

Modelling A User Population for Designing Information Retrieval Metrics

Tetsuya Sakai
(NewsWatch, Inc.)
tetsuyasakai@acm.org

Stephen Robertson
(Microsoft Research Cambridge)
ser@microsoft.com

EVIA 2008, December 16, 2008@NII, Tokyo

TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

Average Precision (AP)

$$AP = \frac{1}{R} \sum_r I(r) \frac{C(r)}{r}$$

Number of relevant docs

1 iff doc at r is relevant

Precision at rank r

- Used widely since the advent of TREC
- Mean over topics is referred to as “MAP”
- Cannot handle graded relevance (but many IR researchers just love it)

Criticisms of (Mean) Average Precision ((M)AP)

- AP may be a poor measure of user performance/satisfaction

[Turpin/Scholer SIGIR 06 etc.]

- **AP lacks a user model**

“there is no single user application that directly motivates MAP” [Buckley/Voorhees TREC book]

“there is no plausible search model that corresponds to MAP, because no user knows in advance the number of relevant answers present in the collection...”

[Moffat/Webber/Zobel SIGIR 07]

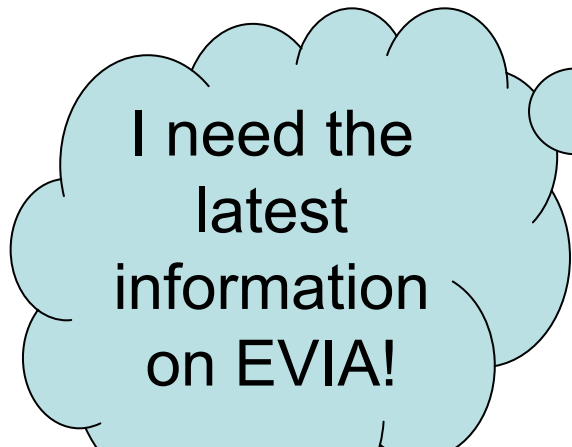
“AP lacks a user model?”




Rubbish!
[Robertson
SIGIR 08]

Objectives

- Robertson showed that AP is a special case of Normalised Cumulative Precision (NCP) which models a *population of users*.
- We generalise NCP and introduce Normalised Cumulative Utility (NCU), and show that
 - AP and Q-measure are in fact reasonable metrics!
 - A version of NCU, which utilises graded relevance in a novel way, has high discriminative power!



Information need



日本語のページだけを表示

ウェブ 764,000 件中 1-10 件 · 高度な検索
 検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#) Beta, [相談箱](#), [すべて表示...](#)

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access
research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (Day 1 of the NTCIR-7 Meeting) Workshop Venue:
ntcir.nii.ac.jp/index.php/ja/EVIA-2008 · [キャッシュ ページ](#)

[Eviaのホテル](#)

Evia ホテル予約 - 直前格安価格スペシャル ... ホテル名称 ロケーション ホテルランク 価格(〜から)
www.ratestogo.com/City/JP/Evia · [キャッシュ ページ](#)

[\[ntcir:240\] EVIA 2007: Call for Participation](#)

To: ntcir@xxxxxxxxx Subject: [ntcir:240] **EVIA** 2007: Call for Participation From: Atsushi Fujii <fujii@xxxxxxxxxxxxxxxxxxxx> Date: Wed, 02 May 2007 15:46:00 +0900 (JST)
research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

[アンフィニ・コメルス株式会社 高級輸入家具リリース ...](#)

Evia
www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction
www.eviaweb.gr · [キャッシュ ページ](#)

[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)
voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

 日本語のページだけを表示

ウェブ 764,000 件中 1-10 件 · 高度な検索
 検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

L1 (partially relevant)

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)

L3 (highly relevant)

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (Day 1 of the NTCIR-7 Meeting) Workshop Venue:

ntcir.nii.ac.jp/index.php/ja/EVIA-2008 · [キャッシュ ページ](#)

L0 (not relevant)

[Eviaのホテル](#)

Evia ホテル予約 - 直前格安価格スペシャル ... ホテル名称 ロケーション ホテルランク 価格(〜から)

www.ratestogo.com/City/JP/Evia · [キャッシュ ページ](#)

L2 (relevant)

[\[ntcir:240\] EVIA 2007: Call for Participation](#)

To: ntcir@xxxxxxxxx Subject: [ntcir:240] **EVIA** 2007: Call for Participation From: Atsushi Fujii <fujii@xxxxxxxxxxxxxxxxxxxx> Date: Wed, 02 May 2007 15:46:00 +0900 (JST)

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0 (not relevant)

[アンフィニ・コムルス株式会社 高級輸入家具リース ...](#)**Evia**

www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

L0 (not relevant)

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)

L3 (highly relevant)

[検索メインアック! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

Where do users stop scanning the list?



I stop at rank 1



L1

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)

I stop at rank 2



L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (Day 1 of the NTCIR-7 Meeting) Workshop Venue:

ntcir.nii.ac.jp/index.php/ja/EVIA-2008 · [キャッシュ ページ](#)

I stop at rank 4



L2

[Eviaのホテル](#)

Evia ホテル予約 - 直前格安価格スペシャル ... ホテル名称 ロケーション ホテルランク 価格(〜から)

www.ratestogo.com/City/JP/Evia · [キャッシュ ページ](#)

[\[ntcir:240\] EVIA 2007: Call for Participation](#)

To: ntcir@xxxxxxxxx Subject: [ntcir:240] **EVIA** 2007: Call for Participation From: Atsushi Fujii <fujii@xxxxxxxxxxxxxxxxxxxx> Date: Wed, 02 May 2007 15:46:00 +0900 (JST)

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0

[アンフィニ・コメルス株式会社 高級輸入家具リース ...](#)

Evia

www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)

I stop at rank 7



L3

[検索メインアック! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

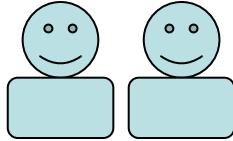
フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

p_u : Uniform Distribution over Relevant Documents



ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

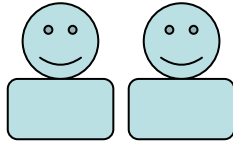


L1

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmww-en.html · [キャッシュ ページ](#)



L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (

ntcir.nii.ac.jp

ASSUMPTIONS:

- Users stop at a relevant doc;

