

大数据知识管理服务平台技术与实践

杜一

中国科学院计算机网络信息中心 大数据知识工程实验室

2020年6月2日

学术搜索

Search any topic, author, journal, etc. or any combination of these



Google Scholar



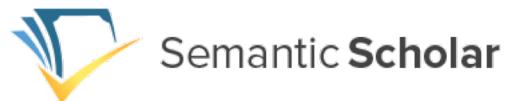
Baidu 学术

学术搜索

Search any topic, author, journal, etc. or any combination of these



Google Scholar



学术搜索

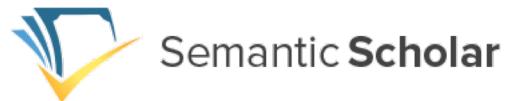
Search any topic, author, journal, etc. or any combination of these



Google Scholar



Baidu 学术



Academia.edu
share research

AMiner



WEB OF SCIENCE

Scopus®

万方数据
WANFANG DATA

CnKI 中国知网
www.cnki.net
中国知识基础设施工程

Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding 论文结论

Chenyan Xiong, Russell Power, James P. Callan · Published in WWW 2017 · Computer Science ·

Proceedings of the 26th International Conference on World Wide Web

This paper introduces Explicit Semantic Ranking (ESR), a new ranking technique that leverages knowledge graph embedding. [...] Experiments demonstrate ESR's ability in improving Semantic Scholar's online production system, especially on hard queries where word-based ranking fails. [\[Expand Abstract\]](#)

[View On ACM](#)

[Alternate Sources](#)

[In Library](#)

[Create Alert](#)

[Cite](#)

[In Feed](#)



Semantic Scholar

87 Citations

5 Highly Influenced Papers

64 Cite Background

20 Cite Methods

2 Cite Results

引文分类

ABSTRACT	FIGURES, TABLES, AND TOPICS	87 CITATIONS	7 REFERENCES	RELATED PAPERS
----------	-----------------------------	--------------	--------------	----------------

Figures, Tables, and Topics from this paper.

论文主题

Figures and Tables

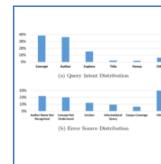


Figure 1

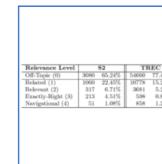


Table 1

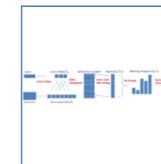


Figure 2

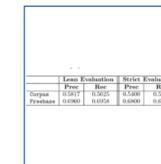


Table 2

[SHOW MORE \(4\)](#)

Explore Further: Topics Discussed in This Paper

- Knowledge Graph
- Graph embedding
- Academic Search
- Semantic Scholar
- Production system (computer science)
- Web search engine

Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding 论文结论

Chenyan Xiong, Russell Power, James P. Callan · Published in WWW 2017 · Computer Science · Proceedings of the 26th International Conference on World Wide Web

This paper introduces Explicit Semantic Ranking (ESR), a new ranking technique that leverages knowledge graph embedding. [...] Experiments demonstrate ESR's ability in improving Semantic Scholar's online production system, especially on hard queries where word-based ranking fails. [\[Expand Abstract\]](#)

[View On ACM](#)

[Alternate Sources](#)

[In Library](#)

[Create Alert](#)

[Cite](#)

[In Feed](#)



Semantic Scholar

87 Citations

5 Highly Influenced Papers

64 Cite Background

20 Cite Methods

2 Cite Results

引文分类

ABSTRACT FIGURES, TABLES, AND TOPICS 87 CITATIONS 7 REFERENCES RELATED PAPERS

Figures, Tables, and Topics from this paper.

论文主题

Figures and Tables

[2020 ICLR 2020] LambdaNet: Probabilistic Type Inference using Graph Neural Networks

Neural Network

[2019 Arxiv] Predicting Variable Types in Dynamically Typed Programming Languages

[2017 NIPS] Premise Selection for Theorem Proving by Deep Graph Embedding

[2016 FSE 2016] Python probabilistic type inference with natural language support

[2014 ICLR] Adam: A Method for Stochastic Optimization

[2014 AAAI] Convolutional Neural Networks over Tree Structures for Programming Language Processing

[2013 Empirical Software Engineering] An empirical study on the impact of static typing on software maintainability

[2011 APLAS] Extending Hindley-Milner Type Inference with Coercive Structural Subtyping
[2007 ECOOP] Gradual Typing for Objects
[2004 POPL '04] Principal typings for Java-like languages
[1998 POPL '98] Local type inference

Deep Learning

[2018 ESEC/FSE 2018] Deep learning type inference

[2016 ACL] Pointing the Unknown Words

[2016 KDD '16] Label Noise Reduction in Entity Typing by Heterogeneous Partial-Label Embedding

[2013 ICLR 2014] Zero-Shot Learning for Semantic Utterance Classification

Local Context

[2017 CoNLL] Named Entity Disambiguation for Noisy Text

Explore Further: Topics Discussed in This Paper

- Knowledge Graph
- Graph embedding
- Academic Search
- Semantic Scholar
- Production system (computer science)
- Web search engine

