

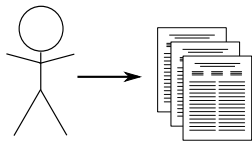
Keyquery-Based Recommendation of Related Work

Anna Beyer

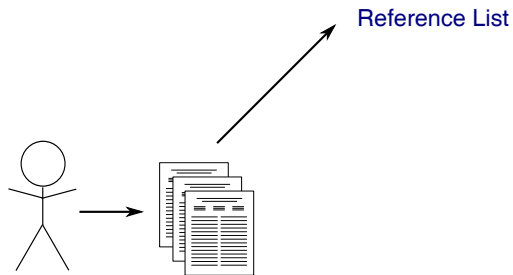
Bauhaus-Universität Weimar

Defense of Master's Thesis
2014-10-17

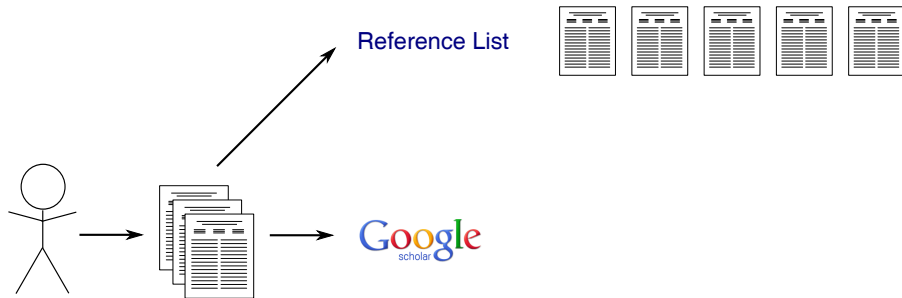
Motivation



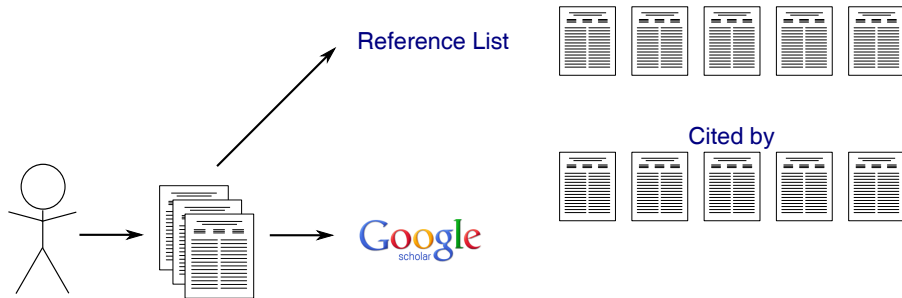
Motivation



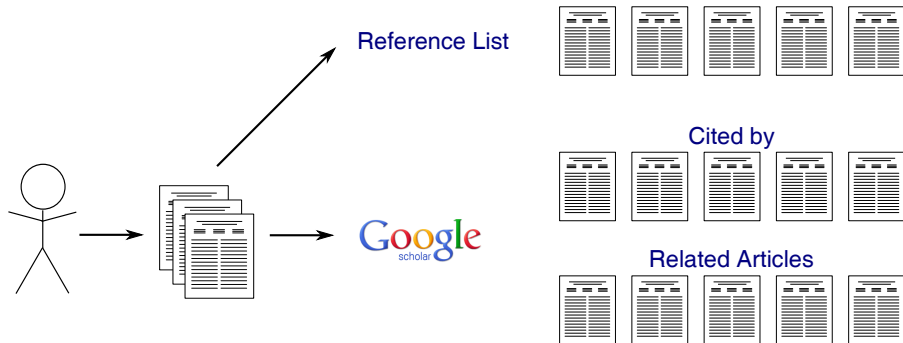
Motivation



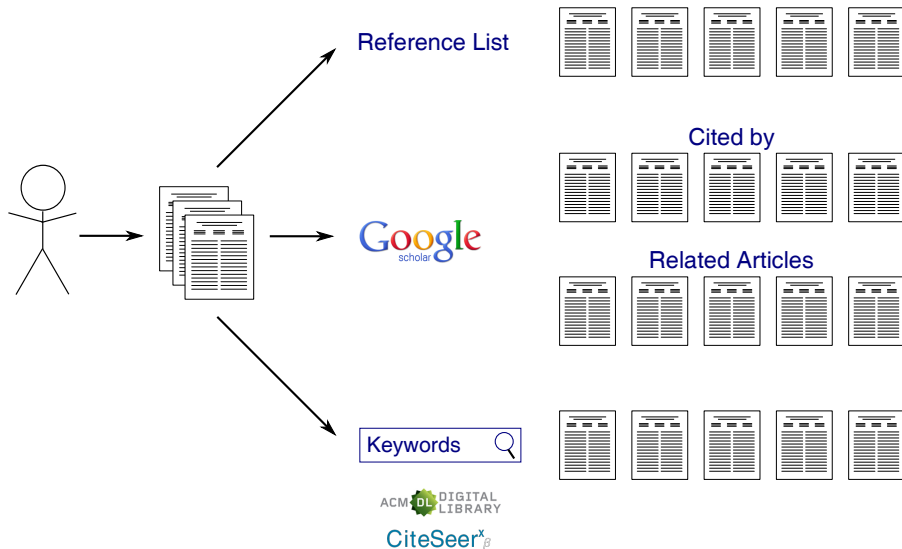
Motivation



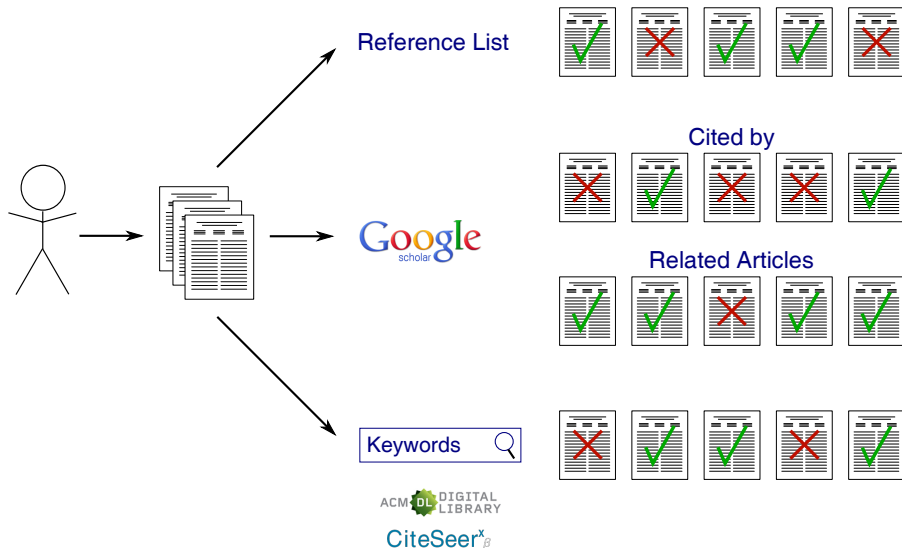
