

Recommender System for Citations and Illustrations

Mandar Mitra

Indian Statistical Institute

Create tool to assist an author who is writing paper / creating presentation

- Provide bibliographic details of paper to be cited tools exist
- Suggest papers to cite techniques exist
- Suggest appropriate illustrations / images for presentation exploratory

What we have

File Edit Element Settings Help

Online Search

☐ Query Terms
 ☒ Engines

Search engines **Google Scholar**, **ACM Digital Library**, and **ScienceDirect** are selected. [Change](#)

Free test:

Title:

Author:

Year:

Number of Results:

Use Entry

Search Results

Type	Title or Book Title	Author or Editor	Year
Publication in Conference...	Cross-lingual Keyword Recommendation Using Latent Topics	Takasu, Atsuhiko	2010
Publication in Conference...	Combining Social Information for Academic Networking	Heck, Tamara	2013
Publication in Conference...	Research Paper Recommender System Evaluation: A Quantitative Literature	Beel, Joeran and Langer, Stefan a...	2013
Publication in Conference...	REFEREE: An Open Framework for Practical Testing of Recommender...	Cosley, Dan and Lawrence, Steve...	2002
Publication in Conference...	Automatically Building Research Reading Lists	Ekstrand, Michael D. and Kannan...	2010

Reference Preview

Source

```
#InCollection[caragea_citeSeer_2014,
author = "Caragea, Cornelia and Wu, Jian and Ciobanu, Alina and Williams, Kyle and Fernández-Ra
booktitle = "Advances in Information Retrieval (Retrieval)",
file = "ecir2014_caragea.pdf"/home/maimdar/.mozilla/firefox/shxot7d.default/zotero/storage/SEN
pages = "311-327",
publisher = "Springer",
shorttitle = "CiteSeer x",
x]
```

Type	Title or Book Title	Author or Editor	Year	DOI
Miscellaneous	A Discriminative Approach to Topic-Based Cit...	Jie Tang, Jing Zhang	2009	
Publication in...	A Language Modeling Approach to Informatio...	Ponte, Jay M. and C...	1998	10.114...
Publication in...	A source independent framework for researc...	Nascimento, Cristian...	2011	
Publication in...	A study of smoothing methods for language ...	Zhai, Chengxiong an...	2001	
Journal Article	Automated Learning of Decision Rules for Te...	Ayd, Chidanand an...	1994	
Publication in...	Can't see the forest for the trees?: a citation r...	Caragea, Cornelia a...	2013	
Part of a Book...	CiteSeer x: A Scholarly Big Dataset	Caragea, Cornelia a...	2014	
Publication in...	CiteSeer: An automatic citation indexing syst...	Giles, C. Lee and Bo...	1998	
Publication in...	CiteSpace: supporting contextual citation reco...	Line, Arvid and ...	2014	10.114...
Publication in...	CiteSeerX: AI in a digital library search engine	Wu, Jian and Willia...	2014	
Publication in...	Comparing citation contexts for information r...	Ritchie, Anna and R...	2008	
Publication in...	Context-aware citation recommendation	He, Qi and Pei, Jian ...	2010	
Publication in...	Creating a test collection for citation-based l...	Ritchie, Anna and Te...	2006	
Publication in...	Cumulative Citation Recommendation: A Fea...	Gebreneskkel, Gebre...	2014	10.110...
Publication in...	Distributed representations of words and phr...	Mikolov, Tomas and ...	2013	
Publication in...	Exploiting potential citation papers in scholar...	Sugiyama, Kazunari ...	2013	
Miscellaneous	GEISIS: Bibliometrics-enhanced Information ...			
Publication in...	On the recommending of citations for researc...	McNee, Sean M. an...	2002	
Part of a Book...	Personalized information access in a bibliogr...	Haase, Peter and Ehl...	2006	
Publication in...	REFEREE: An open framework for practical t...	Cosley, Dan and La...	2002	
Part of a Book...	Readings in Information Retrieval	Porter, M. F.	1997	
Publication in...	Recommending citations for academic papers	Strohman, Trevor an...	2007	
Publication in...	Recommending citations with translation model	Lu, Yang and He, Jin...	2011	
Publication in...	Recommending citations: translating papers I...	Huang, Wenyi and K...	2012	
Publication in...	Research Paper Recommender Systems: A...	Gel, M. and Pucco, A...	2006	10.110...
Journal Article	Research paper recommender systems - a li...	Beel, Joeran and Gl...	2015	
Miscellaneous	Reuters-21578 text categorization test collect...	Lewis, David D.	2004	
Miscellaneous	Scholarly Paper Recommendation Datasets (...			
Publication in...	Scholarly big data information extraction and...	Williams, Kyle and ...	2014	
Publication in...	SimSeerX: a similar document search engine	Williams, Kyle and ...	2014	10.114...
Journal Article	The Probabilistic Relevance Framework: BM...	Robertson, Stephen ...	2009	
Journal Article	The Scope and Importance of Human Interu...	McFarlane, Daniel C...	2002	10.120...
Publication in...	Towards building a scholarly big data platfo...	Wu, Zhongshu and W...	2014	
PHD Thesis	Using language models for information retrieval	Hienstra, D.	2001	
Part of a Book...	Using terms from citations for IR: some first r...	Ritchie, Anna and Te...	2008	