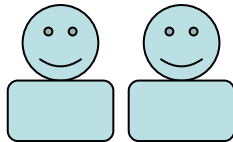
- The stopping probability is uniform across all relevant docs

L0

[Eviaのオ](#)

[Evia ホテル](#)

www.rates.com



L2

[\[ntcir:24\(](#)

To: ntcir@

fujii@xxx

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0

[アンフィニ・コメルス株式会社 高級輸入家具リリース ...](#)

[Evia](#)

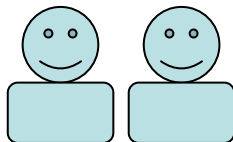
www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)



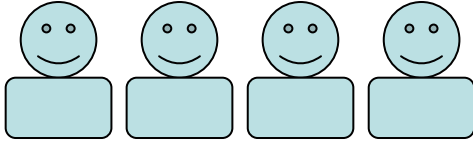
L3

[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

prb: Rank-Biased Distribution over Relevant Docs



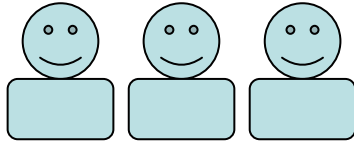
L1

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)



L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2006

ntcir.nii.ac.jp/

ASSUMPTIONS:

- Users stop at a relevant doc;

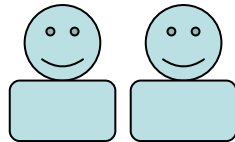
- **Users tend to stop near the top than near the bottom**

L0

[Evia](#)

Evia ホームページ

www.rat.ac.jp/~evia/



L2

[\[ntcir:2\]](#)

To: ntcir

<fuji@ntcir.nii.ac.jp>

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0

[アンフィニ・コムルス株式会社 高級輸入家具リリース ...](#)

Evia

www.infini-commerce.co.jp/collection/modern/catteran/evia.html · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)



L3

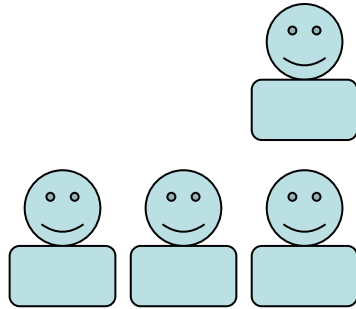
[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

p_{gu} : Graded-Uniform Distribution over Relevant Docs

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#) Beta, [相談箱](#), [すべて表示...](#)



L1

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)

L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (Day 1) ...
ntcir.nii.ac.jp/index.html

L0

[Eviaのホテル](#)

Evia ホテル予約 ...
www.ratestogo.co.jp

L2

[\[ntcir:240\] EVI](#)

To: ntcir@xxxxxx
<fujii@xxxxxxxxx>
research.nii.ac.jp

L0

[アンフィニ・コメルス株式会社 同級判別人家庭ソリューション...](#)

Evia

www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction
www.eviaweb.gr · [キャッシュ ページ](#)

L3

[検索メインアック! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

ASSUMPTIONS:

- Users stop at a relevant doc;
- Users tend to stop at a highly relevant doc than at a partially relevant doc

TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

Robertson's Normalised Cumulative Precision (NCP)

Probability that the user stops at the (relevant) document at rank n

$$NCP = \sum_{n=1}^{\infty} p_s(n) P(n)$$

Cost: #docs seen so far

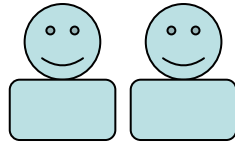
$$P(n) = C(n)/n$$

Expectation over a user population

Utility/Cost given the stopping point (precision at n)

Utility: #relevant seen so far

p_u : Uniform Distribution over Relevant Documents



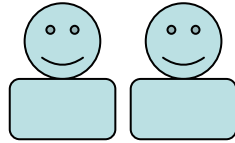
L1

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#) Beta, [相談箱](#), [すべて表示...](#)

[Evia](#)
provis
lingua
resea

ASSUMPTIONS:

- Users stop at a relevant doc;
- The stopping probability is uniform across all relevant docs

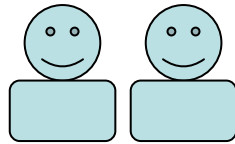


L3

[NTC](#)
NTCIF
16, 20
ntcir.i

Let $p_s(n) = p_u(n) = 1/R$

for every rank n that has a relevant doc.



L2

[Intcir](#)
To: nt
<fuji@
resea

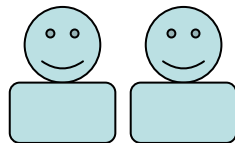
Then NCP reduces to AP (=NCP $_u$)!



L0

[Evia](#)
www.

$$NCP_u = \frac{1}{R} \sum_{n=1}^{\infty} I(n)P(n)$$



L3

[検索メーアツン!](#) (Information Retrieval Maniac!): [Evia 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

That is,

- AP is a special case of NCP.
- It is an expectation of utility/cost over a user population whose stopping probability is uniform across all relevant documents.
- Hence, AP is in fact a reasonable metric!

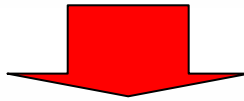


TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

We generalise NCP in two ways

$$NCP = \sum_{n=1}^{\infty} p_s(n)P(n)$$



$$NCU = \sum_{n=1}^{\infty} p_s(n)NU(n)$$

Stopping probability:

pu (uniform)

prb (rank-biased)

pgu (graded-uniform)

Normalised Utility:

BR(n) (blended ratio)

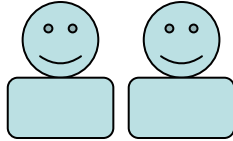
which generalises

P(n)

p_u : Uniform Distribution over Relevant Documents



ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

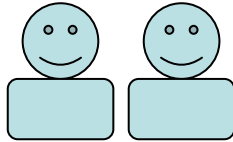


L1

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)

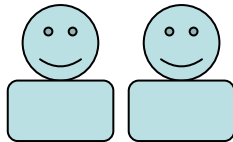


L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (

ntcir.nii.ac



L0

[Eviaのオ](#)

[Evia ホテル](#)

www.rates

L2

[\[ntcir:24\(](#)

To: ntcir@

fujii@xxx

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0

[アンフィニ・コメルス株式会社 高級輸入家具リリース ...](#)

[Evia](#)

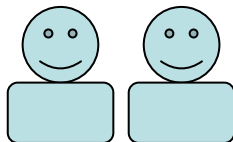
www.infini-commerce.co.jp/collection/modern/catteran/evia.ht... · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)