Zero-Shot
[2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition] Zero-Shot Recognition via Semantic Embeddings and Knowledge Graphs

Source Code
[2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)] NL2Type: Inferring JavaScript Function Types from Natural Language Information

[2017 ICLR] Learning to Represent Programs with Graphs

Type System
[2018 ECOOP] Kafka: Gradual Typing for Objects

[2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)] To Type or Not to Type: Quantifying Detectable Bugs in JavaScript

[2016 FSE 2016] Python probabilistic type inference with natural language support

[2014 ECOOP] Understanding TypeScript

[2014 DLS] Design and evaluation of gradual typing for python
[2013 Empirical Software Engineering] An empirical study on the impact of static typing on software maintainability

[2011 APLAS] Extending Hindley-Milner Type Inference with Coercive Structural Subtyping

[2007 ECOOP] Gradual Typing for Objects
[2004 POPL '04] Principal typings for Java-like languages
[1998 POPL '98] Local type inference

[2009 IEEE Conference on Computer Vision and Pattern Recognition]
Describing objects by their attributes

Zero-Shot

Source Code

Type System

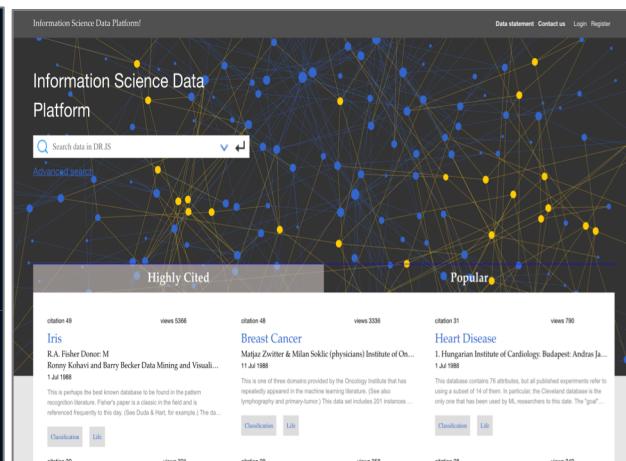
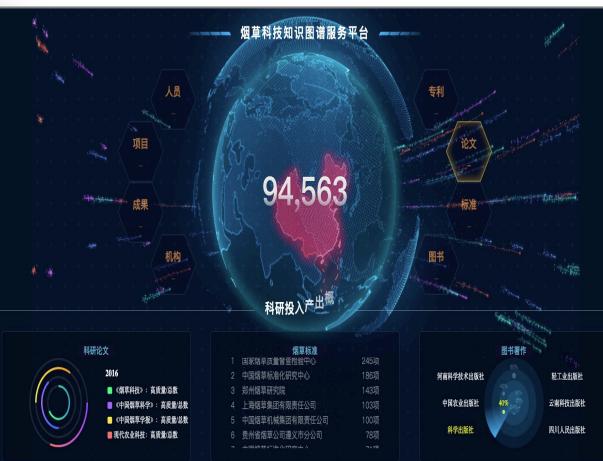
Neural Network

Deep Learning

Local Context

Aminer 论文溯源树

大数据知识管理服务平台



国家自然科学基金
大数据知识
管理服务平台

烟草科技知识图谱
服务平台

Information Science Data
Platform

大数据知识管理服务平台

中国科学院
中国科学院国家空间科学中心
国家科技基础条件平台
中国科学数据云
WODS
lihui 管理 注销 English

空间科学领域知识图谱与决策支持

首页 基础知识 合作网络 关联路径 科研社区 研究热点 态势感知

请您输入要查找人才姓名...

人才信息

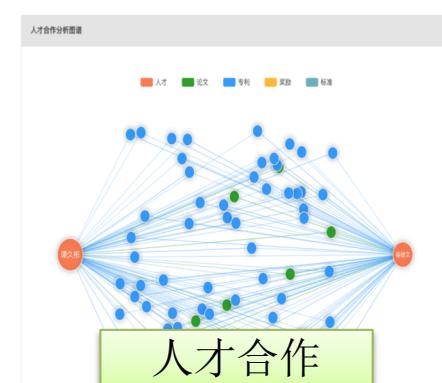
姓名: 王赤
机构: 中国科学院国家空间科学中心
影响力: 12730

项目成果列表

1. 日冕物质抛射事件通过日球... 【项目】
2. 太阳风-磁层相互作用全景... 【项目】
3. 子午工程:空间天气领域e... 【论文】
4. 无振荡、无自由参数格式在... 【论文】
5. 篓环中扭转Alfvén波... 【论文】
6. 基于T96模型的极尖区位... 【论文】
7. 太阳活动与全球气候变化... 【论文】
8. 火星空间环境探测设计与实... 【论文】
9. 萤火一号火星探测器有效载... 【论文】
10. 中国科研信息化蓝皮书20【专著】
11. 进入太空——日地空间探测【专著】

人才
论文
专利
奖励
项目
专著

大数据知识管理服务平台



大数据知识管理服务平台

来自内部的需求逐渐增多

基金委

科学院

空间中心

烟草总局

....