What we have

The screenshot shows the CiteSeer X search interface. A red oval labeled "Search interface" highlights the top section containing the search engine selection (Google Scholar, ACM Digital Library, ScienceDirect), the search query "bibliographic recommender", and the "Use Entry" button. Another red oval labeled "Search results" highlights the "Search Results" table, which lists search results with columns for Type, Title or Book Title, Author or Editor, Year, and DOI. A third red oval labeled "Current .bib file" highlights the "Reference Preview" window, which displays the BibTeX entry for the selected result: "CiteSeer x: A Scholarly Big Dataset".

Search interface

File Edit Element Settings Help
Online Search
Query Terms Engines
Search engines Google Scholar, ACM Digital Library, and ScienceDirect are selected. [Change](#)
Free text: bibliographic recommender
Title:
Author:
Year:
Number of Results: 5
Use Entry

Search Results

Type	Title or Book Title	Author or Editor	Year
Publication in Conference...	Cross-lingual Keyword Recommendation Using Latent Topics	Takasu, Atsuhiko	2010
Publication in Conference...	Combining Social Information for Academic Networking	Heck, Tamara	2013
Publication in Conference...	Research Paper Recommender System Evaluation: A Quantitative Literature Study	Beel, Joeran and Langer, Stefan A.	2013
Publication in Conference...	REFEREE: An Open Framework for Practical Testing of Recommender Systems	Cosley, Dan and Lawrence, Steve	2002
Publication in Conference...	Automatically Building Research Reading Lists	Ekstrand, Michael D. and Kannan, Arun	2010

Reference Preview

```
#inccollection[caraaga_citeSeer_2014,  
author = "Caraaga, Cornelia and Wu, Jian and Ciobanu, Alina and Williams, Kyle and Fernandez-Ra  
booktitle = "Advances in Information Retrieval (Retrieval)",  
file = "ecir2014_caraaga.pdf"/home/maider/.mozilla/firefox/shxot7d.default/zotero/storage/SEN  
pages = "311-322",  
publisher = "Springer",  
shorttitle = "CiteSeer x",  
url = "http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.1.1.1"]
```

