On Improving Wikipedia Search using Article Quality

Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady W. Lauw, Ba-Quy Voung

Meiqun Hu

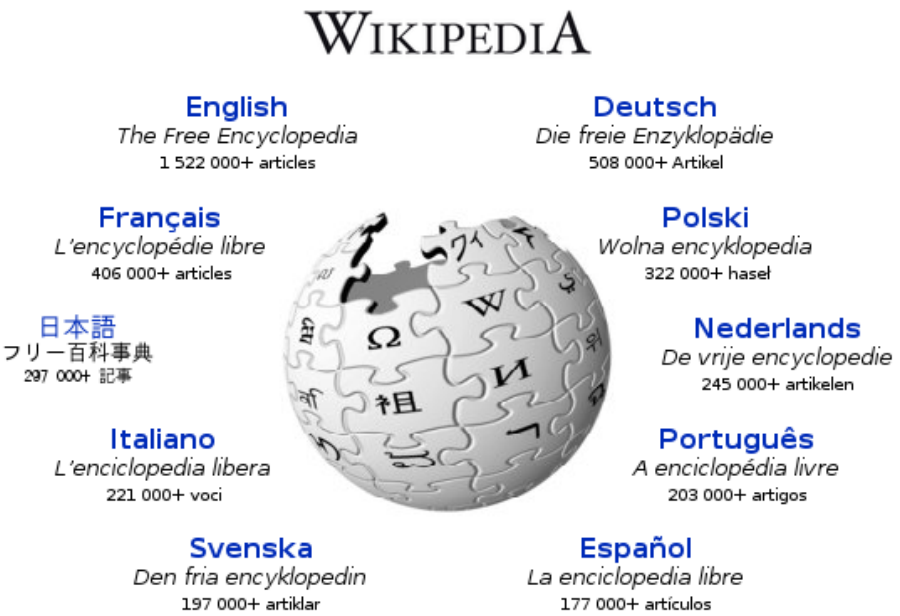
Nanyang Technological University

ACM *WIDM* 2007, Lisboa, Portugal

Outline

Road Map

Wikipedia



Wikipedia

- Web 2.0 service, aim for collaboration and interaction.
- Launched on January 15, 2001.
- Written collaboratively by volunteers.
- Has 236 language editions.
- Contains over 2 million articles in English Edition alone, marked on September 9, 2007.
- Top ten most-visited website worldwide.

Quality in Search

Open & Free	Any one can edit and create articles
	Any one can over–write content contributed by other people
Criticism on:	Information Accuracy
	Reputability of Third-party Sources
	Editorial and Systemic Bias
	Vandalism
	Uneven Quality

Issue

Searching performance compromised by poor quality articles.

Quality in Search

Open & Free	Any one can edit and create articles
	Any one can over–write content contributed by other people
Criticism on:	Information Accuracy
	Reputability of Third-party Sources
	Editorial and Systemic Bias
	Vandalism
	Uneven Quality

Issue

Searching performance compromised by poor quality articles.

Related Work on Incorporating Quality in IR

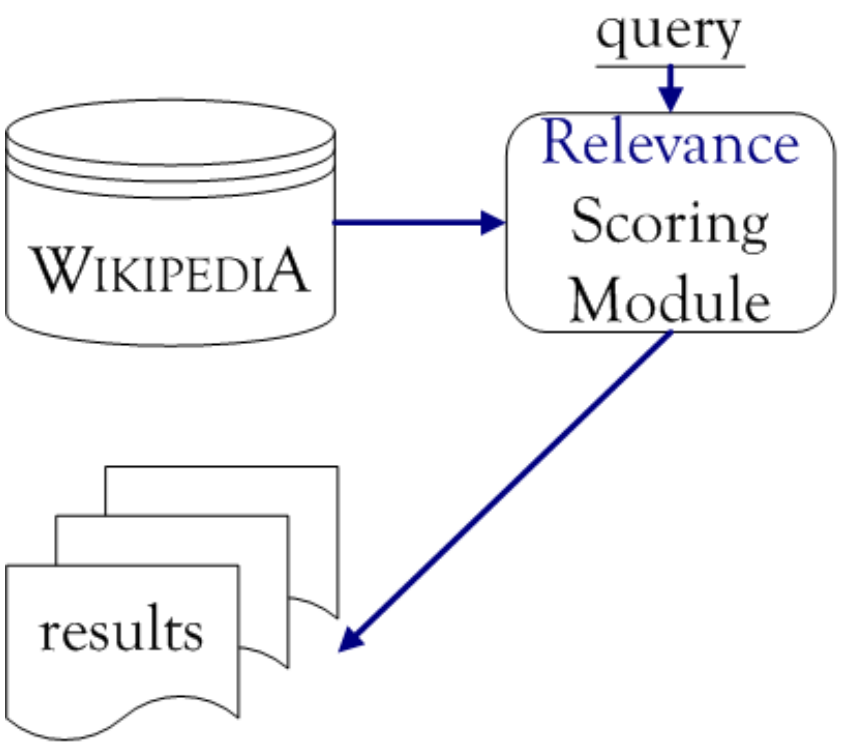
X. Zhu and S. Gauch.
Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web.
In *Proc. of SIGIR'00*, pages 288–295, July 2000.