交叉学科

出版社

学术团队

热词矩阵

研究热点

资助方向

专家画像

人才合作

机构画像

大数据知识管理服务平台

来自内部的需求逐渐增多

学术深度挖掘需求迫切

人员关系

科研社区

学术画像

人才引进

学科交叉

资助方向

专家画像

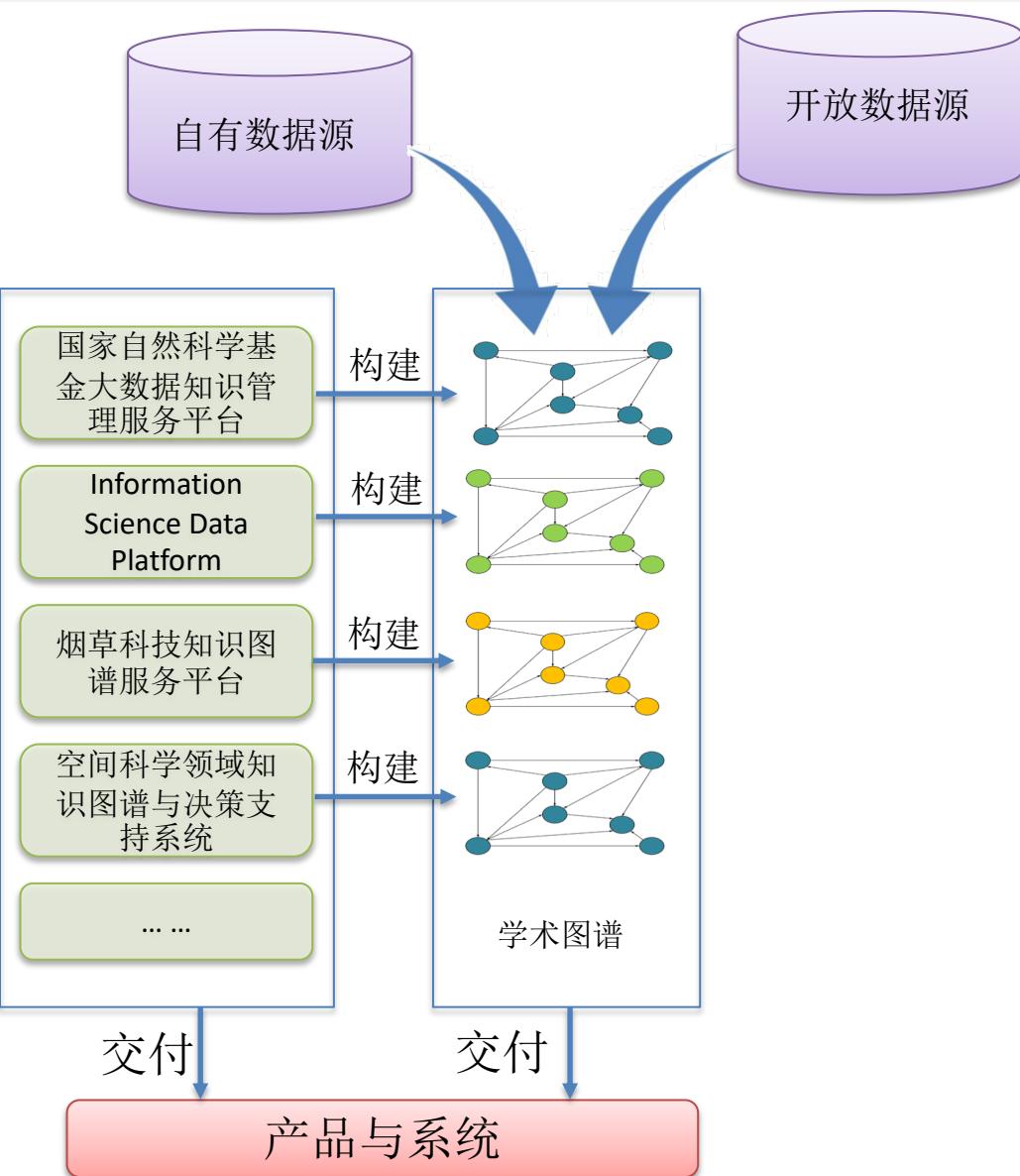
人才合作

机构画像

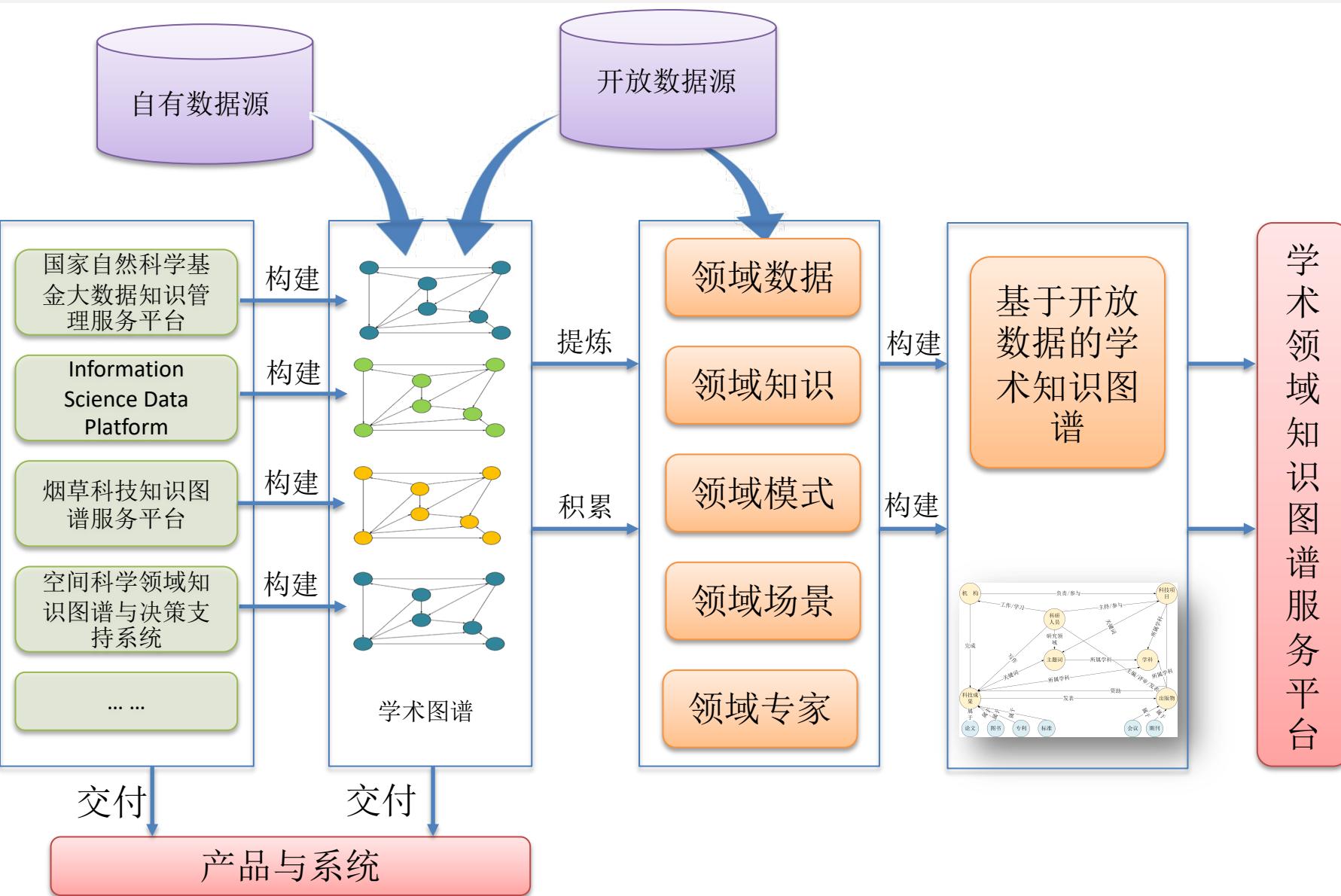


构建科技领域知识图谱，
建设知识管理服务平台，
满足内外部科研学术需求