L3

[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

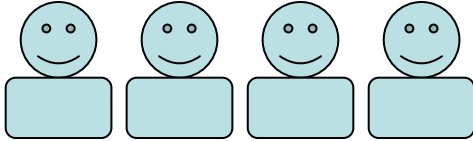
フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

ASSUMPTIONS:

- Users stop at a relevant doc;
- The stopping probability is uniform across all relevant docs

prb: Rank-Biased Distribution over Relevant Docs



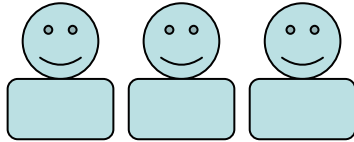
L1

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

[EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)

provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross-lingual information access

research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)



L3

[NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)

NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2006

ntcir.nii.ac.jp/

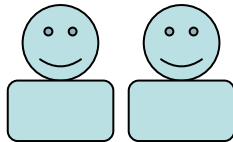
ASSUMPTIONS:

- Users stop at a relevant doc;

- **Users tend to stop near the top than near the bottom**

L0

[Evia](#)
Evia ホームページ
www.rat.ac.jp/~evia/



L2

[\[ntcir:2\]](#)
To: ntcir@nii.ac.jp
<fujii@nii.ac.jp>

research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)

L0

[アンフィニ・コメルス株式会社 高級輸入家具リリース ...](#)

Evia

www.infini-commerce.co.jp/collection/modern/catteran/evia.html · [キャッシュ ページ](#)

L0

[Υπό κατασκευή - Under construction](#)

Υπό κατασκευή - Under construction

www.eviaweb.gr · [キャッシュ ページ](#)



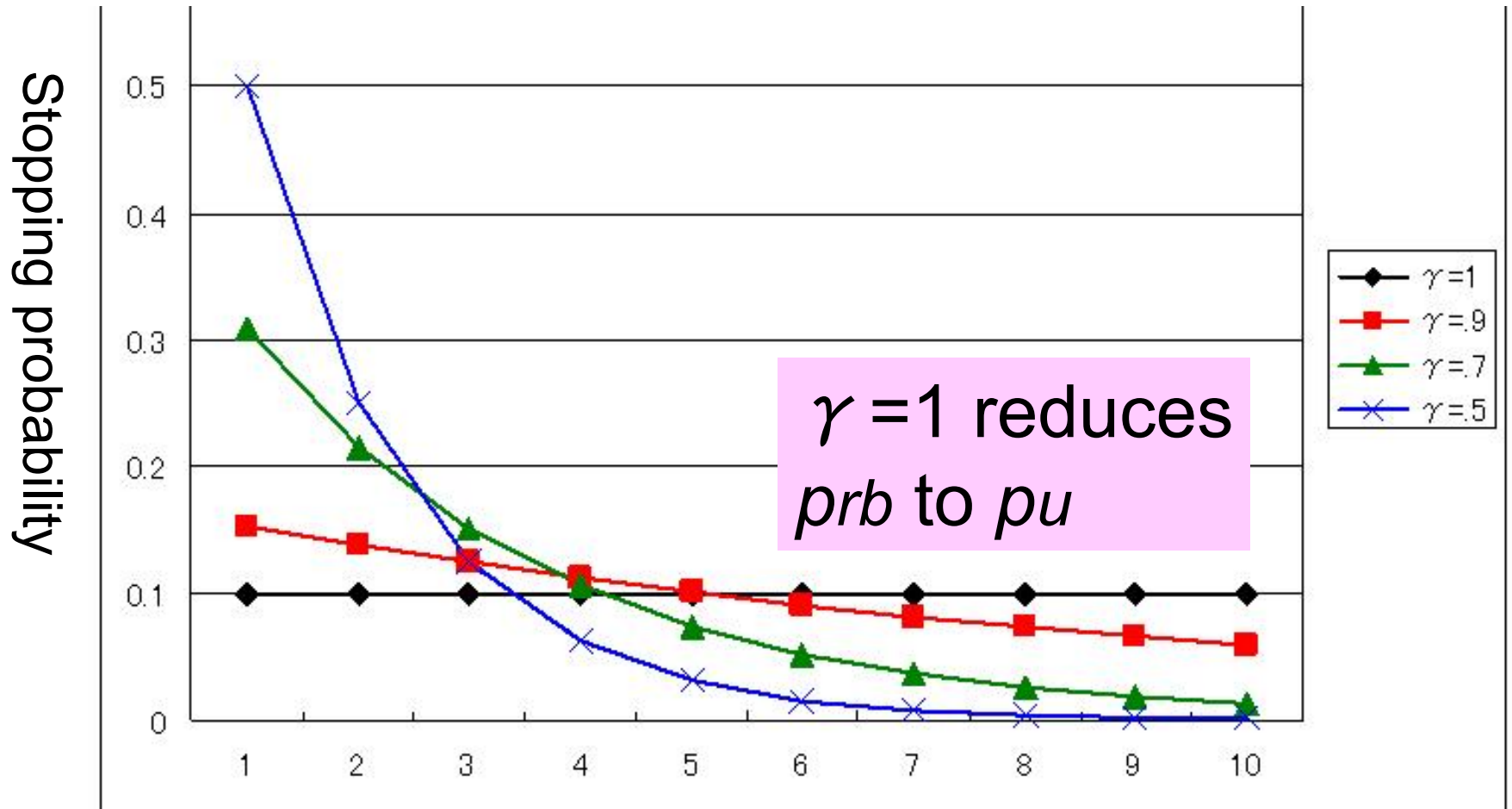
L3

[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

γ : top-heaviness parameter for prb

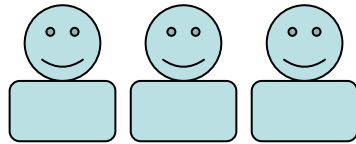


Relevant documents found in the ranked list

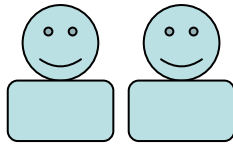
p_{gu} : Graded-Uniform Distribution over Relevant Docs



L1



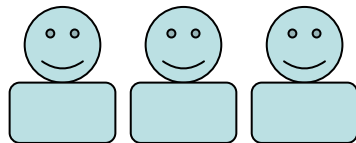
L3



L2

L0

L0



L3

ウェブ 764,000 件中 1-10 件 · 高度な検索
検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#) Beta, [相談箱](#), [すべて表示...](#)

[Evia](#) 2007 (NTCIR.6 Pre-Meeting Workshop)

provis
lingua
resea

ASSUMPTIONS:

- Users stop at a relevant doc;
- **Users tend to stop at a highly relevant doc than at a partially relevant doc**

[Ntcir](#)
NTCIR
16, 20
ntcir.i

[Evia](#)
Evia
www.