Metrics:

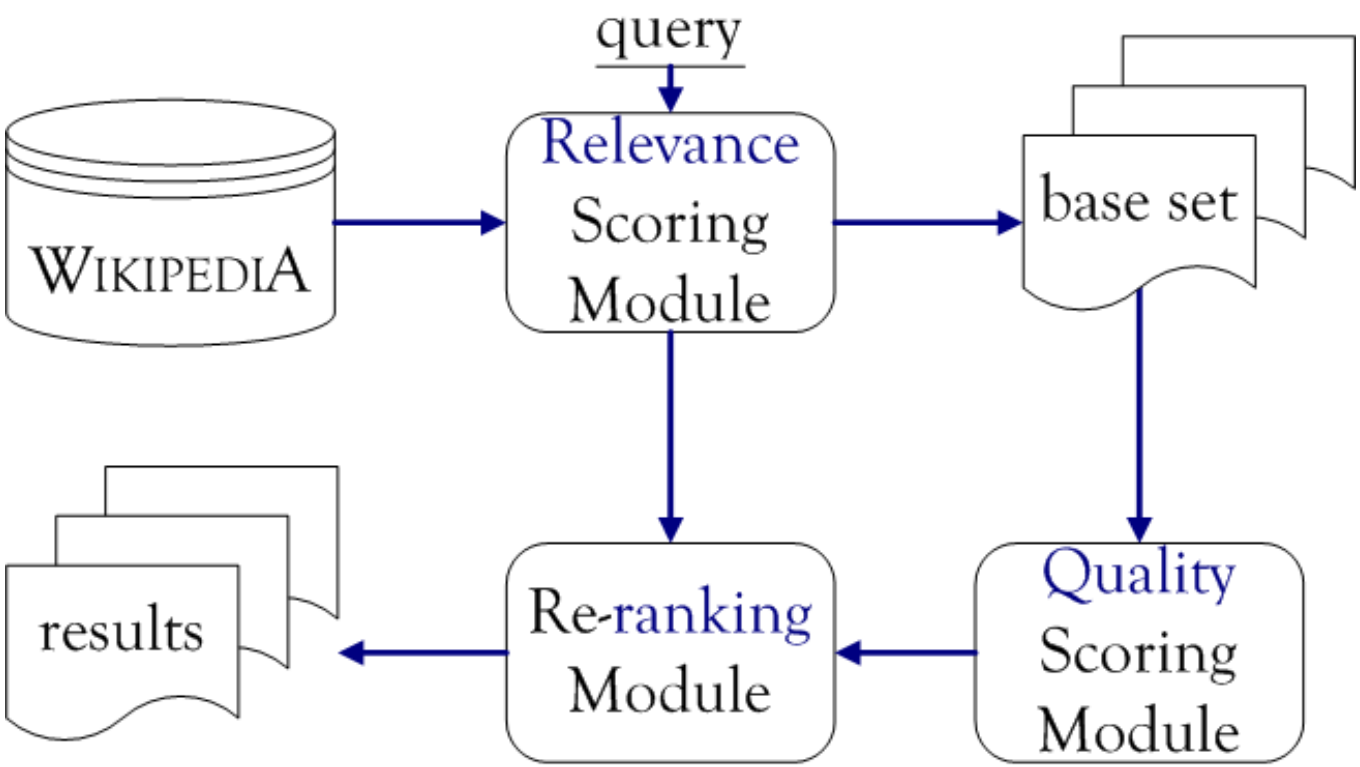
- currency
- availability
- information–to–noise ratio
- authority
- popularity
- cohesiveness