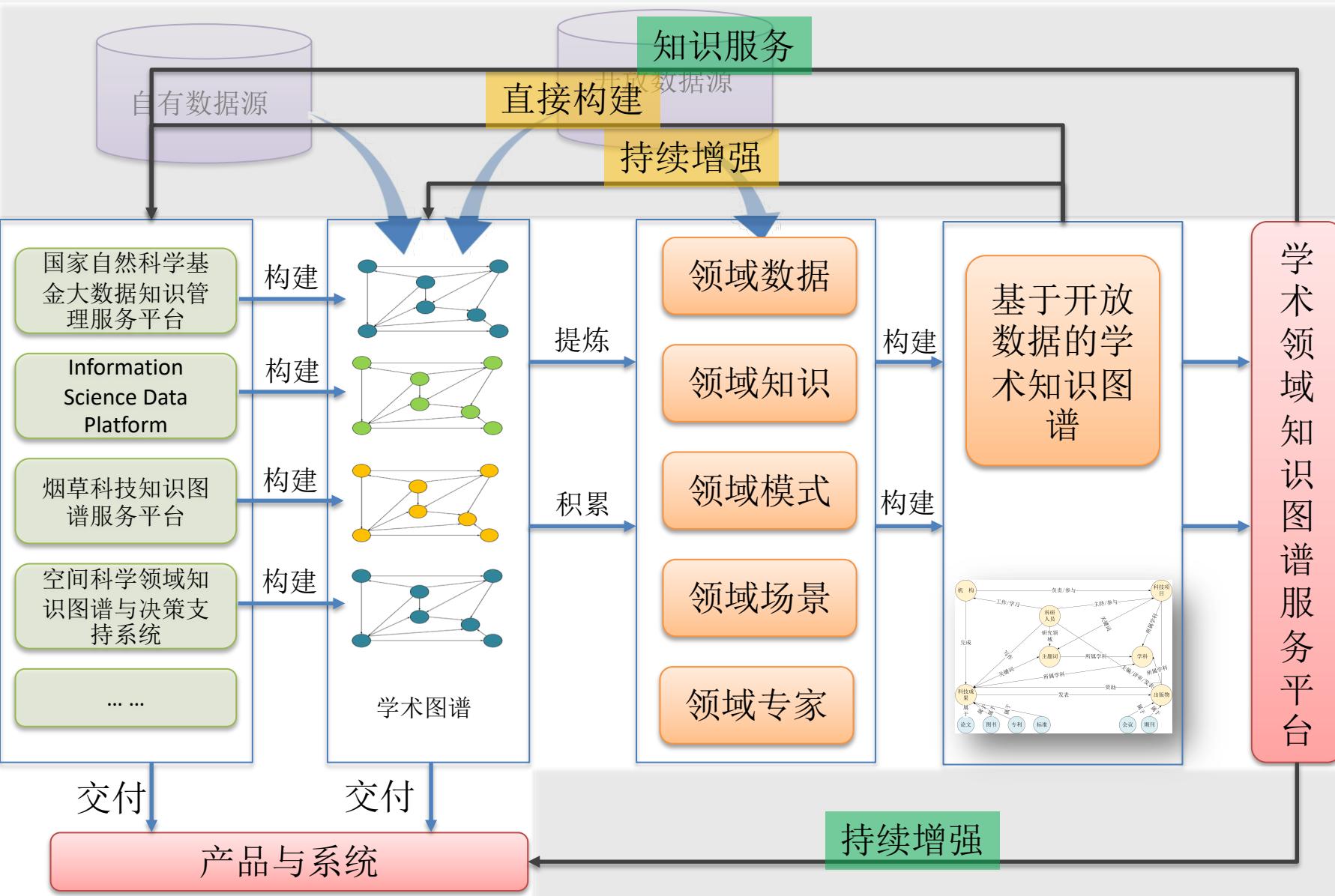
科技领域知识图谱



科技领域知识图谱

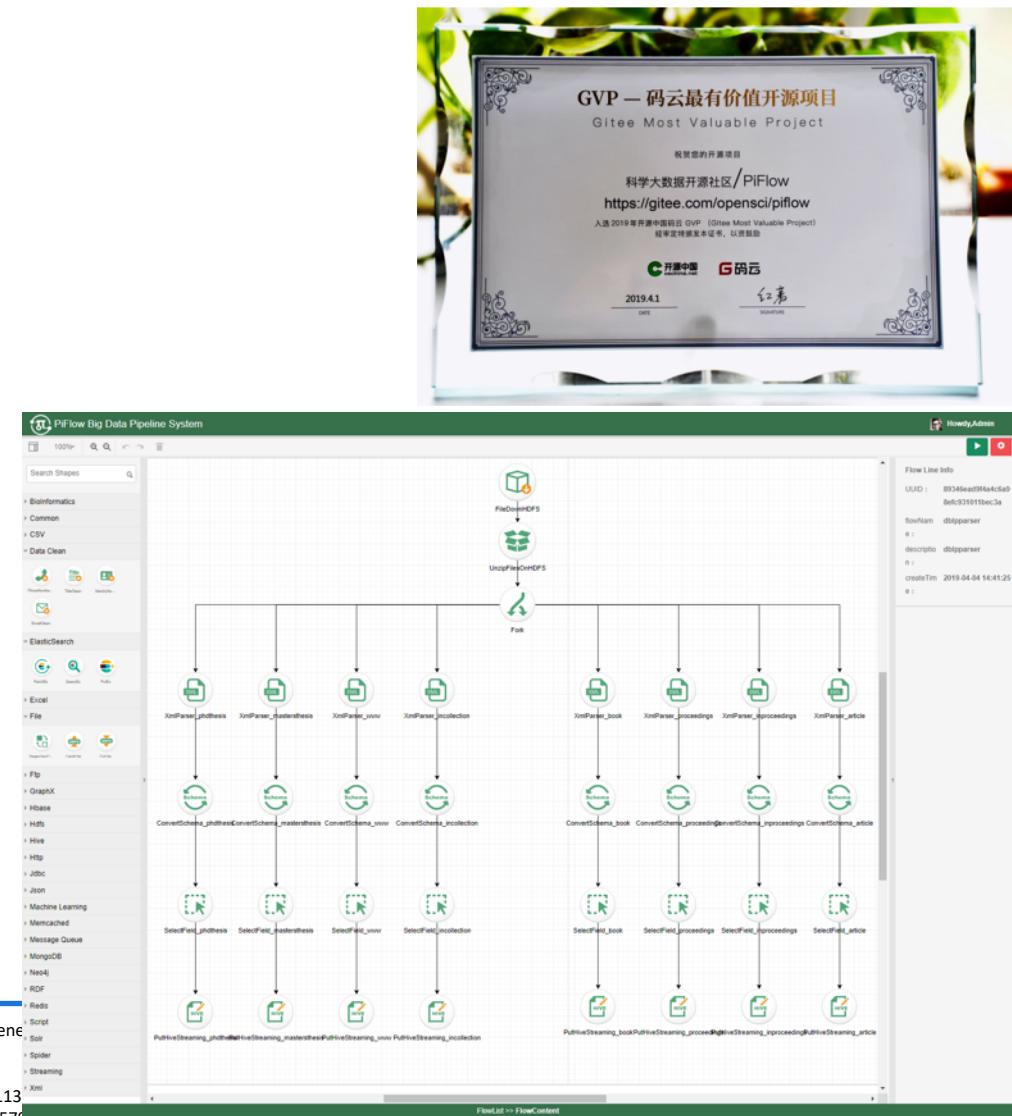
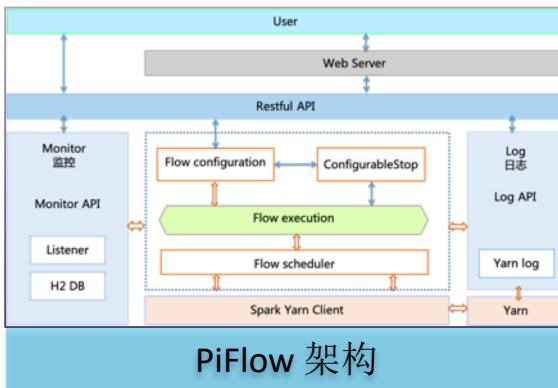