Motivation



Motivation



Motivation



Problem Definition

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

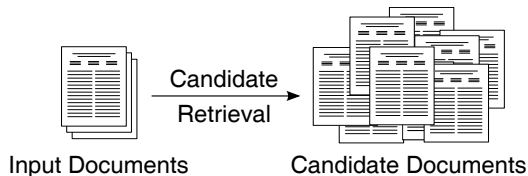
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

Basic Pipeline



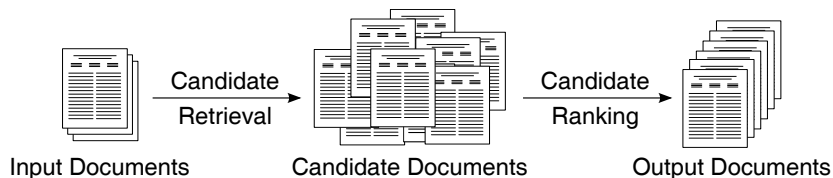
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

Basic Pipeline



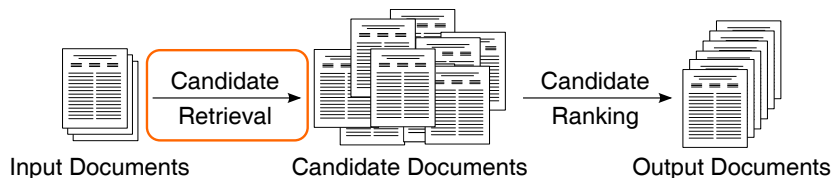
Problem Definition

Related Work Search

Given: A set of research papers.

