126909021 篇文章和 114698044 个作者， 24024 个机构和约 5.29 亿

**Ranking + profiling**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Ranking | Profiling | Relevance Ranking | Search engine | storage |
| Google Scholar | Mainly based on citation | 具有作者画像；**缺乏**关键字、机构、刊会画像 | Not found | Not found | Not found |
| Microsoft Academic | Not found | 具有作者，关键字，机构，刊会画像。但是**缺乏**可视化展示 | Not found | Not found | Azure Data Lake(hadoop) |
| Aminer | Random walk | 具有作者、关键字画像。但是**缺乏**刊会、机构画像 | LDA + LM | Lucene  (Speculate) | Mysql |
| Semantic Scholar | Citation Velocity | 具有作者、关键字画像。但是**缺乏**刊会、机构画像 | Tf\*idf | Elastic Search | RDBMS  (Speculate) |
| Acemap | citation | 具有作者，关键字，机构，刊会画像。 | Tf\*idf | Solr | Spark |
| citeseer | weighted pagerank for articles | 具有作者、关键字画像。但是**缺乏**刊会、机构画像 | Tf\*idf | Solr | Mysql |
| SARank | Heterogenous entity ranking | 具有作者，关键字，机构，刊会画像。 |  | Lucene | Graph database |

**系统提供的功能和对比**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 检索功能 | | | | 排名功能 | | | | 作者画像 | | 论文画像 | | 机构画像 | 刊会画像 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Google Scholar | **√** | **√** | **×** | **√** | **√** | **√** | **×** | **×** | **√** | **×** | **√** | **×** | **×** | **×** |
| Microsoft Academic | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **×** | **√** | **×** | **√** | **√** |
| Aminer | **√** | **√** | **×** | **×** | **√** | **√** | **×** | **×** | **√** | **√** | **√** | **×** | **×** | **×** |
| Semantic Scholar | **√** | **√** | **×** | **×** | **√** | **√** | **×** | **×** | **√** | **×** | **√** | **×** | **×** | **×** |
| Acemap | **√** | **√** | **√** | **√** | **√** | **√** | **×** | **×** | **√** | **×** | **√** | **√** | **√** | **√** |
| citeseer | **√** | **√** | **×** | **×** | **√** | **√** | **×** | **×** | **√** | **×** | **√** | **×** | **×** | **×** |
| SARank | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** | **√** |