Road Map

A Sketch on the Existing Search Engine

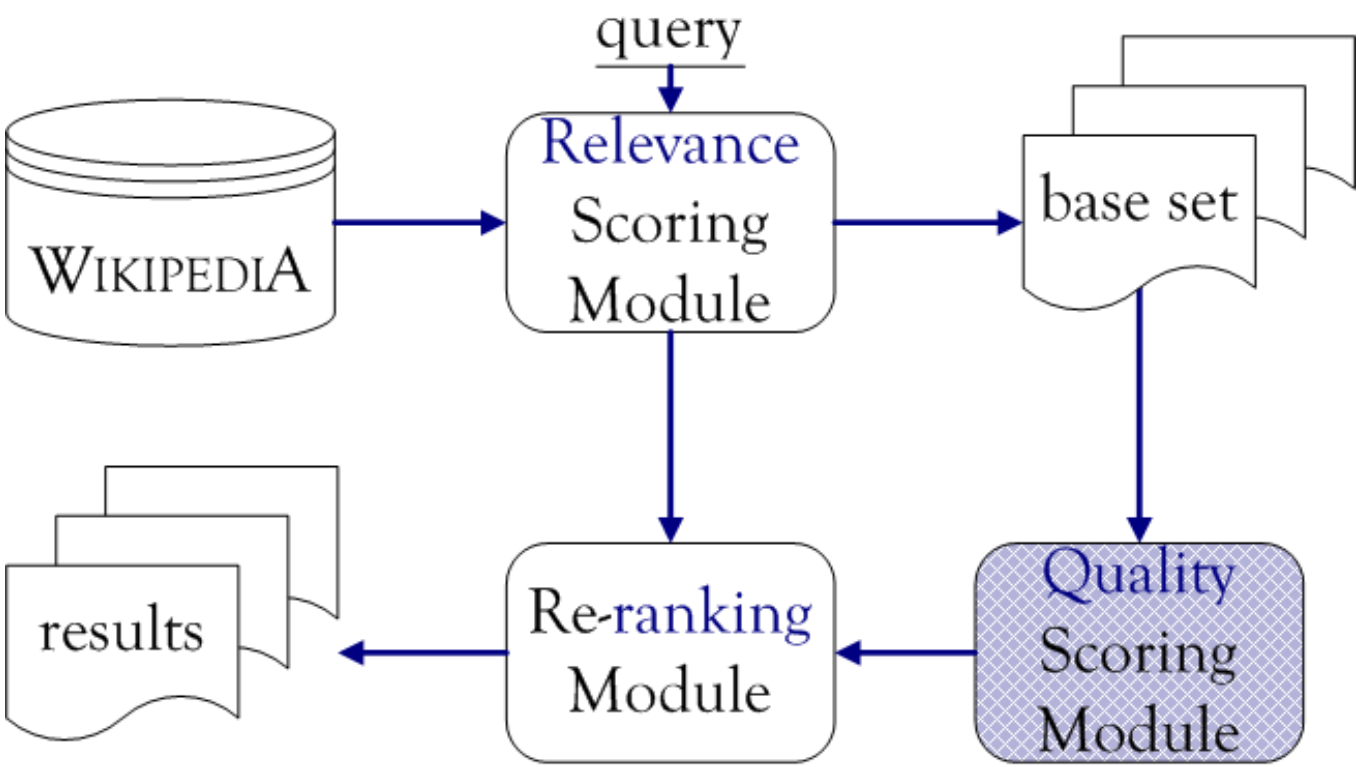


A Sketch on the Quality–aware Search Engine



Road Map

Quality Assessment Models



Naïve model

Naïve

The more words the articles has, the better the quality.

Drawback Not reliable

Easily be fooled

Naïve model

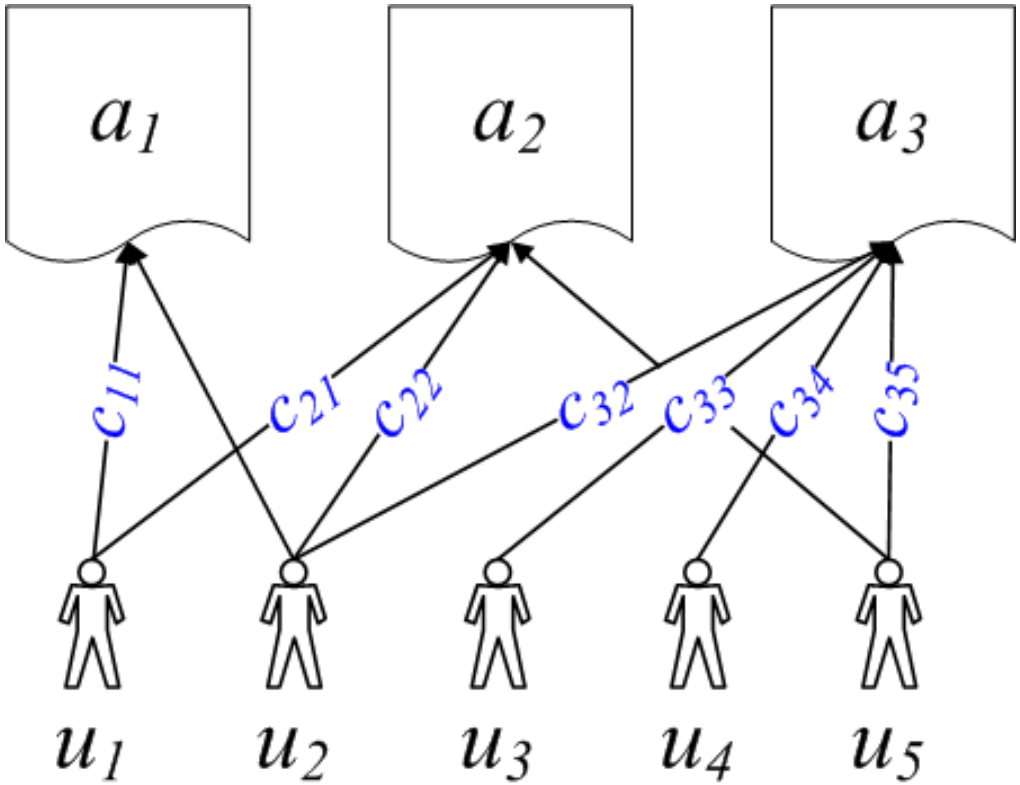
Naïve

The more words the articles has, the better the quality.

Drawback Not reliable

Easily be fooled

Article–Contributor Interaction



Basic model

Mutual Dependency between Quality and Authority

Good authors write good articles;
Good articles are written by good authors.

