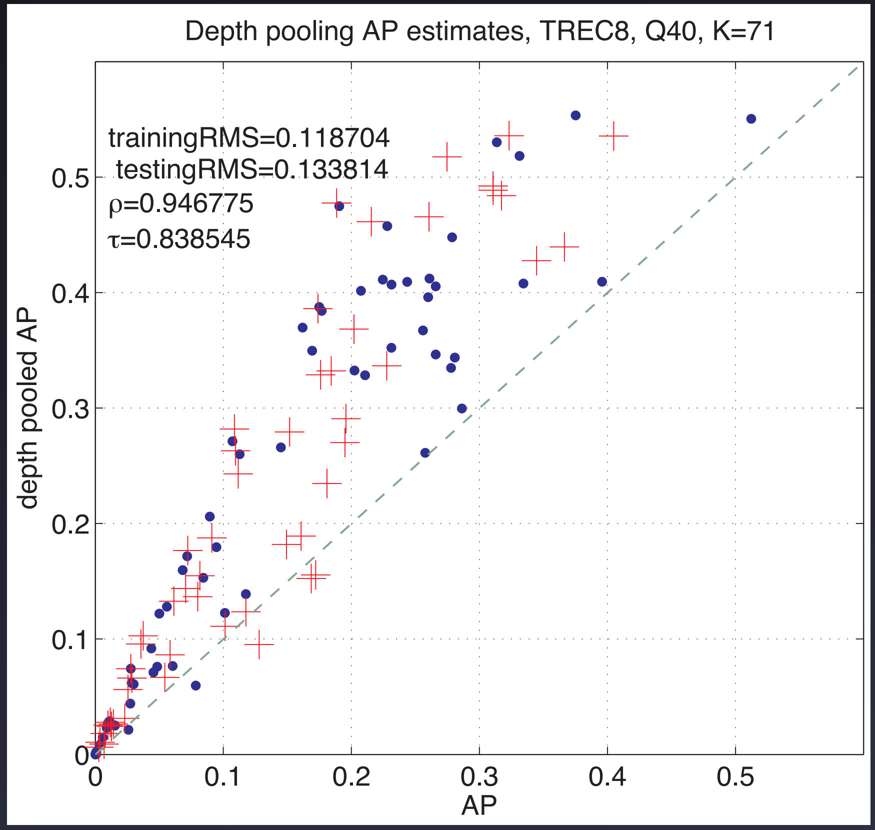
sampling results: one system



add deterministic judgments

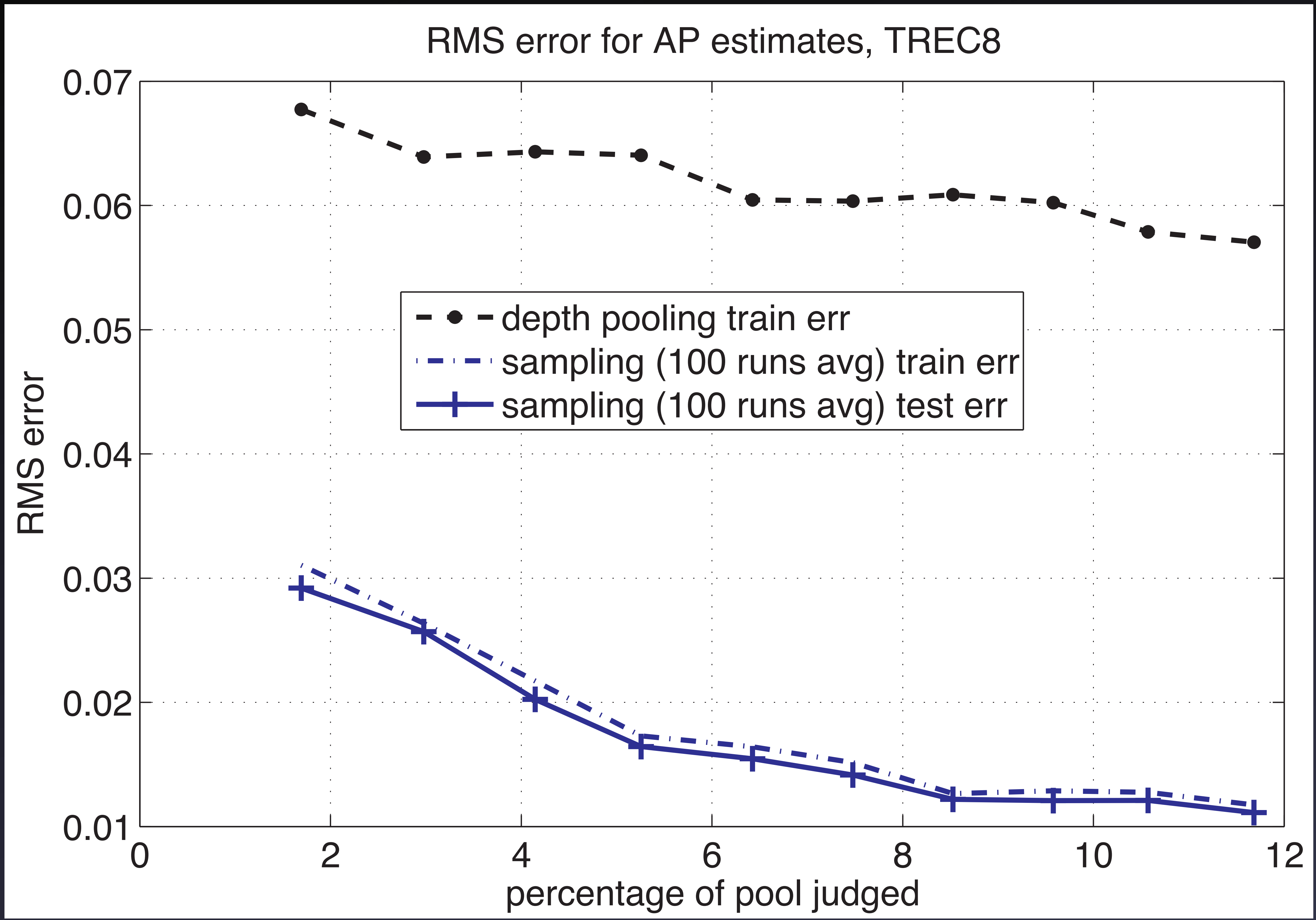




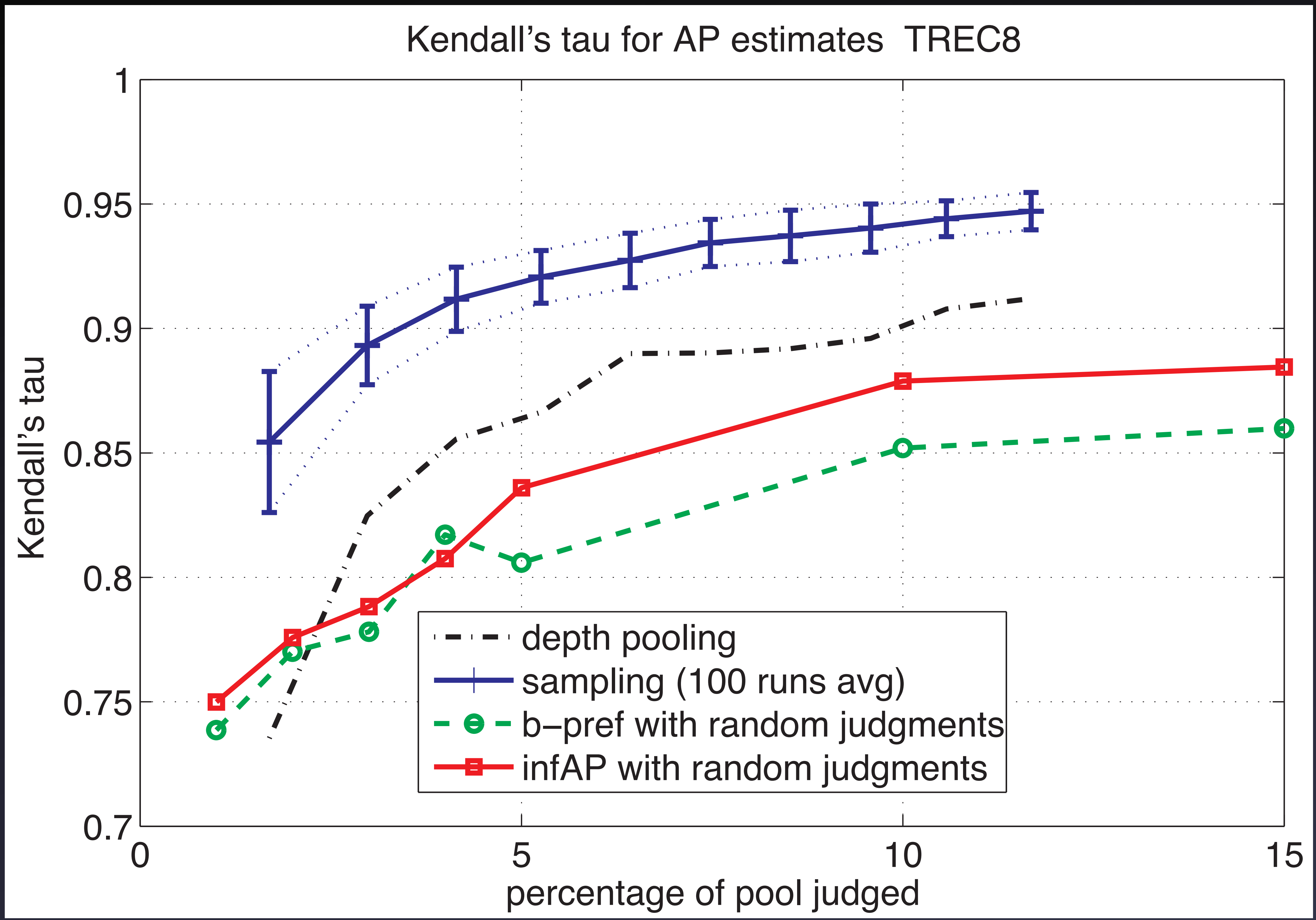


49

sampling results : trends(RMS)



sampling results : trends(τ)



overview

- Introduction
- Relevance Prior
- Hedge
- Sampling
- Future work

future work: Hedge optimization

- Bound known to be optimal for that class of algorithms
  - but performance much better than the bound