Task: Find a set of topically similar research papers.

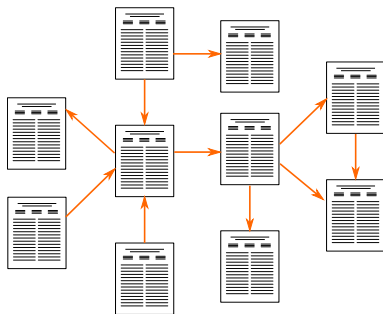
Basic Pipeline



Candidate Retrieval

Candidate Retrieval

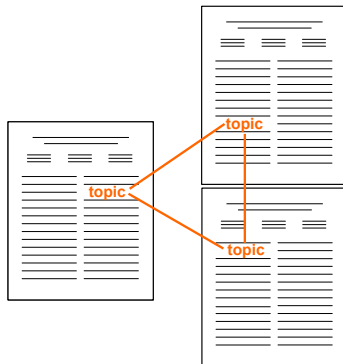
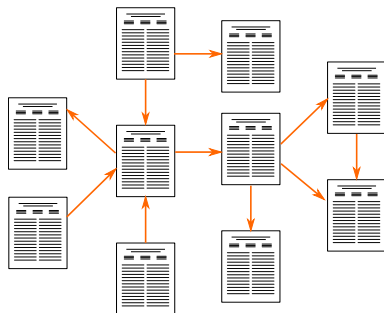
Citation Graph



Candidate Retrieval

Citation Graph

Content



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]

Citation Graph-Based Methods

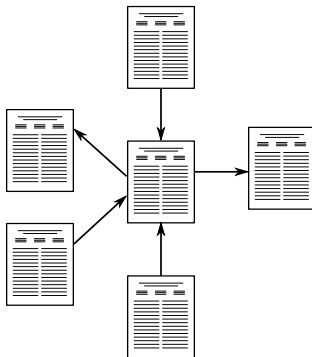
Simple Approaches	[Golshan et al., SIGMOD 2012]
Collaborative Filtering	[Sugiyama and Kan, JCDL 2013] [Caragea et al., JCDL 2013]
Link Ranking Algorithm	[Ekstrand et al., RecSys 2010] [Küçüktunç et al., JCDL 2013]

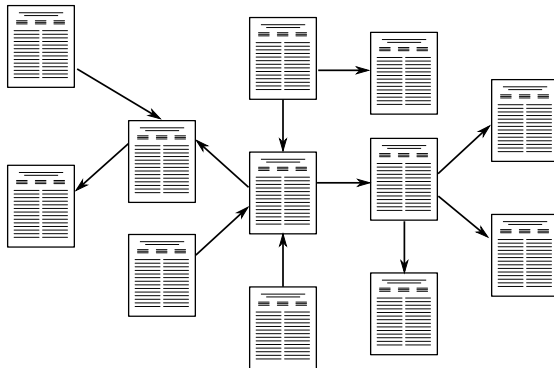
Content-Based Methods

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]









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.

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”

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”

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.

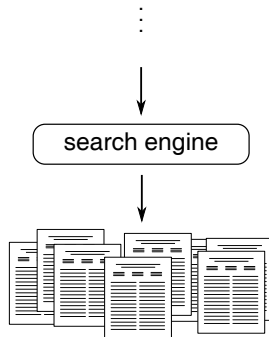
- ▶ “source independent”
- ▶ “independent framework”
- ▶ “framework research”

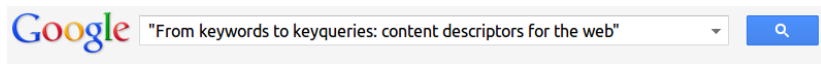
⋮

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.