Basic

$$Q_i = \sum_j c_{ij} \times A_j \quad (1)$$

$$A_j = \sum_i c_{ij} \times Q_i \quad (2)$$

Revision Evolution and Effect of Reviewers

In collaborative editing, contributors will, in general,

1. read the article
2. examine on the various parts of the article
3. edit based on existing revision of the article

Assumption

- If content from earlier revision remains in current revision, then we say the editor of the current revision
  - is a reviewer of the unchanged content; and
  - agrees with the unchanged content.
- If some content of the article has been reviewed by high authority reviewers, then the content also carries high quality.

Revision Evolution and Effect of Reviewers

In collaborative editing, contributors will, in general,

1. read the article
2. examine on the various parts of the article
3. edit based on existing revision of the article

Assumption

- If content from earlier revision remains in current revision, then we say the editor of the current revision
  - is a reviewer of the unchanged content; and
  - agrees with the unchanged content.
- If some content of the article has been reviewed by high authority reviewers, then the content also carries high quality.

Revision Evolution and Effect of Reviewers

In collaborative editing, contributors will, in general,

1. read the article
2. examine on the various parts of the article
3. edit based on existing revision of the article

Assumption

- If content from earlier revision remains in current revision, then we say the editor of the current revision
  is a reviewer of the unchanged content; and
  agrees with the unchanged content.
- If some content of the article has been reviewed by high authority reviewers, then the content also carries high quality.

PeerReview model

PeerReview

$$q_{ik} = \sum_{w_{ik} \overset{A}{\leftarrow} u_j \vee w_{ik} \overset{R}{\leftarrow} u_j} A_j \quad (3)$$

$$A_j = \sum_{w_{ik} \overset{A}{\leftarrow} u_j \vee w_{ik} \overset{R}{\leftarrow} u_j} q_{ik} \quad (4)$$

and,

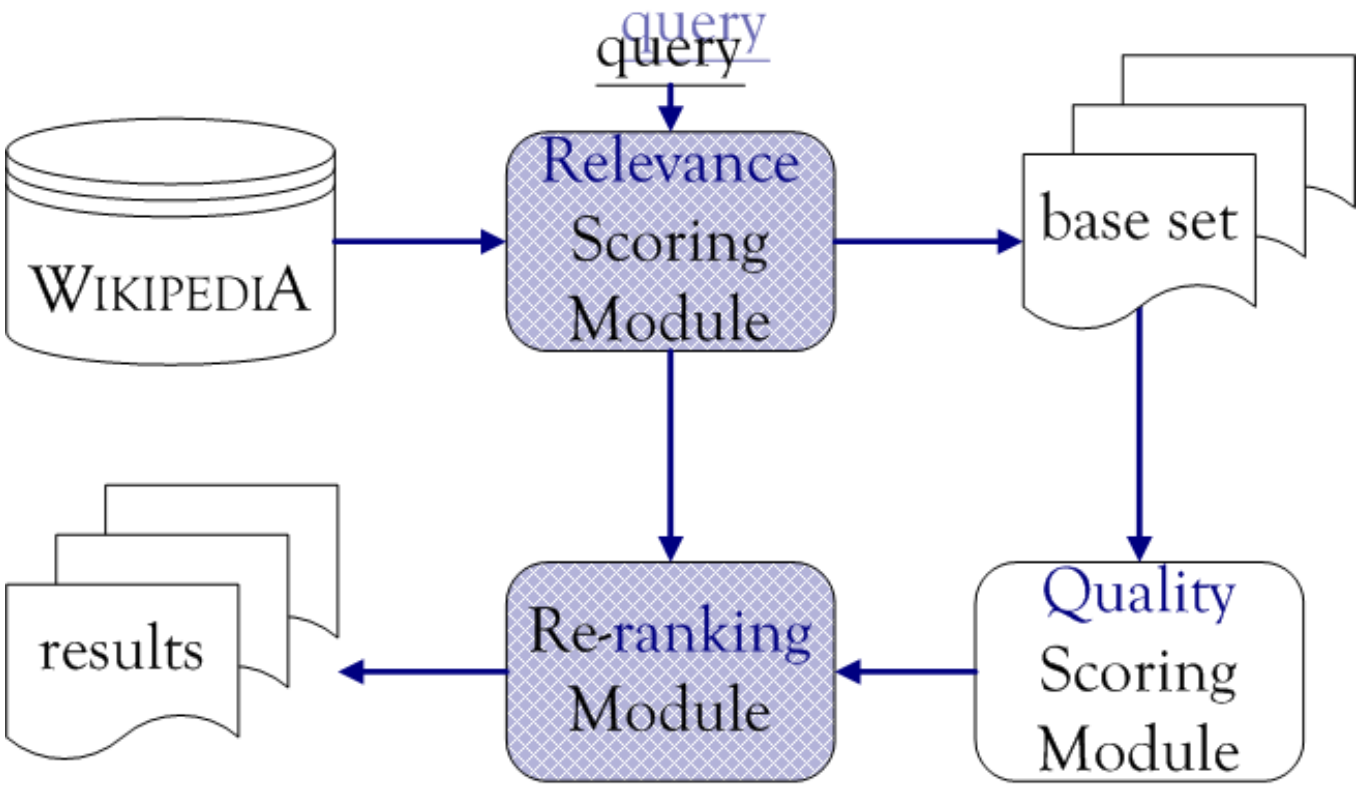
$$Q_i = \sum_{w_{ik} \in a_i} q_{ik}$$

.