1 ->检索关键字

2 ->检索作者

3 ->检索刊会

4 ->检索机构

5->论文排名

6->作者排名

7->刊会排名

8->机构排名

9 –>作者基本信息描述

10->作者深度画像（包括研究领域演变，研究兴趣词云图，可视化等）

11->论文基本信息展示 （包括该文章作者，出版刊会，出版时间）

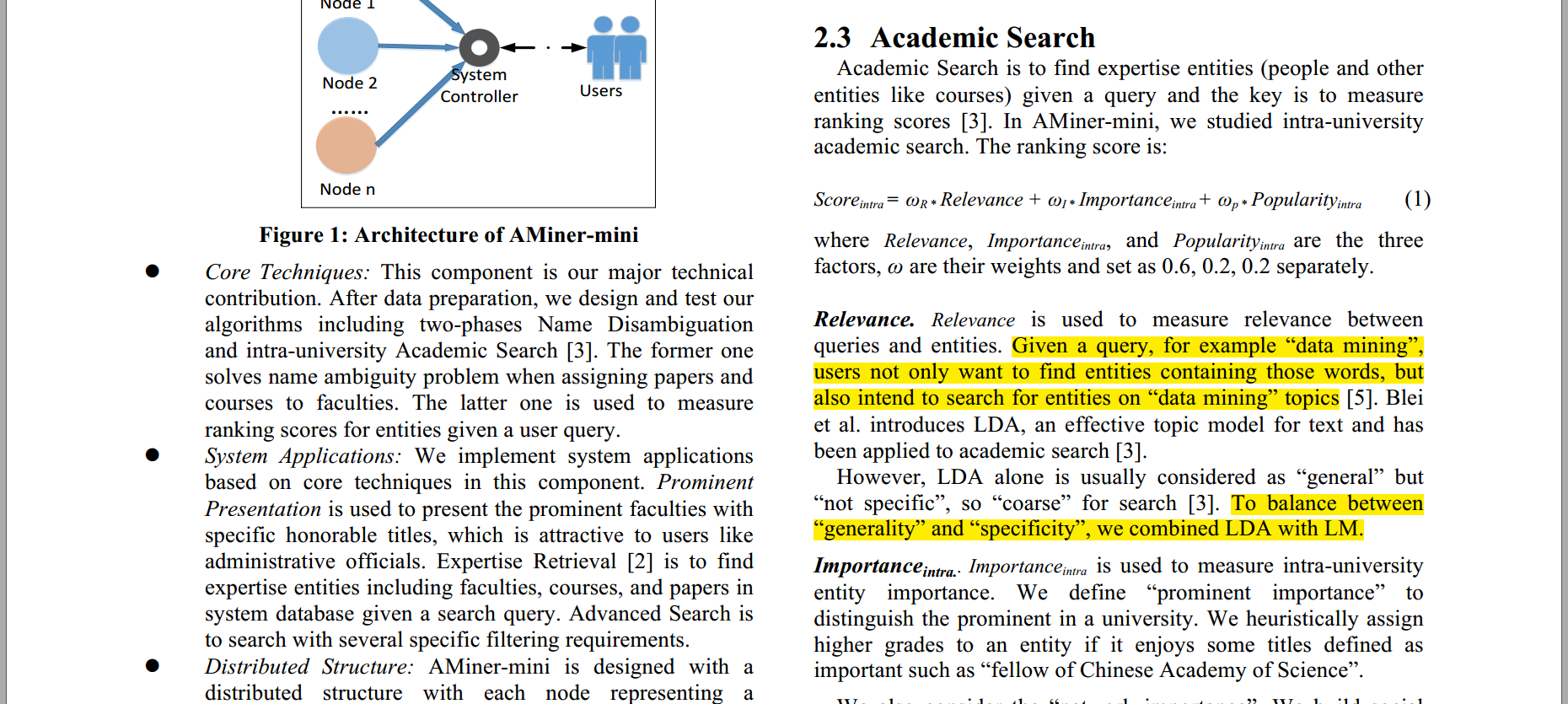
12->论文引用和参考文献

**参考文献**

Aminer academic search: combine LDA with LM

AMiner-mini: A People Search Engine for University