科技领域知识图谱



科技领域知识图谱：关键技术

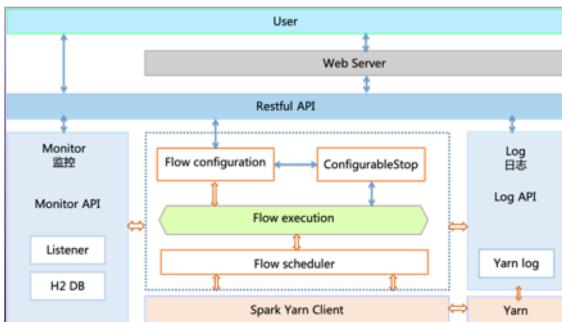
自动及半自动的数据采集、处理流水线机制



1. 乔子越, 杜一*, 傅衍杰, 王鹏飞, 周园春, *Unsupervised Author Disambiguation using Heterogeneous Data*, 2019年“智源-AMiner姓名排歧大赛”第一名
2. PiFlow: 码云2019年最有价值开源项目; 中国科技云首届开源大赛二等奖
3. 杜一, 王寒雪, 乔子越, 周园春, 一种基于网络表征和语义表征的同名作者消歧方法, 2019113
4. 杜一, 乔子越, 周园春, 一种基于异质图卷积神经网络嵌入的作者名字消歧方法, 2019106357934, 2019
5. 杜一, 乔子越, 周园春, 一种基于异质图卷积神经网络嵌入的学者名字消歧方法, 2018112671819, 2018
6. 乔子越, 周园春, 一种基于异质网络嵌入的学者名字消歧方法, 2018112671819, 2018
7. 中国专利: 一种大数据ETL任务的编排方法与系统. 完成人: 朱小杰, 沈志宏, 杜一, 赵子豪, 周园春, 2019
8. 中国专利: 一种大数据ETL任务的调度方法. 完成人: 朱小杰, 沈志宏, 杜一, 赵子豪, 周园春, 2018