$$\textsc{Loss}_{Hedge} \leq \frac{\min_b\{L_b\}\cdot\ln(1/\beta)+\ln(N)}{1-\beta}$$

- Use a variable β
  - can we get a better bound ?
- Experiment on machine learning datasets

future work: sampling bias

- the estimator used is not unbiased
  - but bias very small as sample grows

$$\hat{AP} = \frac{\sum_{k \in S} p_k / \pi_k}{\sum_{k \in S} 1/\pi_k}$$

- Idea: for small sample sizes use a correction

$$\widehat{AP} = \frac{\widehat{SP}}{\widehat{R}} \cdot \frac{1}{1 + \frac{\sigma_R^2}{\widehat{R}^2}}$$

future work: sampling variance

- pair-inclusion probabilities can be computed

$$\pi_{df} = \begin{cases} \pi_d \cdot \pi_f & if \quad \pi_d = \pi_f \\ \frac{|S|-1}{|S|}\pi_d \cdot \pi_f & if \quad \pi_d \neq \pi_f \end{cases}$$

- the variance of the estimator

$$\widehat{var}(\widehat{AP}) = \frac{1}{\widehat{R}^2}\left(\sum_{d \in S, rel} \frac{1-\pi_d}{\pi_d^2} p_d^2 - \frac{1}{|S|-1} \sum_{d,f \in S, rel} \frac{p_d \cdot y_f}{\pi_d \cdot \pi_f}\right)$$