[Intcir](#)
To: nt
<fuji@
resea

[Evia](#)
Evia
www.

[Ypó](#)
Ypó
www.

Stopping weights

$\text{stop}(L3):\text{stop}(L2):\text{stop}(L1)=3:2:1$

$\text{stop}(L3):\text{stop}(L2):\text{stop}(L1)=10:5:1$

$(\text{stop}(L3):\text{stop}(L2):\text{stop}(L1)=1:1:1$

reduces p_{gu} to p_u)

[検索メイニアク! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)

フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(不協和音?)

voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

Blended Ratio (BR)

Precision

Normalised Cumulative Gain
for handling graded relevance

$$BR(n) = \frac{C(n) + \beta \sum_{i=1}^n g(i)}{n + \beta \sum_{i=1}^n g^*(i)}$$

A large β represents a very persistent user; $\beta = 0$ reduces BR to P

BR is suitable as a utility/cost function because, given the stopping point n , it does NOT matter where the relevant documents are within top n .

NCU family

$$NCU = \sum_{n=1}^{\infty} p_s(n) NU(n)$$

Stopping probability:
prb (rank-biased) with
top-heaviness parameter
 γ ($\gamma = 1$ reduces *prb* to *pu*)
pgu (graded-uniform)
with stopping weights
(flat weights
reduces *pgu* to *pu*)

Normalised Utility
given the stopping
point: *BR*(*n*)
(blended ratio) with
persistence
parameter β
($\beta = 0$ reduces
BR(*n*) to *P*(*n*))

TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

Comparing a system ranking by Metric M to that by AP

- Kendall's rank correlation

Monotonic function of the probability that a **randomly chosen system pair** is ordered identically in two rankings

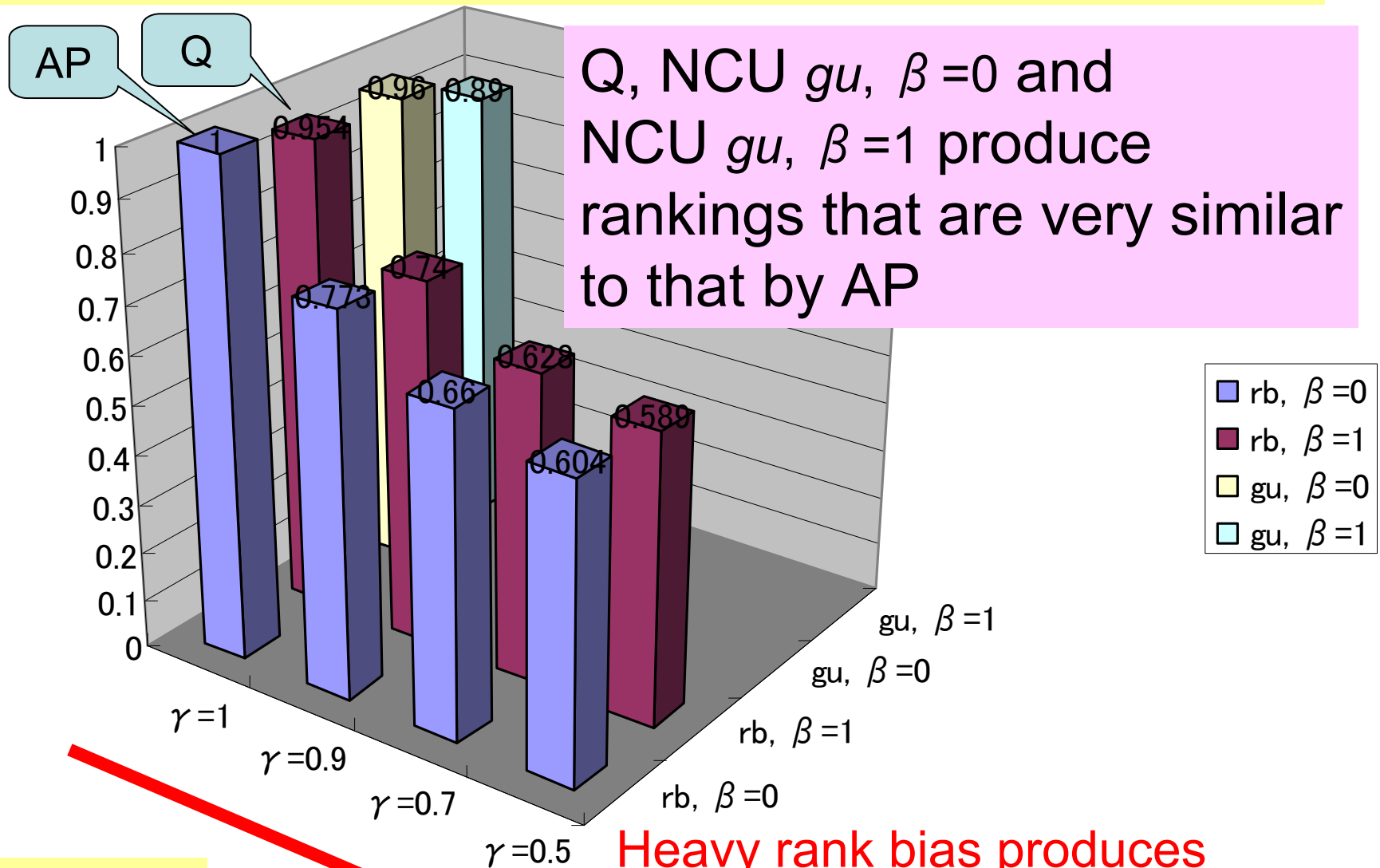
- Yilmaz/Aslam/Robertson (YAR) rank correlation
[SIGIR 08]

Monotonic function of the probability that a **randomly chosen system and one ranked above it** are ordered identically in two rankings

