On Improving Wikipedia Search using Article Quality
Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady W. Lauw, Ba-Quy Young

Meiqun Hu
Nanyang Technological University
ACM WIDM 2007, Lisboa, Portugal
Outline

1. Introduction
2. Quality–aware Search Framework
3. Quality Assessment Models
4. Experimental Design and Results Analysis
5. Conclusion
Road Map

1. Introduction
2. Quality–aware Search Framework
3. Quality Assessment Models
4. Experimental Design and Results Analysis
5. Conclusion
Wikipedia

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.
X. Zhu and S. Gauch.

Metrics:

- currency
- availability
- information-to-noise ratio
- authority
- popularity
- cohesiveness
A Sketch on the Existing Search Engine

Wikimedia

query

Relevance Scoring Module

results
A Sketch on the Quality–aware Search Engine
Road Map

1. Introduction
2. Quality-aware Search Framework
3. Quality Assessment Models
4. Experimental Design and Results Analysis
5. Conclusion
Quality Assessment Models

On Improving Wikipedia Search using Article Quality
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

\[ a_1 \rightarrow C_{11} \rightarrow u_1 \]
\[ a_2 \rightarrow C_{21} \rightarrow u_2 \]
\[ a_2 \rightarrow C_{22} \rightarrow u_3 \]
\[ a_3 \rightarrow C_{32} \rightarrow u_4 \]
\[ a_3 \rightarrow C_{33} \rightarrow u_5 \]
\[ a_3 \rightarrow C_{34} \rightarrow u_1 \]
\[ a_3 \rightarrow C_{35} \rightarrow u_2 \]
Basic model

**Mutual Dependency between Quality and Authority**

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

**Basic**

\[ Q_i = \sum_j c_{ij} \times A_j \]  \hspace{1cm} (1)

\[ A_j = \sum_i c_{ij} \times Q_i \]  \hspace{1cm} (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.
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.
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

\[ q_{ik} = \sum_{w_{ik} \leftarrow u_j \lor w_{ik} \xrightarrow{R} u_j} A_j \]  
\[ A_j = \sum_{w_{ik} \leftarrow u_j \lor w_{ik} \xrightarrow{R} u_j} q_{ik} \]  

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

1. Introduction
2. Quality–aware Search Framework
3. Quality Assessment Models
4. Experimental Design and Results Analysis
5. Conclusion
Experimental Design

- Query input
- Relevance Scoring Module
- Re-ranking Module
- Quality Scoring Module
- Wikipedia database
- Results
- Base set
Query Set

- 10 single-term queries
- 10 double-term queries

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

Sources for the 20 Queries

Title: Relevance Scoring and the Base Set

- Wiki
- Google
- Wikiseek

**Base Set**

*Union of the top 500 (maximum) results from the three search engines.*
Assess and label **top 10** results from each method.

<table>
<thead>
<tr>
<th>Relevant</th>
<th>Quality</th>
<th>Label</th>
<th>$r(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>high</td>
<td>Highly Recommended</td>
<td>2.0</td>
</tr>
<tr>
<td>yes</td>
<td>moderate</td>
<td>Recommended</td>
<td>1.0</td>
</tr>
<tr>
<td>yes</td>
<td>poor</td>
<td>Not Recommended</td>
<td>0.0</td>
</tr>
<tr>
<td>no</td>
<td>-</td>
<td>Not Recommended</td>
<td>0.0</td>
</tr>
</tbody>
</table>
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,

$\text{HR} \ldots \text{HR} \prec R \ldots R \prec \text{NR} \ldots \text{NR}$

$\left\{ \begin{array}{c}
\underbrace{\text{HR}_q} \prec \underbrace{R_q} \prec \underbrace{\text{NR}_q} \\
_{n^H_q} & _{n^R_q} & _{n^N_q}
\end{array} \right\}$

top $k$ ranked results

K. Jarvelin and J. Kekalainen.
IR evaluation methods for retrieving highly relevant documents.
## Methods to be Evaluated

<table>
<thead>
<tr>
<th>Method Type</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>relevance-only</td>
<td>Wiki, Google, Wikiseek</td>
</tr>
<tr>
<td>quality-only</td>
<td>Naïve, Basic, PeerReview</td>
</tr>
<tr>
<td>average-rank</td>
<td>Wiki + {N,B,P}</td>
</tr>
<tr>
<td></td>
<td>Google + {N,B,P}</td>
</tr>
<tr>
<td></td>
<td>Wikiseek + {N,B,P}</td>
</tr>
</tbody>
</table>

### Re-ranking

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

### Average-Rank Method

\[
\gamma_q = \frac{1}{2} \quad \text{for all } q
\]

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

\[s^{qual}(a_i) \quad \text{normalized quality rank for } a_i \text{ from the quality ranking} \]
Methods to be Evaluated

<table>
<thead>
<tr>
<th>Method Type</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>relevance-only</td>
<td>Wiki, Google, Wikiseek</td>
</tr>
<tr>
<td>quality-only</td>
<td>Naïve, Basic, PeerReview</td>
</tr>
<tr>
<td>average-rank</td>
<td>Wiki + {N,B,P}</td>
</tr>
<tr>
<td></td>
<td>Google + {N,B,P}</td>
</tr>
<tr>
<td></td>
<td>Wikiseek + {N,B,P}</td>
</tr>
</tbody>
</table>

Re–ranking

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

Average–Rank Method

\[ \gamma_q = \frac{1}{2} \text{ for all } q \]

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

\[ s^{qual}(a_i) \text{ normalized quality rank for } a_i \text{ from the quality ranking} \]
Observations
- Relevance supersedes Quality, esp., at small $k$
- Relevance alone, Google best
- Quality alone, PeerReview best
Experimental Results

Improvement over Wiki Method

![Graph showing improvement over Wiki Method](image)
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

![Graph showing comparison between Wikiseek and quality-aware methods]

- Wikiseek + Naive
- Wikiseek + Basic
- Wikiseek + PeerReview
- Wikiseek
Road Map

1. Introduction
2. Quality–aware Search Framework
3. Quality Assessment Models
4. Experimental Design and Results Analysis
5. Conclusion
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
B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. 
Real life information retrieval: a study of user queries on the Web. 

A framework to predict the quality of answers with non-textual features. 

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