- derive confidence intervals

assuming we can fix the bias

future work: sampling + active learning

- experiment with mixing the samples
  - we know how to mix a random sample with a deterministic one
  - try to combine two samples obtained random
- try an active learning strategy
  - mixed within sampling process

future work: sampling + active learning

- experiment with mixing the samples
  - we know how to mix a random sample with a deterministic one
  - try to combine two samples obtained random
- try an active learning strategy
  - mixed within sampling process

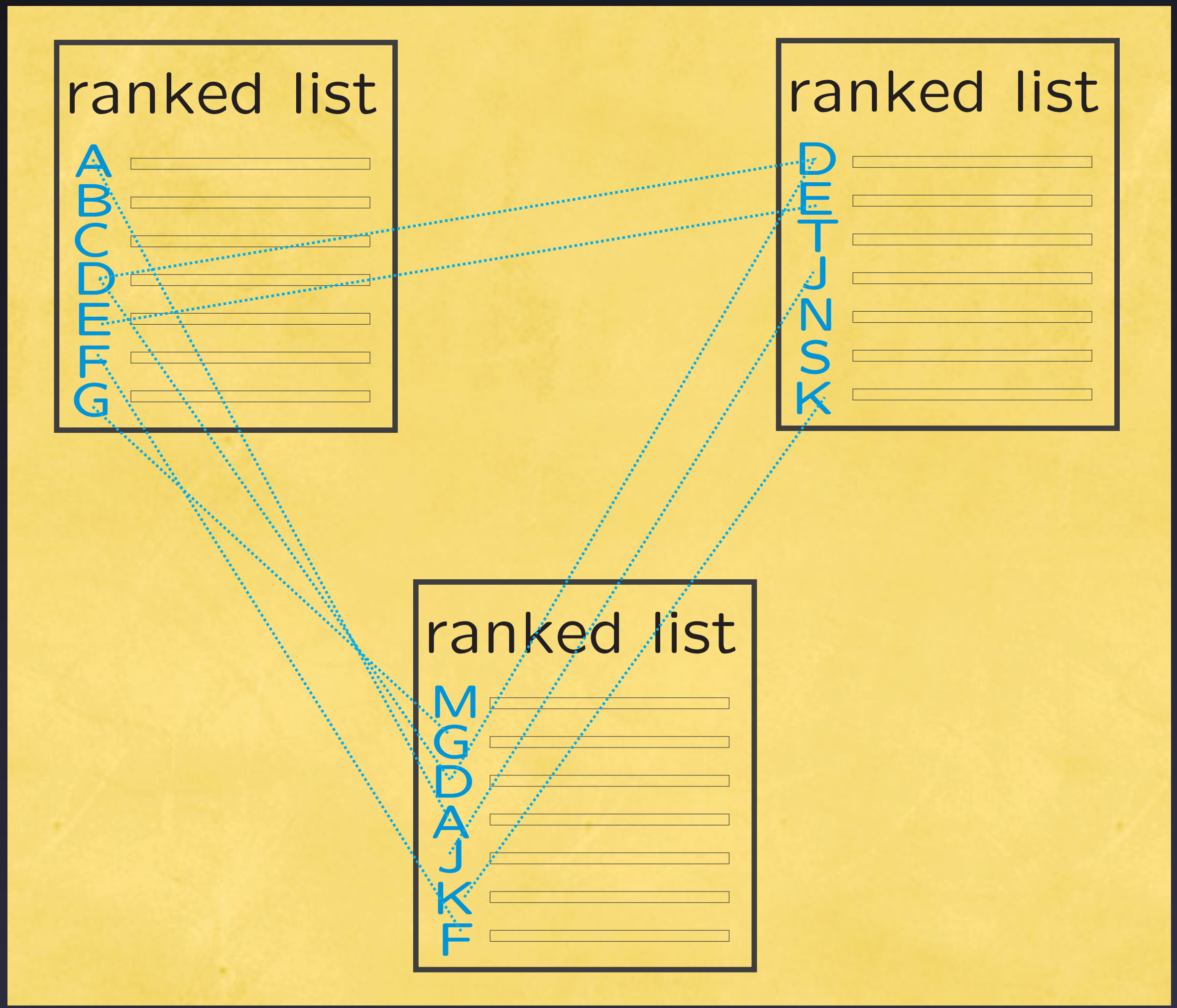
	online	batch
random	?	sampling
deterministic	Hedge	depth-pool