Assumes that the top ranks are the most important

Not symmetrical, but is almost symmetrical in practice

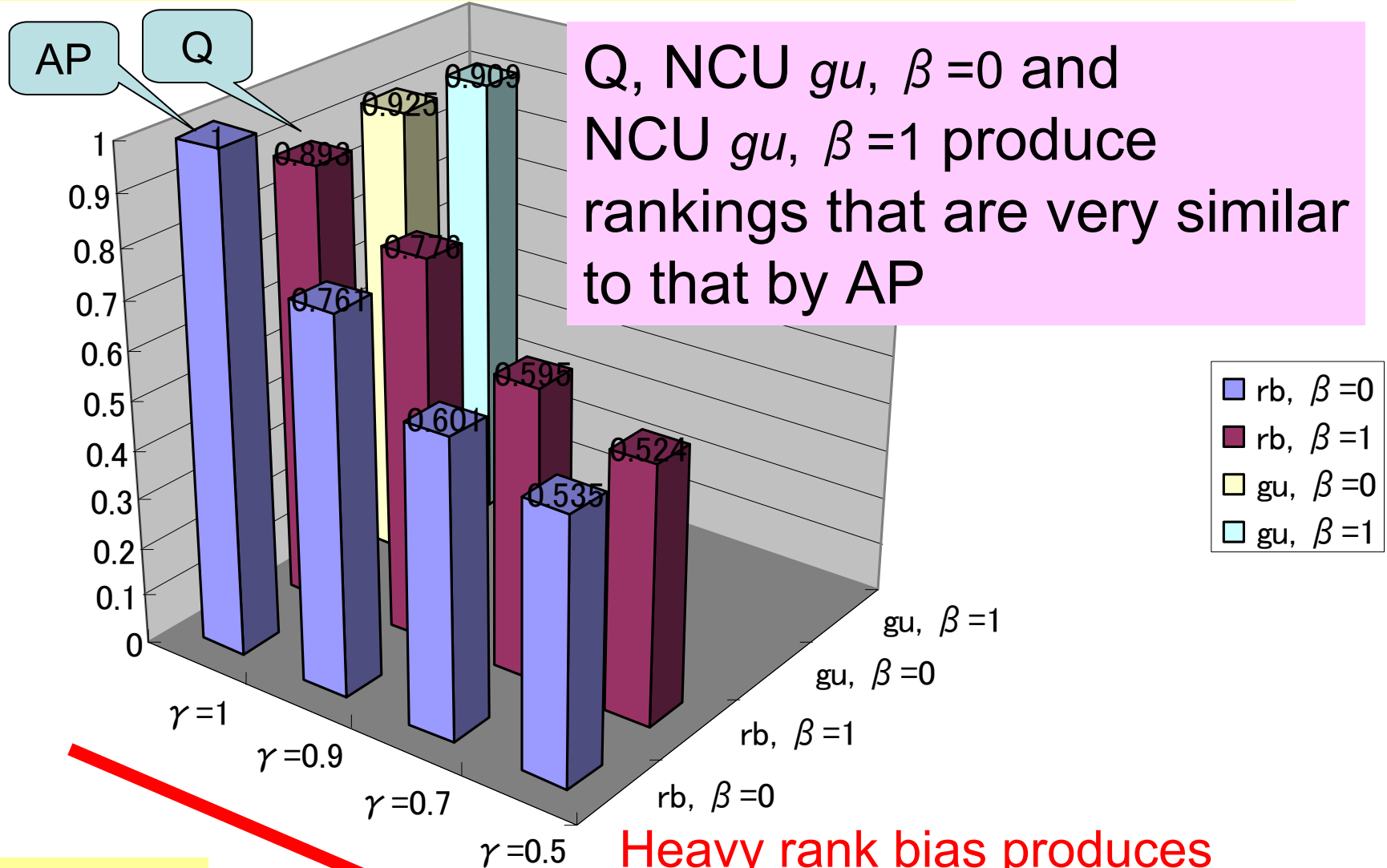
YAR rank correlation with AP (NCU $u, \beta = 0$): NTCIR-6J