科技领域知识图谱：关键技术

自动及半自动的数据采集、处理流水线机制

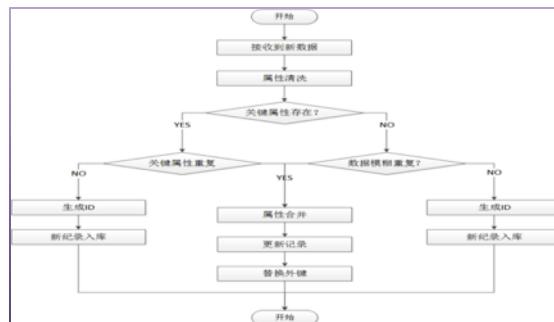


PiFlow 架构



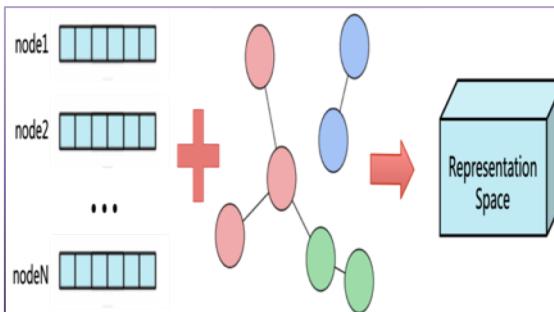
流水线实例

基于统计规则及深度学习方法的实体融合方法

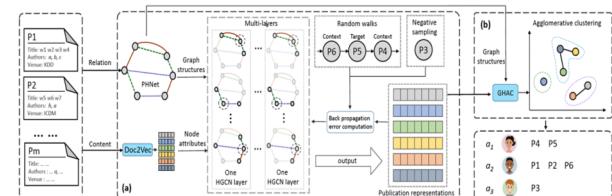


基于规则组合的实体融合

Name	Our method	Component	Zhang et al. 2018	Xu et al. 2018	Zhang et al. 2017
Ajay Gupta	0.750	0.329	0.568	0.552	0.618
Alok Gupta	1	0.690	0.689	0.892	0.590
Bin Yu	0.696	0.292	0.431	0.585	0.614
David Cooper	0.900	0.327	0.737	0.884	0.931
David Nelson	0.944	0.219	0.750	0.735	0.556
Fei Su	1	0.648	0.933	0.630	0.941
Hao Wang	0.604	0.086	0.403	0.557	0.543
Jie Tang	0.982	0.883	0.657	0.522	0.910
Thomas Wolf	0.860	0.502	0.703	0.522	0.352
Yang Wang	0.548	0.118	0.273	0.574	0.409
Avg.	0.786	0.507	0.715	0.681	0.680



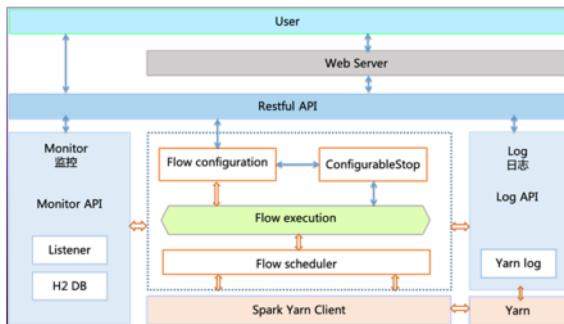
基于GraphEmbedding实体融合



- 乔子越,杜一*,傅衍杰,王鹏飞,周园春,Unsupervised Author Disambiguation using Heterogeneous Graph Convolutional Network Embedding, IEEE Big Data, 2019
- 2019年“智源-AMiner姓名排歧大赛”第一名
- PiFlow: 码云2019年最有价值开源项目;中国科技云首届开源大赛二等奖
- 杜一,王寒雪,乔子越,周园春,一种基于网络表征和语义表征的同名作者消歧方法, 2019113223833 | PCT/CN2019/128642, 2019
- 杜一,乔子越,周园春,一种基于异质图卷积神经网络嵌入的作者名字消歧方法,2019106357994, 2019
- 杜一,乔子越,周园春,一种基于异质网络嵌入的学者名字消歧方法,2018112671819, 2018
- 中国专利: 一种大数据ETL任务的编排方法与系统. 完成人: 朱小杰,沈志宏,杜一,赵子豪,周园春,2019
- 中国专利: 一种大数据ETL任务的调度方法. 完成人: 朱小杰,沈志宏,杜一,赵子豪,周园春,2018