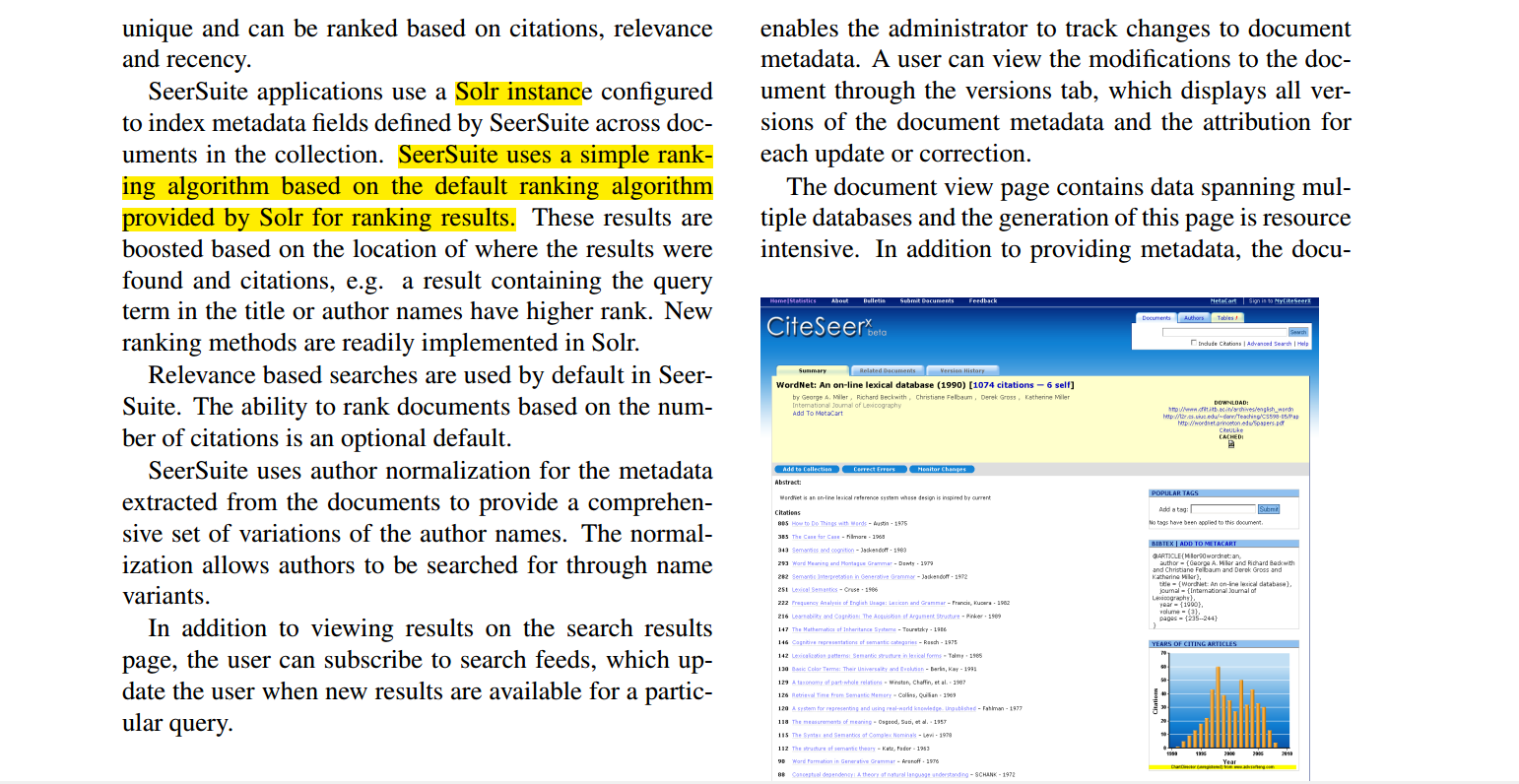
Jingyuan Liu\*, Debing Liu\*, Xingyu Yan\*, Li Dong#, Yutao Zhang\*, and Jie Tang



Citeseerx中使用全文检索引擎Solr，并且citeseer中使用solr中提供的默认排名算法。（tf\*idf模型）

SeerSuite: Developing a Scalable and Reliable Application Framework for Building Digital Libraries by Crawling the Web