- Authority of the reviewers are as important as that of the author;
- Authority of the contributors aggregate the quality of both authored and reviewed words.

Road Map

Experimental Design



Query Set

$$\underbrace{\text{single–term queries}}_{10} + \underbrace{\text{double–term queries}}_{10}$$

Queries carry general meaning.
Double–term queries are more specific than single–term queries.

Sources for the 20 Queries

P. Tsaparas.
Using non-linear dynamical systems for Web searching and ranking.
In *Proc. of PODS'04*, pages 59–70, June 2004.

C. Dwork, R. Kumar, M. Naor, and D. Sivakumar.
Rank aggregation methods for the Web.
In *Proc. of WWW'05*, pages 613–622, May 2005.

Relevance Scoring and the Base Set

Wiki

Google

Wikiseek

Base Set

Union of the top 500 (maximum) results from the three search engines.

Search Results Labeling

Assess and label top 10 results from each method.

Table: Decision Rules in User Assessment

Relevant	Quality	Label	$r(p)$
yes	high	Highly Recommended	2.0
yes	moderate	Recommended	1.0
yes	poor	Not Recommended	0.0
no	–	Not Recommended	0.0

Evaluation Metric

Normalized Discounted Cumulative Gain at top k

NDCG@k

$$G_q = \frac{1}{N_q} \sum_{p=1}^{k} \frac{2^{r(p)} - 1}{\log(1+p)}$$

The normalization factor, N_q, is determined such that a perfect ranking of top k articles will yield a NDCG of 1.
That is,

$$\underbrace{\underbrace{\mathsf{HR}\ldots\mathsf{HR}}_{n_q^{HR}} \prec \underbrace{\mathsf{R}\ldots\mathsf{R}}_{n_q^{R}} \prec \mathsf{NR}\ldots\mathsf{NR}}_{\text{top } k \text{ ranked results}}$$

K. Jarvelin and J. Kekalainen.
IR evaluation methods for retrieving highly relevant documents.
In *Proc. of SIGIR'00*, pages 41–48, July 2000.

Methods to be Evaluated

Method Type	Abbreviation		
relevance-only	Wiki,	Google,	Wikiseek
quality-only	Naïve,	Basic,	PeerReview
average-rank	Wiki + {N,B,P}		
	Google + {N,B,P}		
	Wikiseek + {N,B,P}		

Re–ranking

$$\bar{s}_i = \gamma_q \times s^{\text{rel}}(a_i) + (1 - \gamma_q) \times s^{\text{qual}}(a_i)$$

Average–Rank Method

$\gamma_q = \frac{1}{2}$ for all q

$s^{\text{rel}}(a_i)$ relevance rank for a_i from the search engine results

$s^{\text{qual}}(a_i)$ normalized quality rank for a_i from the quality ranking

Methods to be Evaluated

Method Type	Abbreviation		
relevance-only	Wiki,	Google,	Wikiseek
quality-only	Naïve,	Basic,	PeerReview
average-rank	Wiki + {N,B,P}		
	Google + {N,B,P}		
	Wikiseek + {N,B,P}		

Re–ranking

$$\bar{s}_i = \gamma_q \times s^{\text{rel}}(a_i) + (1 - \gamma_q) \times s^{\text{qual}}(a_i)$$