Current .bib file

Type	Title or Book Title	Author or Editor	Year	DOI
Miscellaneous	A Discriminative Approach to Topic-Based Citation Recommendation	Jie Tang, Jing Zhang	2009	
Publication in Conference...	A Language Modeling Approach to Information Retrieval	Ponte, Jay M. and Croft, W. Bruce	1998	10.1145/289541.289542
Publication in Conference...	A source independent framework for research paper recommendation	Nascimento, Cristian	2011	
Publication in Conference...	A study of smoothing methods for language modeling	Zhai, Chengxiong and Lafferty, John D.	2001	
Journal Article	Automated Learning of Decision Rules for Topic Classification	Apostolakis, Christos and Elkan, Charles	1994	
Publication in Conference...	Can't see the forest for the trees?: a citation recommendation framework	Caraaga, Cornelia and Wu, Jian	2013	
Part of a Book...	CiteSeer x: A Scholarly Big Dataset	Caraaga, Cornelia and Wu, Jian	2014	
Publication in Conference...	CiteSeer: An automatic citation indexing system	Giles, C. Lee and Greisdorf, David	1998	
Publication in Conference...	CiteSpace: supporting contextual citation recommendation	Lim, Arshay and Lim, Arshay	2014	10.1145/259541.259542
Publication in Conference...	CiteSeer: AI in a digital library search engine	Wu, Jian and Williams, Kyle	2014	
Publication in Conference...	Comparing citation contexts for information retrieval	Ritchie, Anna and Rasmussen, David	2008	
Publication in Conference...	Context-aware citation recommendation	He, Qi and Pei, Jian	2010	
Publication in Conference...	Creating a test collection for citation-based information retrieval	Ritchie, Anna and Rasmussen, David	2006	
Publication in Conference...	Creating a test collection: Relevance judgement	Ritchie, Anna and Rasmussen, David	2007	
Publication in Conference...	Cumulative Citation Recommendation: A Feature-based Approach	Gebru, Mervin and Greisdorf, David	2014	10.1145/259541.259542
Publication in Conference...	Distributed representations of words and phrases for information retrieval	Mikolov, Tomas and Yih, Wen-Cheng	2013	
Miscellaneous	GEISIS: Bibliometrics-enhanced information science	Suygama, Kazumasa	2013	
Publication in Conference...	On the recommending of citations for research papers	McNeel, Sean M. and Greisdorf, David	2002	
Part of a Book...	Personalized information retrieval	Greisdorf, David and Greisdorf, David	2002	
Publication in Conference...	REFEREE: An Open Framework for Practical Testing of Recommender Systems	Cosley, Dan and Lawrence, Steve	2002	
Part of a Book...	Research Paper Recommender System Evaluation: A Quantitative Literature Study	Beel, Joeran and Langer, Stefan A.	2013	
Publication in Conference...	Recommending citations for research papers	McNeel, Sean M. and Greisdorf, David	2007	
Publication in Conference...	Recommending citations for research papers	McNeel, Sean M. and Greisdorf, David	2011	
Publication in Conference...	Research Paper Recommender System Evaluation: A Quantitative Literature Study	Beel, Joeran and Langer, Stefan A.	2013	10.1145/259541.259542
Journal Article	Research paper recommendation using a probabilistic relevance framework	Beel, Joeran and Langer, Stefan A.	2015	
Miscellaneous	Research paper recommendation using a probabilistic relevance framework	Beel, Joeran and Langer, Stefan A.	2015	
Miscellaneous	Research Paper Recommender System Evaluation: A Quantitative Literature Study	Beel, Joeran and Langer, Stefan A.	2013	
Publication in Conference...	Scholarly paper recommendation using a probabilistic relevance framework	Beel, Joeran and Langer, Stefan A.	2014	
Publication in Conference...	SimSeerX: a similarity-based search engine	Williams, Kyle and Wu, Jian	2014	10.1145/259541.259542
Journal Article	The Scope and Importance of Human Information Retrieval	McFarlane, Daniel C.	2002	10.1201/10.1201/10.1201
Publication in Conference...	Towards building a scholarly big data platform	Wu, Jian and Williams, Kyle	2014	
PhD Thesis	Using language models for information retrieval	Hemstra, D.	2001	
Part of a Book...	Using terms from citations for IR: some first steps	Ritchie, Anna and Rasmussen, David	2008	

What we want

File Edit Element Settings Help

Online Search

☐ Query Terms ☒ Engines

Search engines [Google Scholar](#), [ACM Digital Library](#), and [ScienceDirect](#) are selected. [Change](#)

Free text:

Title:

Author:

Year:

Number of Results:

[Use Entry](#)

[Search](#)

Search Results

Type	Title or Book Title	Author or Editor	Year
Publication in Conference...	Cross-lingual Keyword Recommendation Using Latent Topics	Takasu, Atsuhiko	2010
Publication in Conference...	Combining Social Information for Academic Networking	Heck, Tamara	2013
Publication in Conference...	Research Paper Recommender System Evaluation: A Quantitative Literature Study	Bell, Jordan and Lopez-Solano, ...	2013
Publication in Conference...	REFREE: An Open Framework for Practical Testing of Recommendation Systems	Cosley, Dan and Lawrence, Steve...	2002
Publication in Conference...	Automatically Building Research Reading Lists	Chen, Michael D. and Karmann...	2010

Search interface

Search results

Recommended citations
(you may also be interested in:)

Reference Preview

Source

```
#inccollection{caragea_citeSeer_2014,  
author = "Caragea, Cornelia and Wu, Jian and Ciobanu, Alina and Williams, Kyle and Fern\u00e1ndez-Sa  
booktitle = "[Advances in Information] (Retrieval)",  
file = "ecir2014_caragea.pdf/home/mandar/.mozilla/firefox/shzot7d.default/zotero/storage/8EV  
pages = "331-322",  
publisher = "Springer",  
shorttitle = "[CiteSeer] x",  
}
```

[Open](#) [Save as HTML](#)