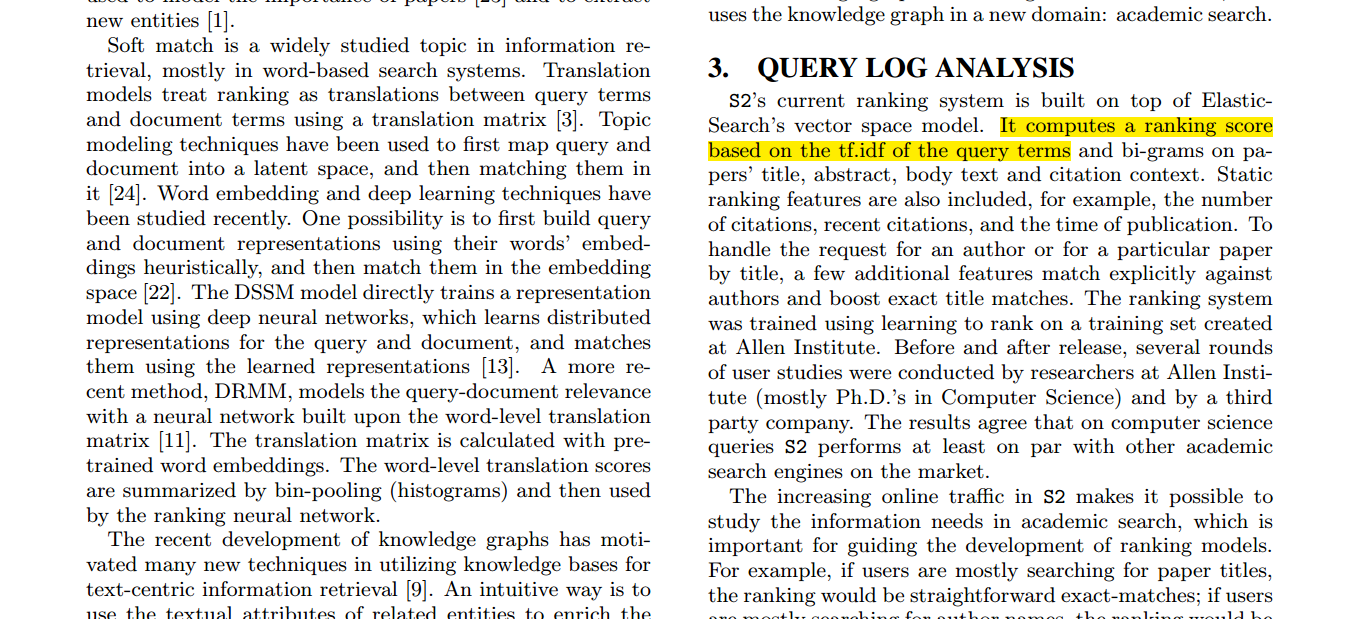
Pradeep B. Teregowda, Isaac G. Councill, ect



Semantic scholar relevance (tf\*idf)

Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding

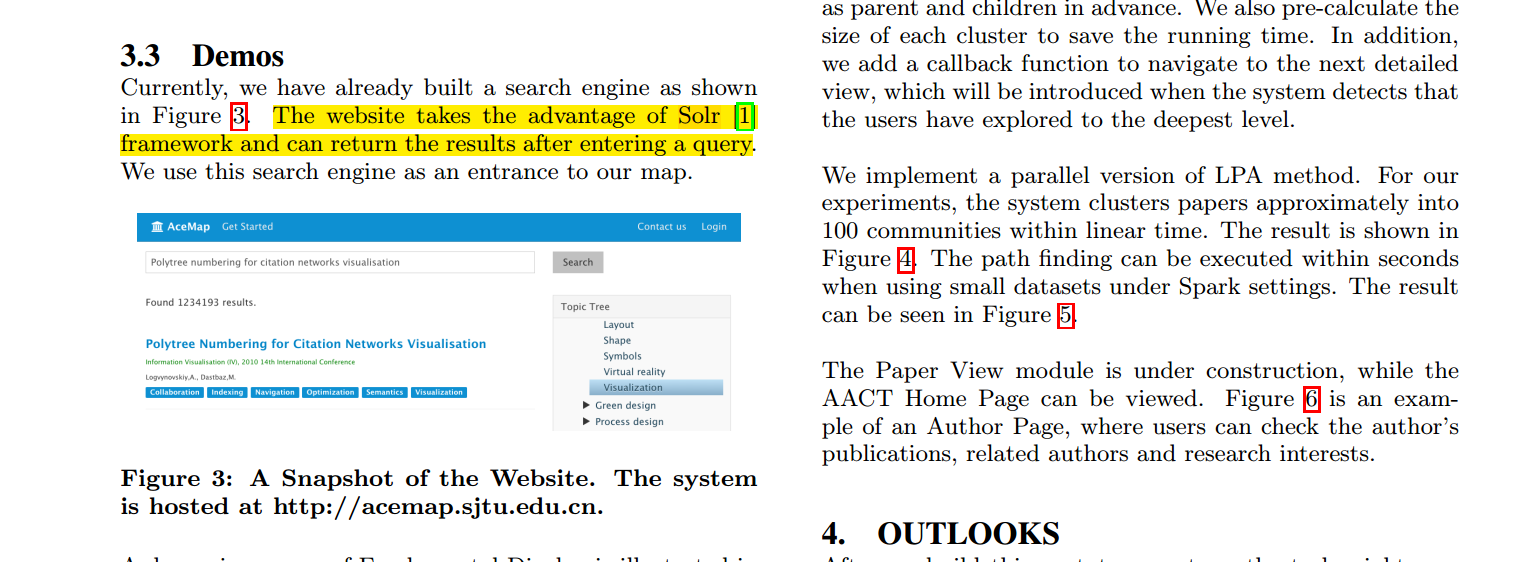
Chenyan Xiong, Russell Power, Jamie Callan