- ▶ “source independent”
- ▶ “independent framework”
- ▶ “framework research”





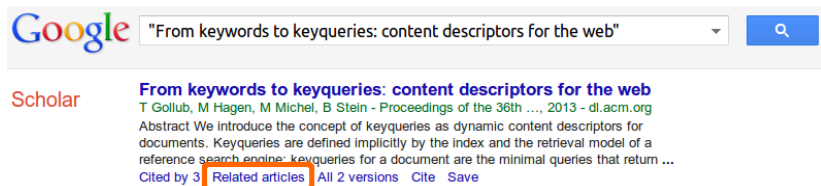
Scholar

From keywords to keyqueries: content descriptors for the web

T Gollub, M Hagen, M Michel, B Stein - *Proceedings of the 36th ...*, 2013 - dl.acm.org

Abstract We introduce the concept of keyqueries as dynamic content descriptors for documents. Keyqueries are defined implicitly by the index and the retrieval model of a reference search engine: keyqueries for a document are the minimal queries that return ...

[Cited by 3](#) [Related articles](#) [All 2 versions](#) [Cite](#) [Save](#)



The screenshot shows a Google Scholar search interface. At the top, the Google logo is on the left, followed by a search bar containing the text "From keywords to keyqueries: content descriptors for the web". To the right of the search bar is a blue button with a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. The search results show the title "From keywords to keyqueries: content descriptors for the web" in blue. Below the title, the authors "T Gollub, M Hagen, M Michel, B Stein" and the publication information "Proceedings of the 36th ..., 2013 - dl.acm.org" are listed. An abstract follows, starting with "Abstract We introduce the concept of keyqueries as dynamic content descriptors for documents. Keyqueries are defined implicitly by the index and the retrieval model of a reference search engine: keyqueries for a document are the minimal queries that return ...". At the bottom of the result, there are links: "Cited by 3", "Related articles" (which is highlighted with an orange box), "All 2 versions", "Cite", and "Save".

Google "From keywords to keyqueries: content descriptors for the web" 🔍

Scholar

From keywords to keyqueries: content descriptors for the web
T Gollub, M Hagen, M Michel, B Stein - Proceedings of the 36th ..., 2013 - dl.acm.org

Abstract We introduce the concept of keyqueries as dynamic content descriptors for documents. Keyqueries are defined implicitly by the index and the retrieval model of a reference search engine: keyqueries for a document are the minimal queries that return ...

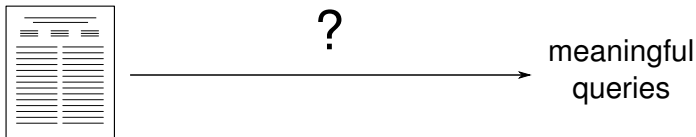
Cited by 3 **Related articles** All 2 versions Cite Save

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]



Overview

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





Scholar

Scholarly **paper recommendation** via user's recent **research interests**

[K Sugiyama](#), [MY 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**

[C Nascimento](#), [AHF Laender](#), [AS da Silva](#)... - Proceedings of the 11th ..., 2011 - dl.acm.org

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**

[SM McNee](#), [I 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

[A Naak](#), [H Hage](#), [E Aimeur](#) - E-Technologies: Innovation in an Open World, 2009 - Springer

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



research paper recommendation



Scholar

Scholarly **paper recommendation** via user's recent **research interests**

[K Sugiyama](#), [MY 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**

[C Nascimento](#), [AHF Laender](#), [AS da Silva](#)... - Proceedings of the 11th ..., 2011 - dl.acm.org

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**

[SM McNee](#), [I 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

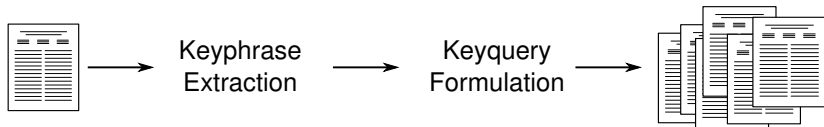
[A Naak](#), [H Hage](#), [E Aimeur](#) - E-Technologies: Innovation in an Open World, 2009 - Springer

