Keyquery-Based Recommendation of Related Work

Anna Beyer

Bauhaus-Universität Weimar

Defense of Master’s Thesis
2014-10-17
Motivation
Motivation

Reference List
Motivation

Reference List

Google Scholar
Motivation

Reference List

Cited by
Motivation

Reference List

Cited by

Related Articles

Google Scholar
Motivation
Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

Basic Pipeline

Input Documents
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

Basic Pipeline

Input Documents → Candidate Retrieval → Candidate Documents
Problem Definition

Related Work Search

Given: A set of research papers.
Task: Find a set of topically similar research papers.

Basic Pipeline
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

Basic Pipeline

Input Documents \(\xrightarrow{\text{Candidate Retrieval}}\) Candidate Documents \(\xrightarrow{\text{Candidate Ranking}}\) Output Documents
Candidate Retrieval
Problem Definition

Candidate Retrieval

Citation Graph
Problem Definition

Candidate Retrieval

Citation Graph

Content
Related Work

Citation Graph-Based Methods

Simple Approaches  [Golshan et al., SIGMOD 2012]
Collaborative Filtering  [Sugiyama and Kan, JCDL 2013]
[Caragea et al., JCDL 2013]
Link Ranking Algorithm  [Ekstrand et al., RecSys 2010]
[Küçüktunç et al., JCDL 2013]

Content-Based Methods

Query-Based Approaches  [Bethard and Jurasky, CIKM 2010]
[He et al., WSDM 2011]
[Nascimento et al., JCDL 2011]
Translation Models  [Lu et al., CIKM 2011]
[Huang et al., CIKM 2012]
[Tang et al., SIGIR 2014]
Topic Models  [Tang and Zhang, PAKDD 2009]
[Kataria et al., AAAI 2010]
[El-Arini and Guestrin, SIGKDD 2011]
Related Work

**Citation Graph-Based Methods**

Simple Approaches  
[Golshan et al., SIGMOD 2012]

Collaborative Filtering  
[Sugiyama and Kan, JCDL 2013]
[Caragea et al., JCDL 2013]

Link Ranking Algorithm  
[Ekstrands et al., RecSys 2010]
[Küçüktunç et al., JCDL 2013]

**Content-Based Methods**

Query-Based Approaches  
[Bethard and Jurasky, CIKM 2010]
[He et al., WSDM 2011]
[Nascimento et al., JCDL 2011]

Translation Models  
[Lu et al., CIKM 2011]
[Huang et al., CIKM 2012]
[Tang et al., SIGIR 2014]

Topic Models  
[Tang and Zhang, PAKDD 2009]
[Kataria et al., AAAI 2010]
[El-Arini and Guestrin, SIGKDD 2011]
Baselines | Sofia Search

[Golshan et al., SIGMOD 2012]
Baselines | Sofia Search

[Golshan et al., SIGMOD 2012]
A Source Independent Framework for Research Paper Recommendation

In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.
A Source Independent Framework for Research Paper Recommendation

In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.
In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.
In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.

- “source independent”
- “independent framework”
A Source Independent Framework for Research Paper Recommendation

In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.

- “source independent”
- “independent framework”
- “framework research”
A Source Independent Framework for Research Paper Recommendation

In this paper we propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers.
Baselines | Google Scholar

"From keywords to keyqueries: content descriptors for the web"
Baselines | Google Scholar
Baselines

Overview

- Google Scholar
- Citation graph-based [Golshan et al., SIGMOD 2012]
- Content-based [Nascimento et al., JCDL 2011]
Overview

- Google Scholar
- Citation graph-based [Golshan et al., SIGMOD 2012]
- Content-based [Nascimento et al., JCDL 2011]
Baselines

Overview

- Google Scholar
- Citation graph-based [Golshan et al., SIGMOD 2012]
- Content-based [Nascimento et al., JCDL 2011]

concept of keyqueries [Gollub et al., SIGIR 2013] meaningful queries
Concept of Keyqueries

Scholarly paper recommendation via user's recent research interests
K. Sugiyma, M. Kan - Proceedings of the 10th annual joint conference ..., 2010 - dl.acm.org
Abstract We examine the effect of modeling a researcher's past works in recommending scholarly papers to the researcher. Our hypothesis is that an author's published works
Cited by 50 Related articles All 29 versions Cite Save

A source independent framework for research paper recommendation
Abstract As the number of research papers available on the Web has increased enormously over the years, paper recommender systems have been proposed to help researchers on
Cited by 21 Related articles Cite Save

On the recommending of citations for research papers
S. M. McNeer, L. Albert, D. Cosley ... - Proceedings of the ..., 2002 - dl.acm.org
... four collaborative filtering-based algorithms along with two other recommendation algorithms in the domain of research papers. ... The outline of the rest of the paper is as follows. ...
Cited by 262 Related articles All 12 versions Cite Save

A multi-criteria collaborative filtering approach for research paper recommendation in papyres
Abstract Graduate students, professors and researchers regularly access, review, and use large amounts of literature. In previous work, we presented Papyres, a Research Paper
Cited by 26 Related articles All 6 versions Cite Save
Concept of Keyqueries