future work: maximum entropy

- How much information can be in one number ?
  - given AP of a list, what can we infer about it?
  - particular interest in Precision-Recall curve
- Maximum entropy method
  - given some constraints and an objective model
  - compute the instance of the model that has the highest entropy and satisfies the constraints
  - usually involves constrained nonlinear optimization

future work : query hardness

- Say we have access to many lists in response to the same query
- Query Hardness Hyp:
  - hard if the lists are diverse
  - easy if the lists are similar
- How to measure diversity?
  - transform lists in distributions
  - compare the distributions



future work: prec-recall models

- Task : construct prec-recall models based on search engine type
- Possible leads:
  - explore connection between AP, RP and MedianP
  - known distributions for rel/nonrel documents, also derived based on engine type [Manmatha et al, Robertson et al]
  - work by S.Robertson on fall-out curves
  - ROC curves
  - $p(r) = (1-r)/(1+\alpha * r)$

time line

Query Hardness Estimation	June 07
Max Entropy Application	June 07
Priors, Prec-Recall Models	October 07
Sampling: bias	January 08
Sampling: variance, CI	January 08
Active learning	April 08
Defend thesis	Summer 08

60

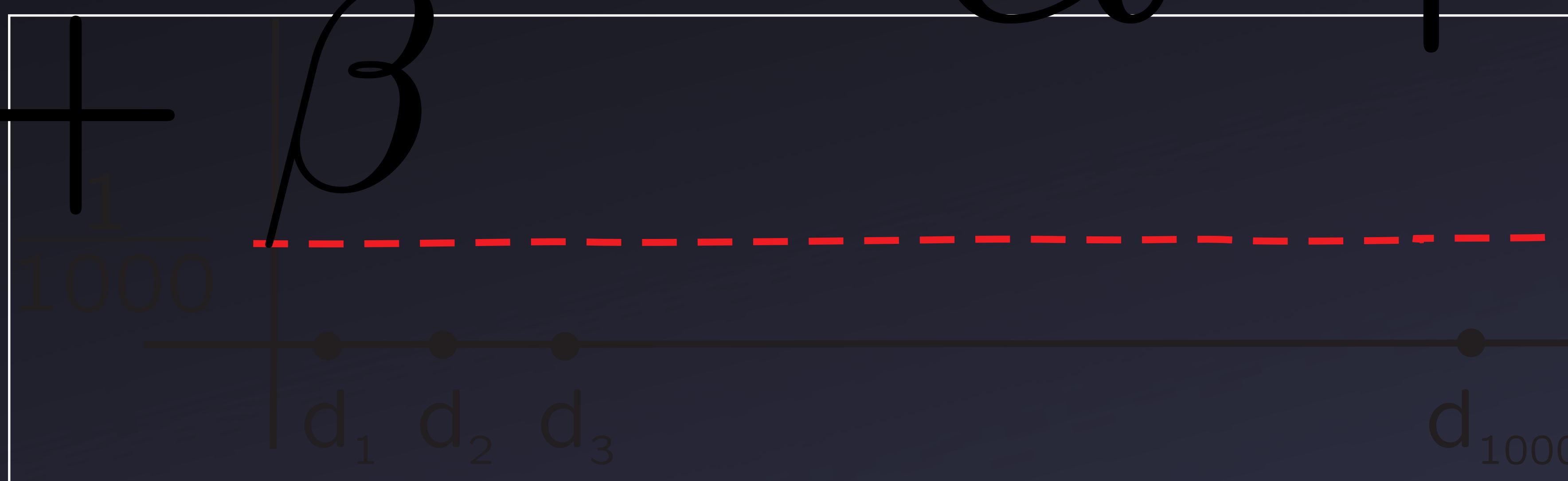
Thank You

$$\alpha + \beta$$

Text

$$\alpha + \beta$$

$$\alpha + \beta$$



$$n = 4p \cdot (1-p)/(0.03)^2 \approx 833$$