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

Keyquery-Based Recommendation

Basic Pipeline



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]

paper recommendation query framework









⋮
 [Nascimento et al., JCDL 2011]







Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework



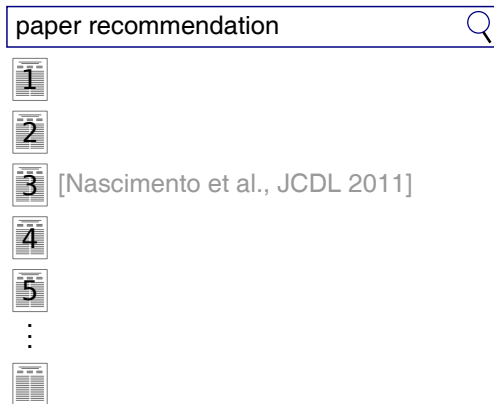


 [Nascimento et al., JCDL 2011]


⋮


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]

paper recommendation query framework

query

1

2

3

4

5

⋮

93

[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]

paper recommendation query framework paper

query framework paper 



[Nascimento et al., JCDL 2011]

⋮




Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework paper

query framework paper 



[Nascimento et al., JCDL 2011]

⋮

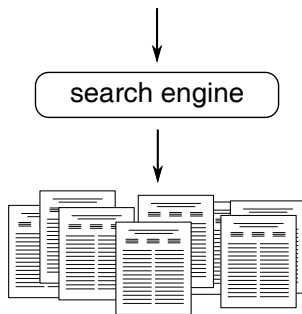


Keyquery-Based Recommendation

Keyquery Formulation

Example: [Nascimento et al., JCDL 2011]

paper recommendation query framework paper



Cranfield-Style Experiment

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

Cranfield-Style Experiment

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

Webis Computer Science Paper Corpus [187,000 papers]

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

Cranfield-Style Experiment

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

Webis Computer Science Paper Corpus [187,000 papers]

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

Cranfield-Style Experiment

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

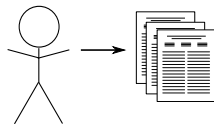
User Study [10 participants]



Cranfield-Style Experiment

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

User Study [10 participants]

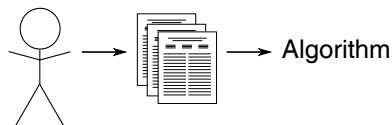


Experiments

Cranfield-Style Experiment

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

User Study [10 participants]

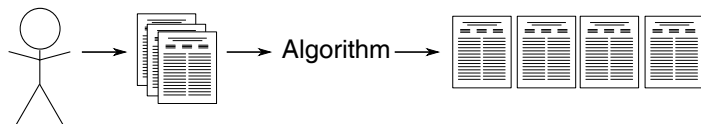


Experiments

Cranfield-Style Experiment

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

User Study [10 participants]

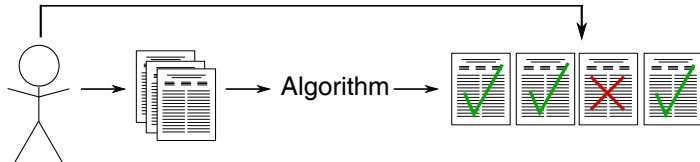


Experiments

Cranfield-Style Experiment

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

User Study [10 participants]

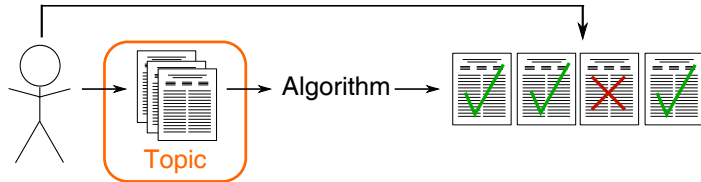


Experiments

Cranfield-Style Experiment

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

User Study [10 participants]