Stop weights=3:2:1

Heavy rank bias produces very unconventional system rankings

YAR rank correlation with AP (NCU $u, \beta = 0$): TREC03



Stop weights=3:2:1

Heavy rank bias produces very unconventional system rankings

TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

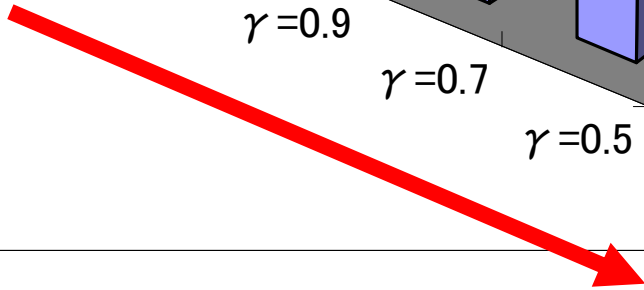
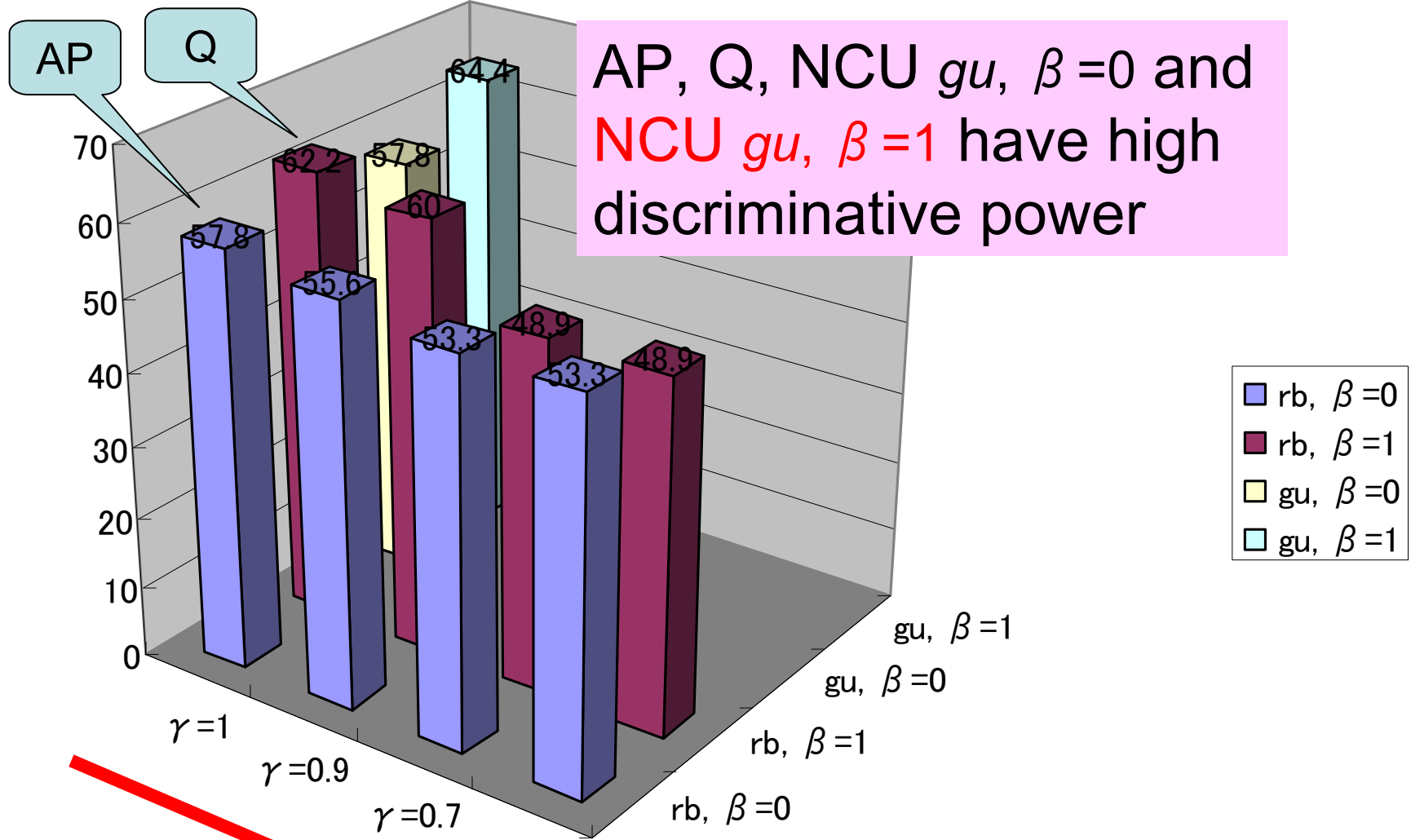
Measuring discriminative power of metrics [Sakai SIGIR06]

- Given a set of systems and a significance level α , for how many system pairs can a metric detect statistical significance?

Probability of Type I error $\alpha = 0.05 \Leftrightarrow 95\%$ confidence

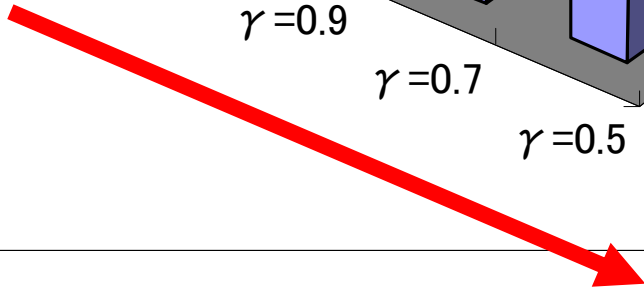
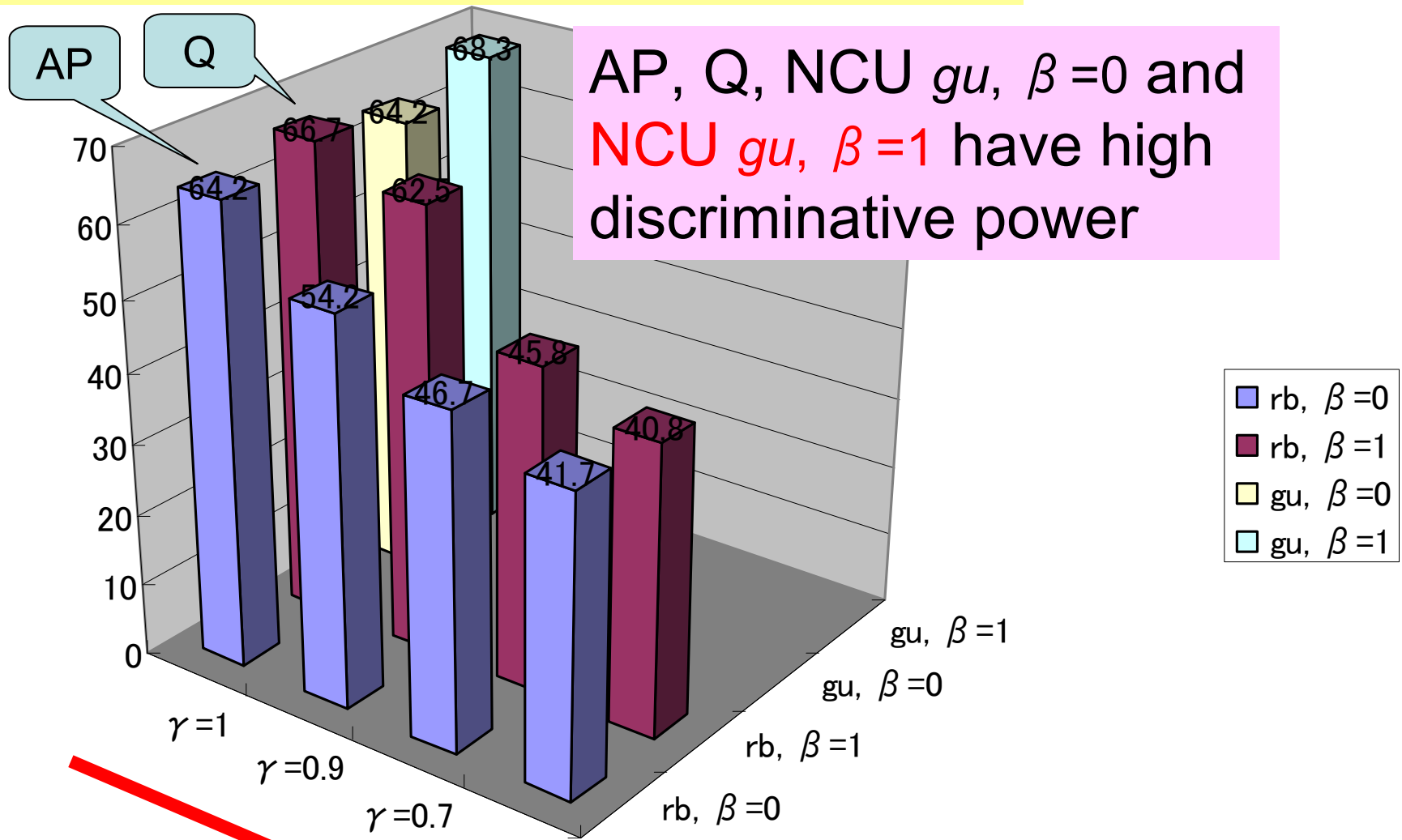
- Sakai's method uses the bootstrap test, and can also estimate the absolute performance difference required to achieve statistical significance (e.g. "a difference of 0.20 is usually statistically significant")
- Sakai's method and the Voorhees/Buckley swap method [SIGIR 02] give similar results in practice