Acemap使用的solr搜索引擎，solr中相关性计算是tf\*idf，推测Acemap使用的是tf\*idf

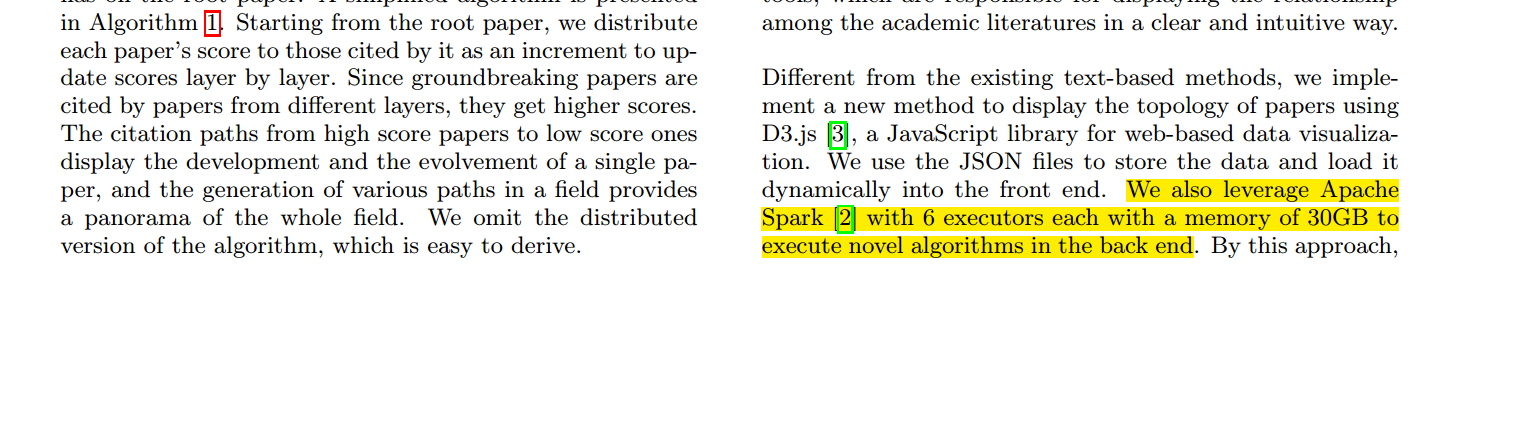
AceMap: A Novel Approach towards Displaying Relationship among Academic Literatures