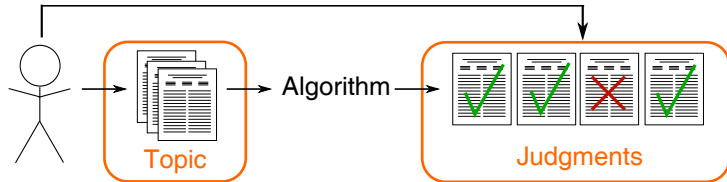


Experiments

Cranfield-Style Experiment

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

User Study [10 participants]



Algorithm	MAP@10
Sofia Search	0.546
Nascimento	0.523
Google Scholar	0.535
Keyquery	0.568
Keyquery+Google Scholar	0.605

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

Summary

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

Outlook

- ▶ Bigger user study
- ▶ Other datasets
- ▶ Keyquery formulation

Thank you!

User Study

STEP 1

Email

Please enter your email address.

anna.beyer@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

User Study

STEP 1

Email

Please enter your email address.

anna.beyer@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



Topic

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.

<i>Highly</i>	The document matches my research task perfectly.
<i>Fairly</i>	The document matches my research task.
<i>Marginally</i>	The document includes only a few aspects of my research task.
<i>Not Related</i>	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.

<i>Familiar</i>	I knew the document before.
<i>Unfamiliar</i>	I didn't know the document before.

Document	Level of Relatedness	Level of Familiarity
ABSTRACT PDF Anchor Text Extraction for Academic Search Shuming Shi, Fei Xing, Mingjie Zhu, Zaqing Nie, Ji-Rong Wen NLP4DL 2009	Highly Fairly Marginally Not Related ✓	Familiar Unfamiliar ✓
ABSTRACT PDF Context-aware citation recommendation Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, Lee Giles WWW 2010	Highly Fairly Marginally Not Related ✓	Familiar Unfamiliar ✓

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.

<i>Highly</i>	The document matches my research task perfectly.
<i>Fairly</i>	The document matches my research task.
<i>Marginally</i>	The document includes only a few aspects of my research task.
<i>Not Related</i>	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.

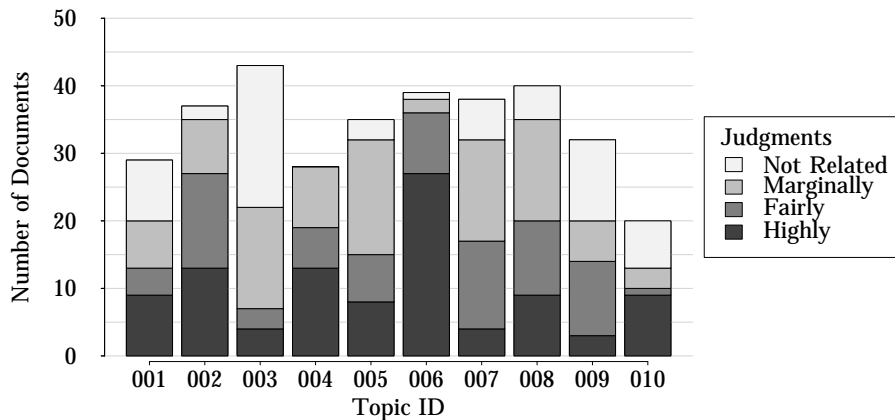
<i>Familiar</i>	I knew the document before.
<i>Unfamiliar</i>	I didn't know the document before.

Document	Level of Relatedness	Level of Familiarity
ABSTRACT PDF Anchor Text Extraction for Academic Search Shuming Shi, Fei Xing, Mingjie Zhu, Zaqing Nie, Ji-Rong Wen NLP4DL 2009	Highly Fairly Marginally Not Related ✓	Familiar Unfamiliar ✓
ABSTRACT PDF Context-aware citation recommendation Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, Lee Giles WWW 2010	Highly Fairly Marginally Not Related ✓	Familiar Unfamiliar ✓

Judgments

Submit

User Study



Result Overlap

	Sofia Search	Nascimento	Google Scholar	Keyquery
Sofia Search	100	26	27	55
Nascimento	26	100	16	25
Google Scholar	27	16	100	30
Keyquery	55	25	30	100