[Gollub et al., SIGIR 2013]
Keyquery-Based Recommendation

Basic Pipeline

Keyphrase Extraction  →  Keyquery Formulation
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]
Keyquery-Based Recommendation

**Keyquery Formulation**

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework
Keyquery-Based Recommendation

**Keyquery Formulation**

Example: [Nascimento et al., JCDL 2011]

document recommendation query framework
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework

paper recommendation

1

2

3 [Nascimento et al., JCDL 2011]

4

5

...
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

document recommendation query framework

paper recommendation
Keyquery-Based Recommendation

**Keyquery Formulation**

Example: [Nascimento et al., JCDL 2011]

**paper recommendation query framework**

query

[1]

[2]

[3]

[4]

[5]

...
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework

query framework

1
2
3
4
5

...
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework paper
Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework paper

query framework paper

1
2
3
4
5
[Nascimento et al., JCDL 2011]
Keyquery-Based Recommendation

**Keyquery Formulation**

Example: [Nascimento et al., JCDL 2011]

---

paper recommendation query framework paper
Experiments

Cranfield-Style Experiment

(1) Dataset
(2) Topics
(3) Judgments
Experiments

Cranfield-Style Experiment

(1) Dataset Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments

Webis Computer Science Paper Corpus [187,000 papers]

<table>
<thead>
<tr>
<th>Field</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM ID</td>
<td>1498835</td>
</tr>
<tr>
<td>Title</td>
<td>Finding Text Reuse on the Web</td>
</tr>
<tr>
<td>Authors</td>
<td>Michael Bendersky, W. Bruce Croft</td>
</tr>
<tr>
<td>Conference</td>
<td>WSDM 2009</td>
</tr>
<tr>
<td>Abstract</td>
<td>With the overwhelming number of reports on […]</td>
</tr>
<tr>
<td>Keywords</td>
<td>Text reuse, information flow, web search</td>
</tr>
<tr>
<td>Text</td>
<td>A sufficiently large archive such […]</td>
</tr>
<tr>
<td>References</td>
<td>1092473, 1341557, 1390432, […]</td>
</tr>
<tr>
<td>Citations</td>
<td>2487688, 1840829, 2399184, […]</td>
</tr>
</tbody>
</table>
Experiments

**Cranfield-Style Experiment**

1. **Dataset**
   - Webis Computer Science Paper Corpus

2. **Topics**
   - User Study

3. **Judgments**
   - User Study

**Webis Computer Science Paper Corpus** [187,000 papers]

<table>
<thead>
<tr>
<th>Field</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM ID</td>
<td>1498835</td>
</tr>
<tr>
<td>Title</td>
<td>Finding Text Reuse on the Web</td>
</tr>
<tr>
<td>Authors</td>
<td>Michael Bendersky, W. Bruce Croft</td>
</tr>
<tr>
<td>Conference</td>
<td>WSDM 2009</td>
</tr>
<tr>
<td>Abstract</td>
<td>With the overwhelming number of reports on […]</td>
</tr>
<tr>
<td>Keywords</td>
<td>Text reuse, information flow, web search</td>
</tr>
<tr>
<td>Text</td>
<td>A sufficiently large archive such […]</td>
</tr>
<tr>
<td>References</td>
<td>1092473, 1341557, 1390432, […]</td>
</tr>
<tr>
<td>Citations</td>
<td>2487688, 1840829, 2399184, […]</td>
</tr>
</tbody>
</table>
Experiments

Cranfield-Style Experiment

(1) Dataset  Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments  \{User Study\}

User Study  [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset       Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments

User Study [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset  Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments  } User Study

User Study [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset  Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments  \{User Study\}

User Study [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset
Webis Computer Science Paper Corpus

(2) Topics

(3) Judgments
User Study

User Study [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset  Webis Computer Science Paper Corpus
(2) Topics
(3) Judgments } User Study

User Study [10 participants]
Experiments

Cranfield-Style Experiment

(1) Dataset  
Webis Computer Science Paper Corpus

(2) Topics

(3) Judgments  
User Study

User Study [10 participants]
## Results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MAP@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sofia Search</td>
<td>0.546</td>
</tr>
<tr>
<td>Nascimento</td>
<td>0.523</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>0.535</td>
</tr>
<tr>
<td>Keyquery</td>
<td>0.568</td>
</tr>
<tr>
<td>Keyquery + Google Scholar</td>
<td>0.605</td>
</tr>
</tbody>
</table>
Summary

- Related work search
- Keyquery-based approach
- User study
- Keyquery + Google Scholar performs best
### Summary

- Related work search
- Keyquery-based approach
- User study
- Keyquery + Google Scholar performs best

### Outlook

- Bigger user study
- Other datasets
- Keyquery formulation
Conclusion

Summary

- Related work search
- Keyquery-based approach
- User study
- Keyquery + Google Scholar performs best

Outlook

- Bigger user study
- Other datasets
- Keyquery formulation

Thank you!
STEP 1

Email
Please enter your email address.

anna.bayer@uni-weimar.de

Research Task
Please describe a research task you are familiar with in a few words (e.g. cluster labeling).

recommendation of research papers

Input Documents
Please enter at least 2 papers by title which match the research task you have specified above. The entered titles will be validated immediately. If the document cannot be found in our data collection, the field will be marked as invalid. Our data collection comprises around 187,000 documents published by ACM in the years 1982-2013.