Zhaowei Tan Changfeng Liu Yuning Mao Yunqi Guo Jiaming Shen Xinbing Wang



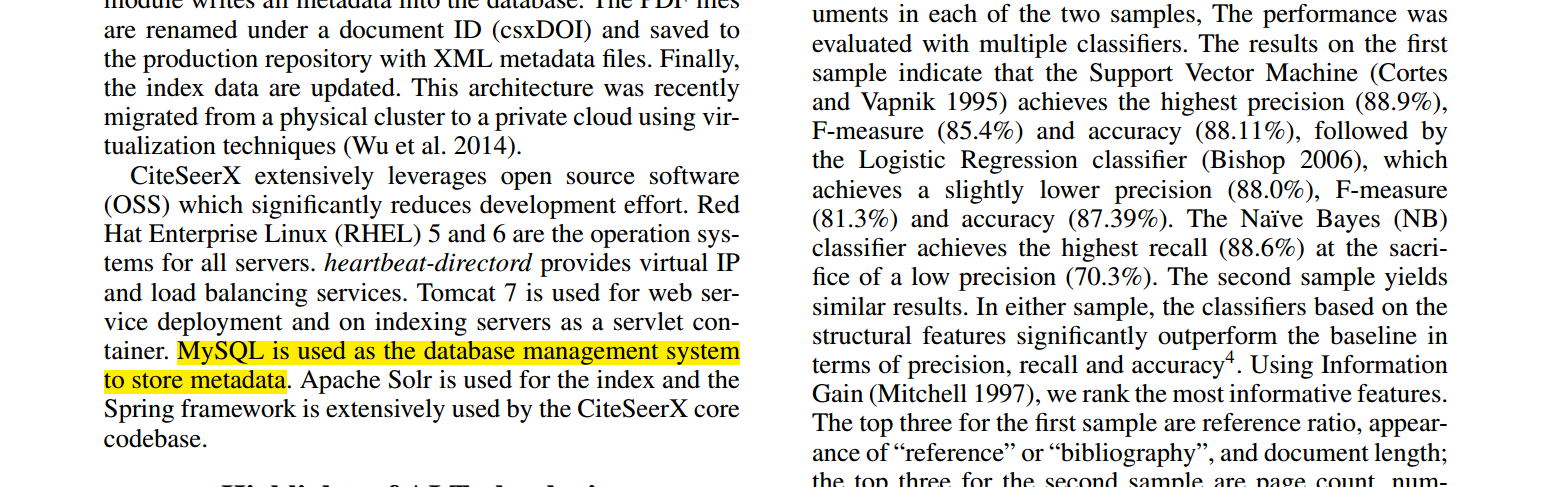
Acemap storage

AceMap: A Novel Approach towards Displaying Relationship among Academic Literatures