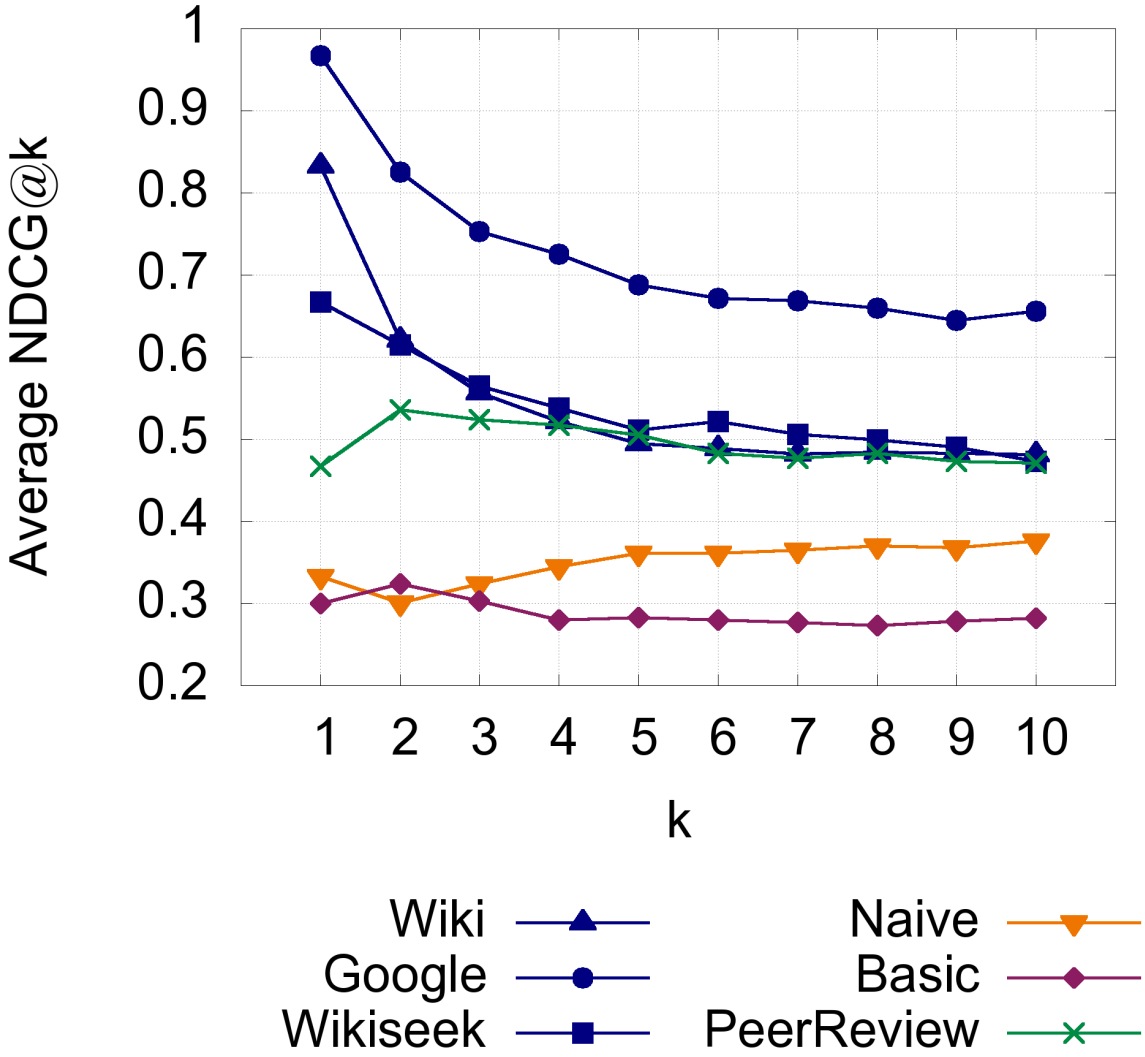
Average–Rank Method

$\gamma_q = \frac{1}{2}$ for all q

$s^{\text{rel}}(a_i)$ relevance rank for a_i from the search engine results

$s^{\text{qual}}(a_i)$ normalized quality rank for a_i from the quality ranking

Experimental Results
Non–combined Methods

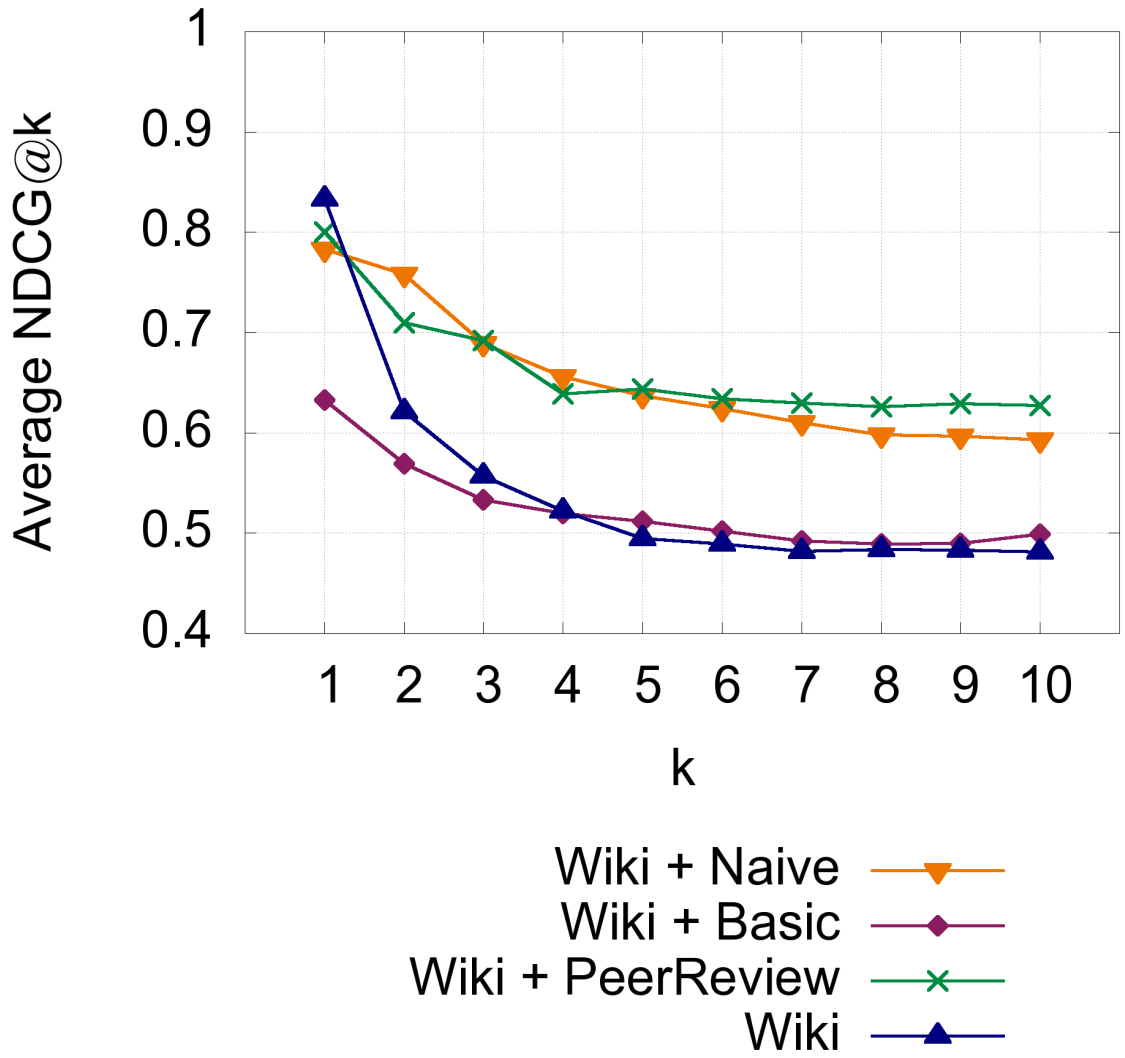


Observations

- Relevance supersede Quality, esp., at small k
- Relevance alone, Google best
- Quality alone, PeerReview best