A source independent framework for research paper recommendation
SOFIA SEARCH: a tool for automating related-work search

Expected Documents
Please enter at least 2 document which are related to your research task and input documents. As with the input documents, your input will be validated against our data collection.

Recommending citations: translating papers into references
Context-aware citation recommendation

Submit
STEP 1

Email
Please enter your email address.
anna.bayer@uni-weimar.de

Research Task
Please describe a research task you are familiar with in a few words (e.g. cluster labeling).

recommendation of research papers

Input Documents
Please enter at least 2 papers by title which match the research task you have specified above.
The entered titles will be validated immediately. If the document cannot be found in our data collection, the field will be marked as invalid. Our data collection comprises around 187,000 documents published by ACM in the years 1962-2013.

A source independent framework for research paper recommendation
SOFIA SEARCH: a tool for automating related-work search

Expected Documents
Please enter at least 2 document which are related to your research task and input documents.
As with the input documents, your input will be validated against our data collection.

Recommending citations: translating papers into references
Context-aware citation recommendation

Submit
**STEP 2**

**Related Documents**

The table below lists the documents found to be related in alphabetical order by title. You can expand a document’s abstract by clicking on **ABSTRACT**, and you can open the document as PDF file by clicking on **PDF**.

Please read the abstract of each paper and rate it regarding two criteria: relatedness and familiarity.

The **Level of Relatedness** indicates how related the document is with respect to your research task and input documents specified in STEP 1.

- **Highly**: The document matches my research task perfectly.
- **Fairly**: The document matches my research task.
- **Marginally**: The document includes only a few aspects of my research task.
- **Not Related**: The document does not match my research task in any respect.

The **Level of Familiarity** indicates whether you knew the document before this user study or not.

- **Familiar**: I knew the document before.
- **Unfamiliar**: I didn’t know the document before.

<table>
<thead>
<tr>
<th>Document</th>
<th>Level of Relatedness</th>
<th>Level of Familiarity</th>
</tr>
</thead>
</table>
| ABSTRACT PDF  
**Anchor Text Extraction for Academic Search**  
Shuming Shi, Fei Xing, Mingjie Zhu, Zaiqing Niu, Ji-Rong Wen  
NLPIR4DL 2009 | Highly | Familiar |
| | Fairly | |
| | Marginally | |
| | Not Related | ✓ |
| | | |
| ABSTRACT PDF  
**Context-aware citation recommendation**  
Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, Lee Giles  
WWW 2010 | Highly | Familiar |
| | Fairly | |
| | Marginally | |
| | Not Related | ✓ |
| | | |

**Submit**
**STEP 2**

**Related Documents**

The table below lists the documents found to be related in alphabetical order by title. You can expand a document's abstract by clicking on **ABSTRACT**, and you can open the document as PDF file by clicking on **PDF**.

Please read the abstract of each paper and rate it regarding two criteria: relatedness and familiarity.

The **Level of Relatedness** indicates how related the document is with respect to your research task and input documents specified in STEP 1.

- **Highly** - The document matches my research task perfectly.
- **Fairly** - The document matches my research task.
- **Marginally** - The document includes only a few aspects of my research task.
- **Not Related** - The document does not match my research task in any respect.

The **Level of Familiarity** indicates whether you knew the document before this user study or not.

- **Familiar** - I knew the document before.
- **Unfamiliar** - I didn't know the document before.

<table>
<thead>
<tr>
<th>Document</th>
<th>Level of Relatedness</th>
<th>Level of Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor Text Extraction for Academic Search</td>
<td>Highly</td>
<td>Familiar</td>
</tr>
<tr>
<td>Shuming Shi, Fei Xing, Mingjie Zhu, Zaiqing Nie, Ji-Rong Wen</td>
<td>Fairly</td>
<td>Unfamiliar</td>
</tr>
<tr>
<td>NLPIR4DL 2009</td>
<td>Marginally</td>
<td></td>
</tr>
<tr>
<td>Context-aware citation recommendation</td>
<td>Highly</td>
<td>Familiar</td>
</tr>
<tr>
<td>Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, Lee Giles</td>
<td>Fairly</td>
<td>Unfamiliar</td>
</tr>
<tr>
<td>WWW 2010</td>
<td>Marginally</td>
<td></td>
</tr>
</tbody>
</table>

Submit
### Result Overlap

<table>
<thead>
<tr>
<th></th>
<th>Sofia Search</th>
<th>Nascimento</th>
<th>Google Scholar</th>
<th>Keyquery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sofia Search</td>
<td>100</td>
<td>26</td>
<td>27</td>
<td>55</td>
</tr>
<tr>
<td>Nascimento</td>
<td>26</td>
<td>100</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>27</td>
<td>16</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>Keyquery</td>
<td>55</td>
<td>25</td>
<td>30</td>
<td>100</td>
</tr>
</tbody>
</table>