Type	Title or Book Title	Author or Editor	Year	DOI
Miscellaneous	A Discriminative Approach to Topic-Based Citations	Jie Tang, Jing Zhang	2009	
Publication in Conference...	A Language Modeling Approach to Information Retrieval	Ponne, Jay M. and C...	1998	10.114...
Publication in Conference...	A source independent framework for research paper recommendation	Nascimento, Cristian	2011	
Publication in Conference...	A study of smoothing techniques for language modeling	Zhai, Chengxian an...	2001	
Journal Article	Automated Learning of Decision Rules for Text Classification	Apt\u00e9, Chidanand an...	1994	
Publication in Conference...	Can't see the forest for the trees? a citation recommendation system	Caragea, Cornelia a...	2013	
Publication in Conference...	CiteSeer: An automatic citation indexing system	Giles, C. Lee and Bo...	1998	
Publication in Conference...	CiteSight: supporting contextual citation recommendation	Livne, Avishay and ...	2014	10.114...
Publication in Conference...	CiteSeer: AI in a digital library search engine	Wu, Jian and Willia...	2014	
Publication in Conference...	Comparing citation contexts for information retrieval	Ritchie, Anna and R...	2008	
Publication in Conference...	Context-aware citation recommendation	He, Q. and Pei, Jian	2010	
Publication in Conference...	Creating a test collection for citation-based information retrieval	Ritchie, Anna and R...	2006	
Publication in Conference...	Creating a test collection: Relevance judgement	Ritchie, Anna and R...	2007	
Publication in Conference...	Cumulative Citation Recommendation: A Feature-based Approach	Gebremeskel, Gebre...	2014	10.110...
Publication in Conference...	Distributed representations of words and phrases	Mikolov, Tomas and ...	2013	
Publication in Conference...	Exploiting potential citation papers in scholarly research	Sugiyama, Kazunari	2013	
Miscellaneous	GEISIS: Bibliometrics-enhanced information retrieval	McNee, Sean M. an...	2002	
Publication in Conference...	On the recommending of citations for research	McNee, Sean M. an...	2006	
Part of a Book...	Personalized information access in academic research	Haase, Peter and Eh...	2006	
Publication in Conference...	REFREE: An open framework for practical testing of recommendation systems	Cosley, Dan and La...	2002	
Part of a Book...	Readings in Information Retrieval	Porter, M. F.	1997	
Publication in Conference...	Recommending citations for academic papers	Strohman, Trevor an...	2007	
Publication in Conference...	Recommending citations with topic-based models	Wang, Hui and He, J...	2011	
Publication in Conference...	Recommending citations: The role of citation context	Ruiz, Wenyi and K...	2012	
Publication in Conference...	Research Paper Recommender Systems: A Survey	Gori, M. and Pucci, A.	2006	10.110...
Journal Article	Research paper recommender systems - a survey	Peel, Joeran and Gl...	2015	
Miscellaneous	Research paper recommendation systems	David D.L.	2004	
Miscellaneous	Scholarly big data information extraction and analysis	Williams, Kyle and ...	2014	
Publication in Conference...	SimSeer: a similar document search engine	Williams, Kyle and ...	2014	10.114...
Journal Article	The Probabilistic Relevance Framework: SimSeer	Hobson, Stephen...	2009	
Journal Article	The Scope and Importance of Human Interactions in Scholarly Research	Hobson, Stephen...	2010	10.120...
Publication in Conference...	Towards building a scholarly big data platform	Wu, Zhaohui and W...	2014	
PHD Thesis	Using language models for information retrieval	Hienstra, D.	2001	
Part of a Book...	Using terms from citations for IR: some first steps	Ritchie, Anna and R...	2008	

Proposed work plan

- Year I:
- Literature survey (update to BEEL ET AL., 2016)
 - Select methods to compare empirically
 - Expand existing dataset

- Year II:
- Implementation and evaluation of methods identified above
 - Design of editor-agnostic recommendation service

- Year III:
- Implementation of service
 - Creation of prototype plugin
 - Debugging and testing

- Automated story illustration dataset
 - ImageCLEF 2010 data set (237,434 different images + metadata)
 - 27 stories taken from Aesop's Fable
- Wikipedia 2016 dataset
 - Significantly larger, more heterogeneous (selected from 12+ million pages)
 - Queries = 300 Wikipedia articles
 - Each selected article has between 10 and 30 images

Overview of the automated story illustration task.

Debasis Ganguly, I. Calixto, and Gareth Jones. In Proc. FIRE Workshop, 2015.

- Survey of existing test collections
- Creation of new test collection based on CiteSeer / CiteSeer^X
 - CiteSeer^X: document collection, with cleaner, more reliable metadata
 - Our collection: 600,000 research papers + 2,500 queries
- Standard techniques
 - Collaborative filtering approaches: user \times item matrix
 - Content based methods (widely used)
 - *Reference Directed Indexing*: use “anchor text” to index an article

From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation.
Dwaipayan Roy, Kunal Ray, and Mandar Mitra. In AAAI Workshop on Scholarly Big Data, 2016.

An Improved Test Collection and Baselines for Bibliographic Citation Recommendation.
Dwaipayan Roy. Proc. ACM CIKM, 2017.