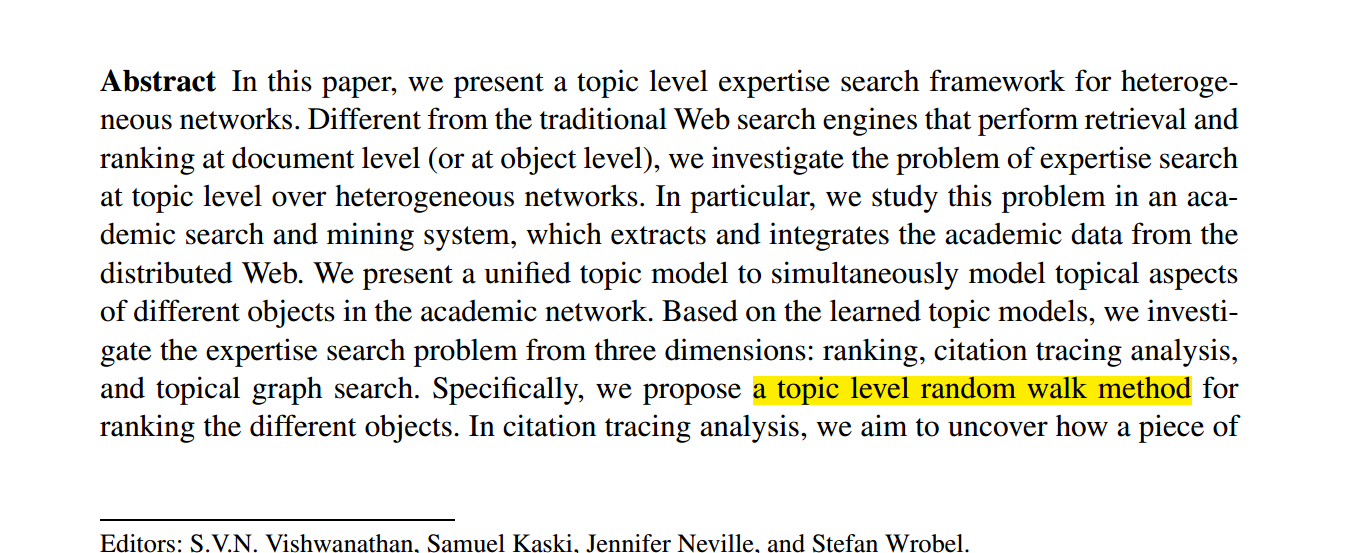
Citeseerx storage

AAAI-2014 citeseex ai in digital library search engine



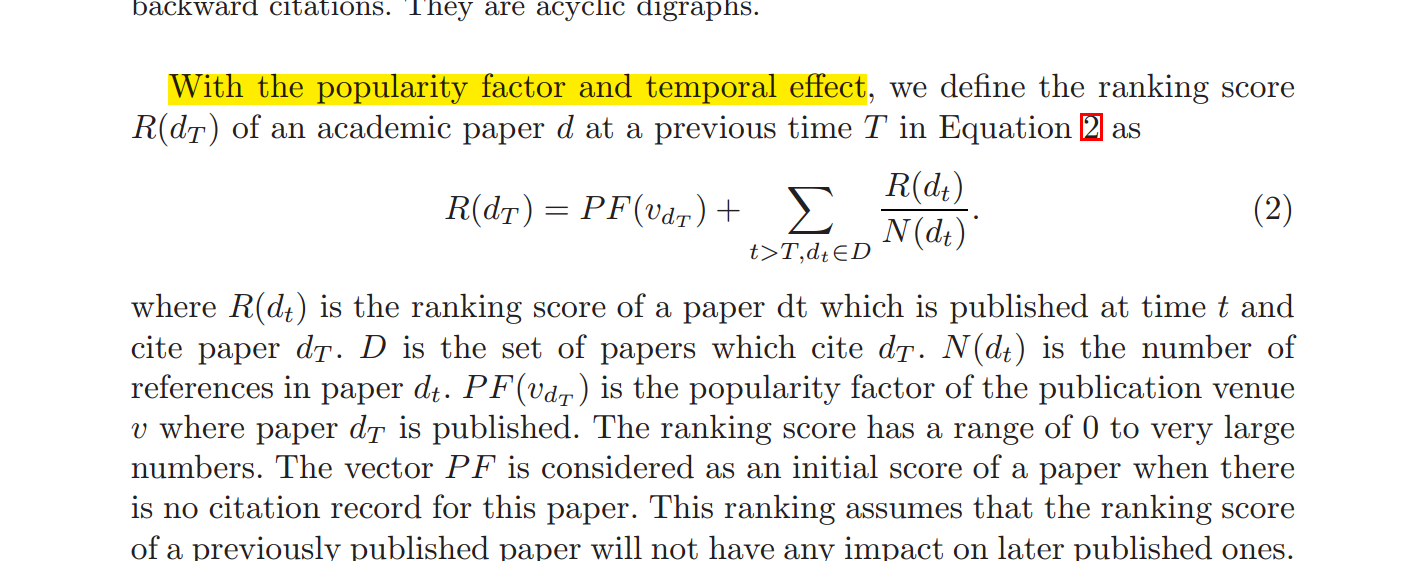
Aminer, ranking algorithm: random walk

Topic level expertise search over heterogeneous networks



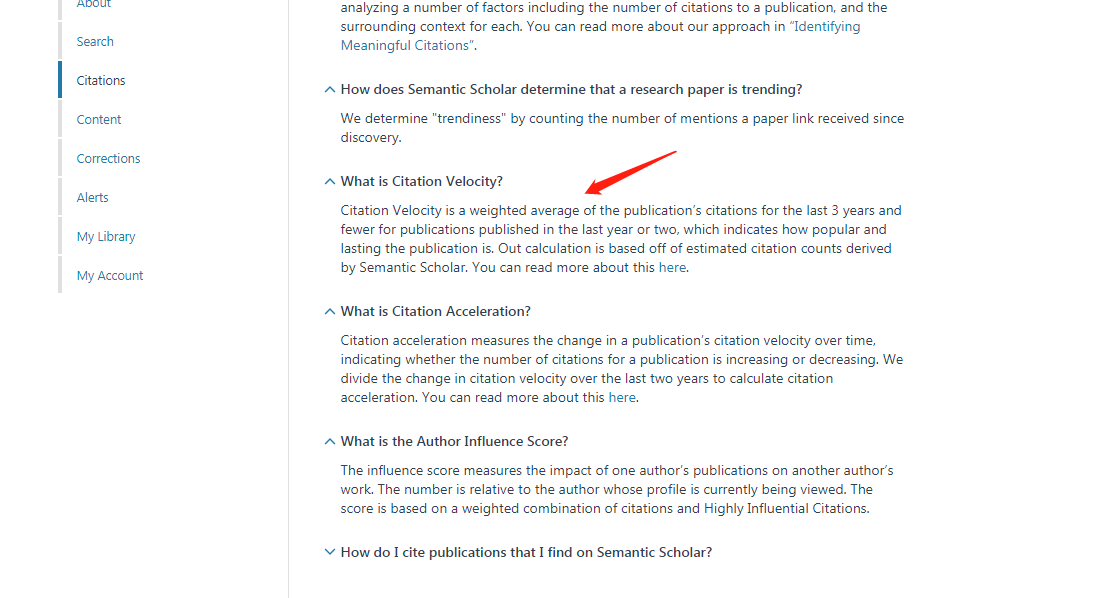
Citeseer: ranking article weighted PageRank

Popularity Weighted Ranking for Academic Digital Libraries



Semantic, ranking articles citation velocity

https://www.semanticscholar.org/faq#trending-papers



本系统中，ranking和profiling的内容

Ranking：

检索页面：对检索出的文章，使用SARank，Relevance Ranking, Time ranking, Citation Ranking。对于右侧Author Conference Journal Affiliation信息使用SARank排名Top-k；检索的同名作者按照作者的SARank分数排名；检索的机构排名按照机构的SARank分数

作者页面：文章按照SARank分数排。Co-author按照合作的文章数量排名，Affiliation按照时间排

机构页面：该机构的文章按照SARank，Time Ranking，Citation Ranking排名，与该机构相关的作者，机构，期刊，会议等按照SARank排Top-k

刊会页面：该刊会相关的文章按照SARank，Time Ranking，Citation Ranking排名，与该刊会相关的作者，机构，期刊，会议等按照SARank排Top-k

Profiling

关键字的画像：该关键字下Top-k排名的文章；该关键字的时间信息；相关的Author，Conference，Journal，Affiliation信息

作者画像：作者研究领域词云画像；作者研究兴趣随时间的变化；作者Top-k的Co-author信息；作者的历史机构信息；作者文章的统计信息（引用信息，文章信息，引用随时间的变化信息等）

机构画像：该机构Top-k排名的文章；该机构的时间信息；该机构Top-k相关的Author，Conference，Journal，Affiliation信息

刊会画像：该刊会Top-k排名的文章；该刊会的时间信息；该刊会的Top-k相关的Author，Conference，Journal，Affiliation信息。