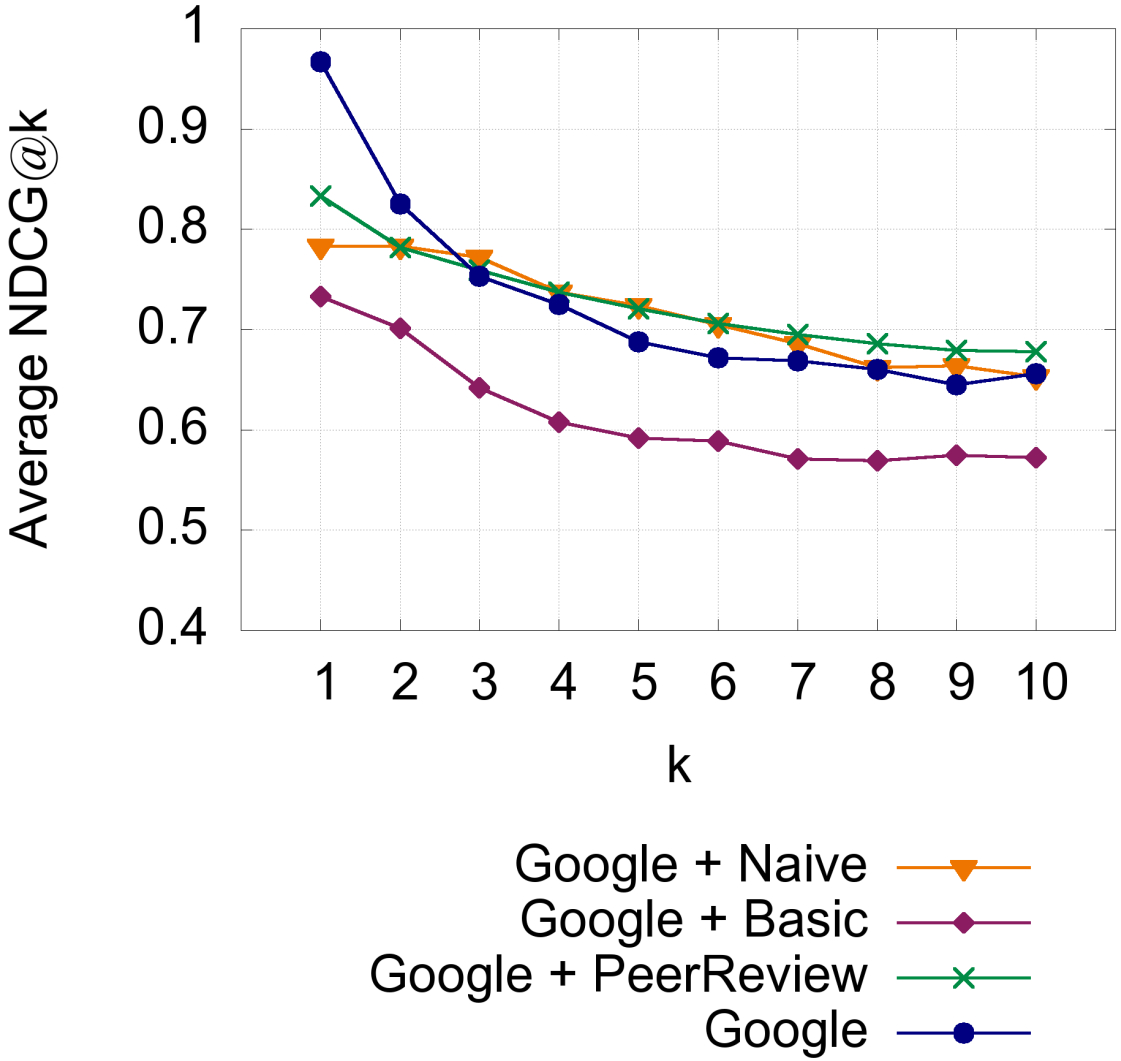
Experimental Results

Improvement over Wiki Method



Experimental Results

Quality–aware Methods compared with Google Method



Quality factor in Google's searching results

backlink

↓

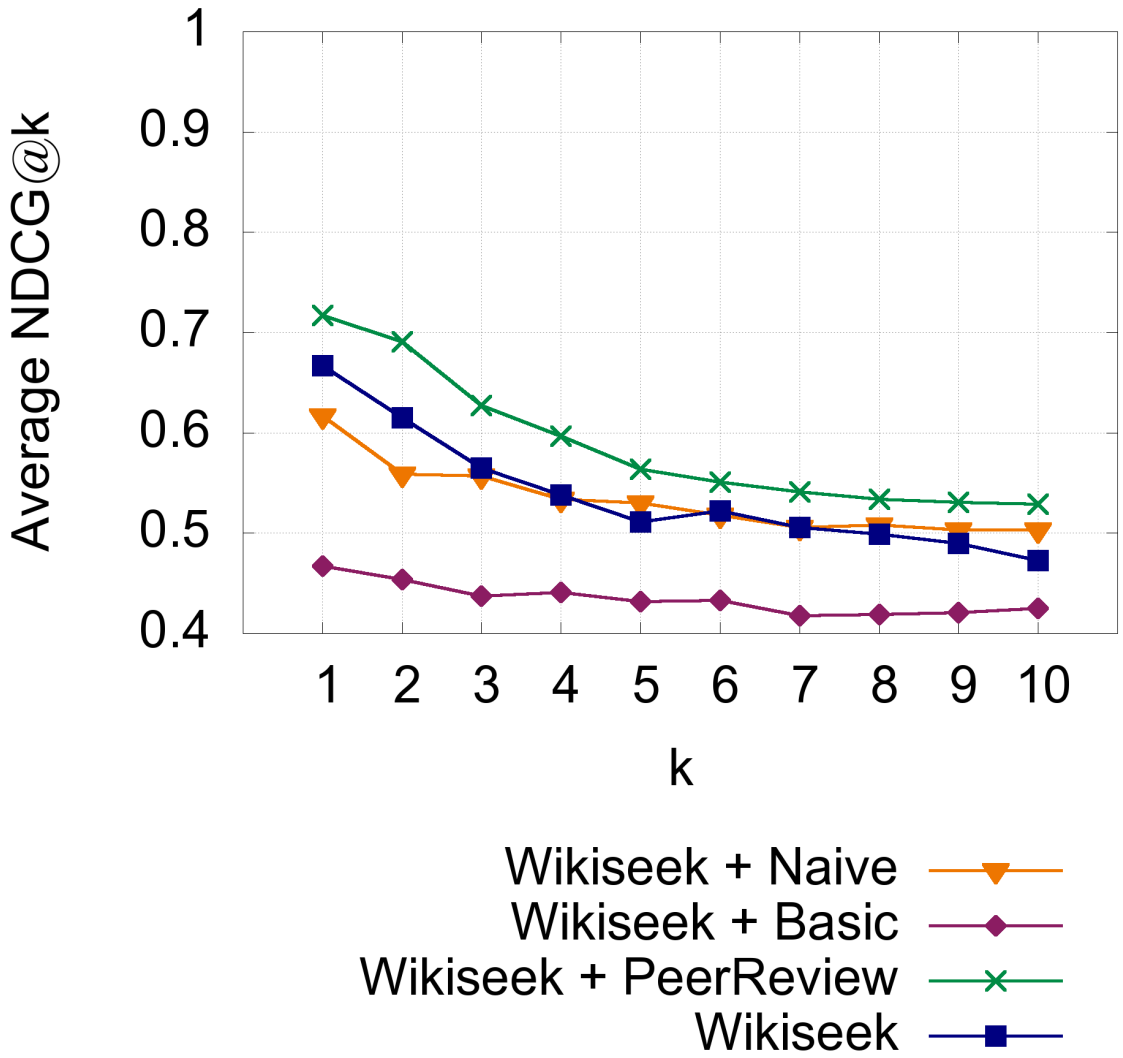
traffic

↓

improvement

Experimental Results

Quality–aware Methods compared with Wikiseek Method



Road Map

Conclusion

- Quality improves search results
- Quality based on the interaction of contributors in collaborative editing
- PeerReview is robust in measuring article quality
- Room for improvement
  - Base Set construction
  - Weighting in re–ranking
  - Authority in contributors

Thank You

Conclusion

- Quality improves search results
- Quality based on the interaction of contributors in collaborative editing
- PeerReview is robust in measuring article quality
- Room for improvement

  Base Set construction
  Weighting in re–ranking
  Authority in contributors

Thank You

Conclusion

- Quality improves search results
- Quality based on the interaction of contributors in collaborative editing
- PeerReview is robust in measuring article quality
- Room for improvement
  - Base Set construction
  - Weighting in re–ranking
  - Authority in contributors

Thank You

Bibliography

B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic.
Real life information retrieval: a study of user queries on the Web.
ACM SIGIR Forum, 32(1):5–17, April 1998.

J. Jeon, W. B. Croft, J. H. Lee, and S. Park.
A framework to predict the quality of answers with non-textual features.
In *Proc. of SIGIR'06*, pages 228–235, August 2006.

T. Mandl.
Implementation and evaluation of a quality-based search engine.
In *Proc. of HYPERTEXT'06*, pages 73–84, August 2006.