Discriminative power at $\alpha = 0.05$: NTCIR-6J



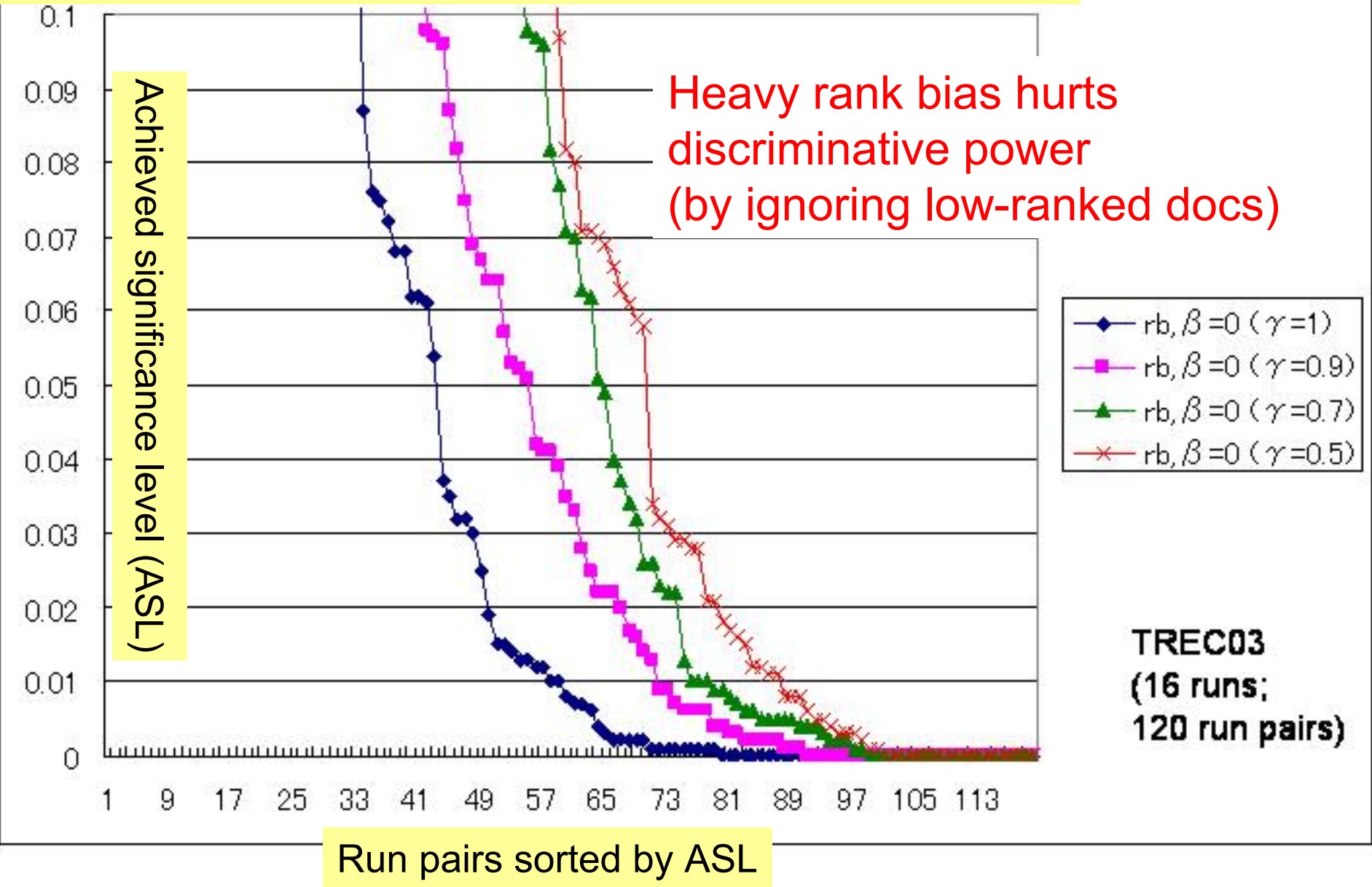
Heavy rank bias hurts discriminative power (by ignoring low-ranked docs)

Discriminative power at $\alpha = 0.05$: TREC03



Heavy rank bias hurts discriminative power (by ignoring low-ranked docs)

Effect of γ on discriminative power: TREC03



TALK OUTLINE

1. Objectives
2. Normalised Cumulative Precision (NCP)
3. Normalised Cumulative Utility (NCU)
4. Evaluating Evaluation Metrics:
Resemblance to Average Precision
5. Evaluating Evaluation Metrics:
Discriminative Power
6. Conclusions

Conclusions

We defined NCU, whose components are:

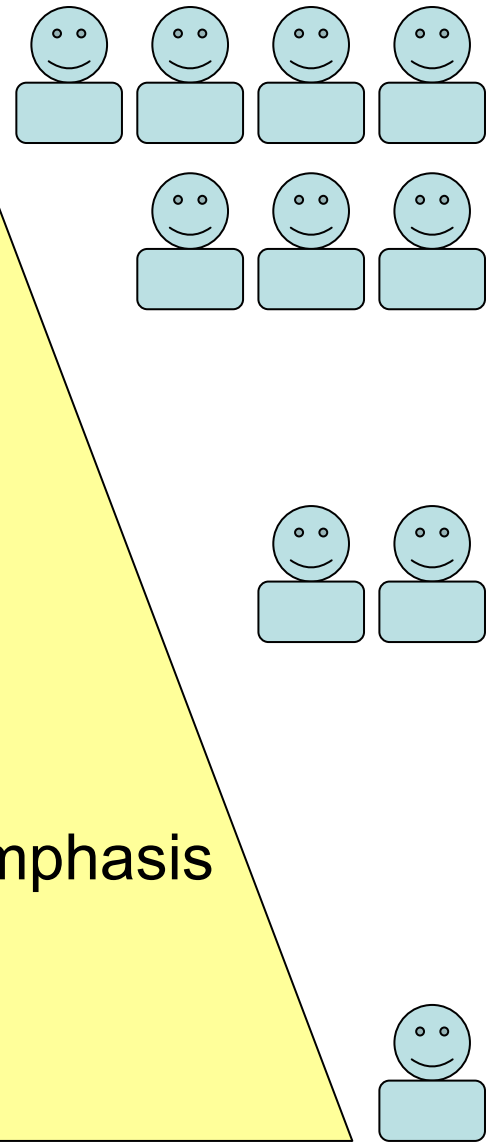
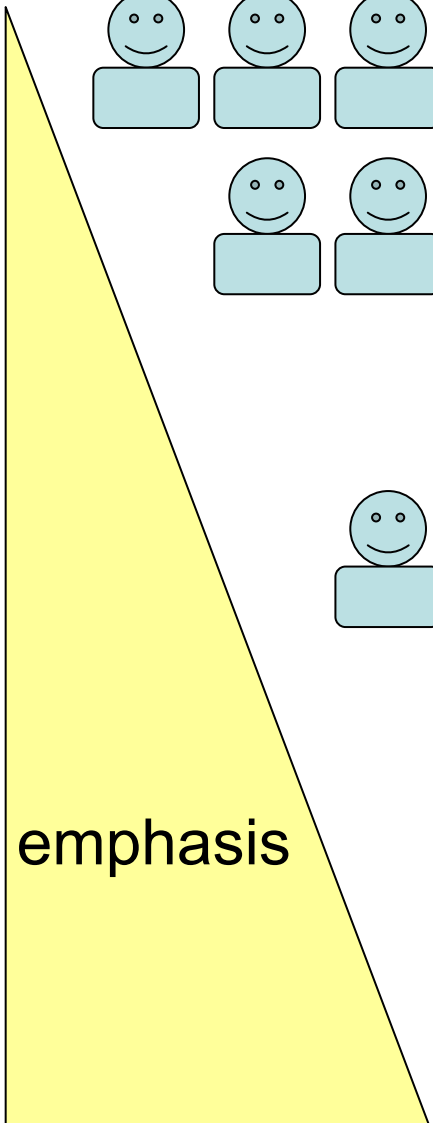
- Probability distribution of the user's stopping behaviour (p_u, p_{rb}, p_{gu})
- Blended Ratio (BR) as the utility/cost function given the stopping point

and showed that:

- Heavy rank-bias (small γ) is not desirable
- AP and Q, which rely on p_u , are reasonable metrics – they emphasize long-tail users who tend to dig deep into the ranked list and achieve high discriminative power
- NCU $g_u, \beta = 1$ has high consistency with AP and has the highest discriminative power (utilises graded relevance for both probability distribution p_{gu} and utility/cost BR)



Uniform distribution of (AP and Q) can be interpreted as...



- ウェブ 764,000 件中 1-10 件 · 高度な検索
 検索サービス: [画像](#), [動画](#), [ニュース](#), [地図](#)^{Beta}, [相談箱](#), [すべて表示...](#)

L1 [EVIA 2007 \(NTCIR-6 Pre-Meeting Workshop\)](#)
 provisional schedule of the 6th NTCIR Workshop, evaluation of IR, QA, Opinion, Trend and Cross lingual information access
research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html · [キャッシュ ページ](#)
- L3** [NTCIR-7 Information Retrieval Evaluation Site - EVIA](#)
 NTCIR-7 Information Retrieval Evaluation Workshop in Tokyo, Japan ... Workshop Date: December 16, 2008 (Day 1 of the NTCIR-7 Meeting) Workshop Venue:
ntcir.nii.ac.jp/index.php/ja/EVIA-2008 · [キャッシュ ページ](#)
- L0** [Eviaのホテル](#)
Evia ホテル予約 - 直前格安価格スペシャル ... ホテル名称 ロケーション ホテルランク 価格(~から)
www.ratestogo.com/City/JP/Evia · [キャッシュ ページ](#)
- L2** [\[ntcir:240\] EVIA 2007: Call for Participation](#)
 To: ntcir@xxxxxxxxx Subject: [ntcir:240] EVIA 2007: Call for Participation From: Atsushi Fujii <fujii@xxxxxxxxxxxxxxxxxxxx> Date: Wed, 02 May 2007 15:46:00 +0900 (JST)
research.nii.ac.jp/ntcir/archive/ntcir/msg00199.html · [キャッシュ ページ](#)
- L0** [アノフィニ・コムルス株式会社 高級輸入家具リース ...](#)
Evia
www.infini-commerce.com/evia.html · [キャッシュ ページ](#)
- L0** [Υπό κατασκευή - Under construction](#)
 Υπό κατασκευή - Under construction
www.eviaweb.gr · [キャッシュ ページ](#)
- L3** [検索メインアック! \(Information Retrieval Maniac!\): EVIA 2008 ...](#)
 フレッシュアイトップ 情報検索の研究者・酒井哲也が奏でる、研究とビジネスと私生活のハーモニー。(ク音?)
voice.fresheye.com/sakai/2008/08/evia_2008.html · [キャッシュ ページ](#)

Long-tail user

ir4qa evaluation scripts

- Simple scripts for computing AP, Q, nDCG, RBP, NCU etc. are available at:

http://research.nii.ac.jp/ntcir/tools/ir4qa_eval-en



NTCIR Project
Tools
ir4qa_eval

[\[JAPANESE\]](#) [\[NTCIR Home\]](#) [\[NTCIR Tools Home\]](#)

ir4qa_eval

These information retrieval (IR) evaluation scripts were developed for the NTCIR-7 ACLIA IR4QA subtask. They can be used for other IR tasks at NTCIR, TREC, etc. The scripts can compute average precision, Q-measure, nDCG and some other evaluation metrics. For more details, please read the README file included in the tar file.

Download

http://research.nii.ac.jp/ntcir/tools/ir4qa_eval.tar.gz

[\[JAPANESE\]](#) [\[NTCIR Home\]](#) [\[NTCIR DATA Home\]](#)

Thank you!