科技领域知识图谱：关键技术

自动及半自动的数据采集、处理流水线机制

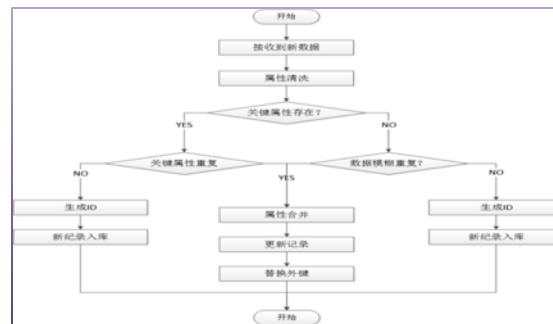


PiFlow 架构

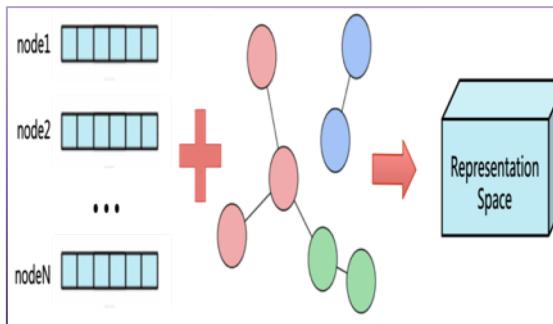


流水线实例

基于统计规则及深度学习方法的实体融合方法

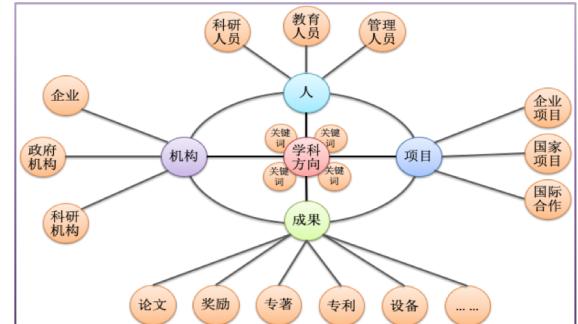


基于规则组合的实体融合

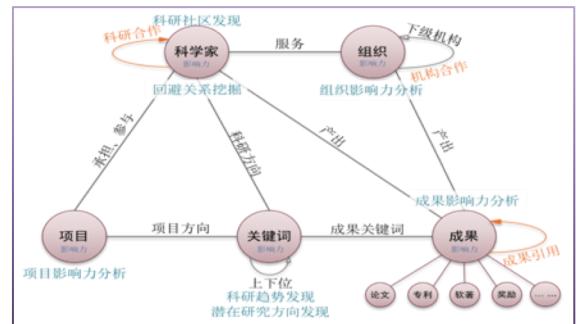


基于GraphEmbedding实体融合

基于属性图的科技领域知识图谱构建方法



基于规则的学术知识图谱



基于推理学习的学术知识图谱

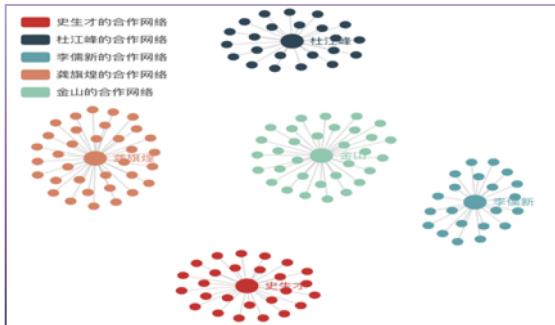
- 乔子越,杜一*,傅衍杰,王鹏飞,周园春,Unsupervised Author Disambiguation using Heterogeneous Graph Convolutional Network Embedding, IEEE Big Data, 2019
- 2019年“智源-AMiner姓名排歧大赛”第一名
- PiFlow: 码云2019年最有价值开源项目;中国科技云首届开源大赛二等奖
- 杜一,王寒雪,乔子越,周园春,一种基于网络表征和语义表征的同名作者消歧方法,2019113223833 | PCT/CN2019/128642, 2019
- 杜一,乔子越,周园春,一种基于异质图卷积神经网络嵌入的作者名字消歧方法,2019106357994, 2019
- 杜一,乔子越,周园春,一种基于异质网络嵌入的学者名字消歧方法,2018112671819, 2018
- 中国专利: 一种大数据ETL任务的编排方法与系统. 完成人: 朱小杰,沈志宏,杜一,赵子豪,周园春,2019
- 中国专利: 一种大数据ETL任务的调度方法. 完成人: 朱小杰,沈志宏,杜一,赵子豪,周园春,2018

科技领域知识图谱：关键技术

基于学科的科研社区发现算法

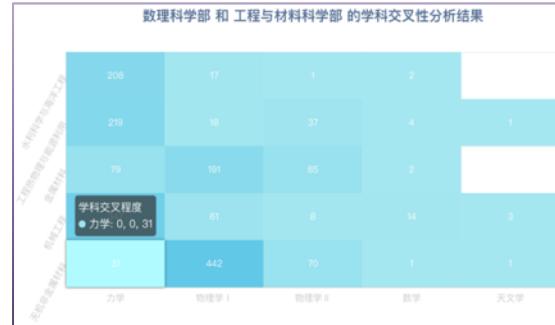


基于PageRank+SVD的高影响力人员关系网络挖掘

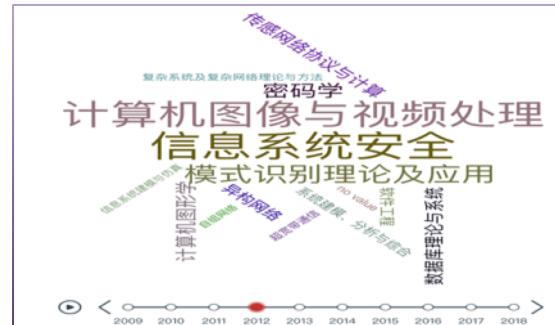


基于网络度量指标+标签传播的 科研社区发现

学科交叉性评估及学科 趋势预测方法



基于统计指标的学科 交叉性评价

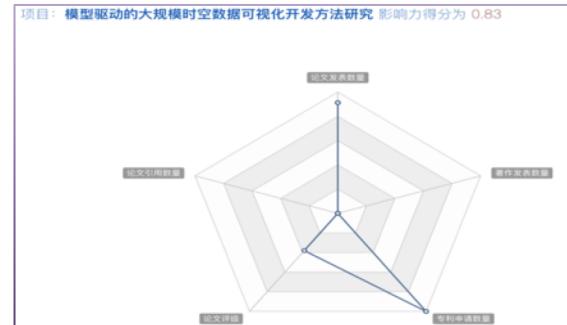


基于时序+图数据库的 学科趋势评估

学术影响力评价方法



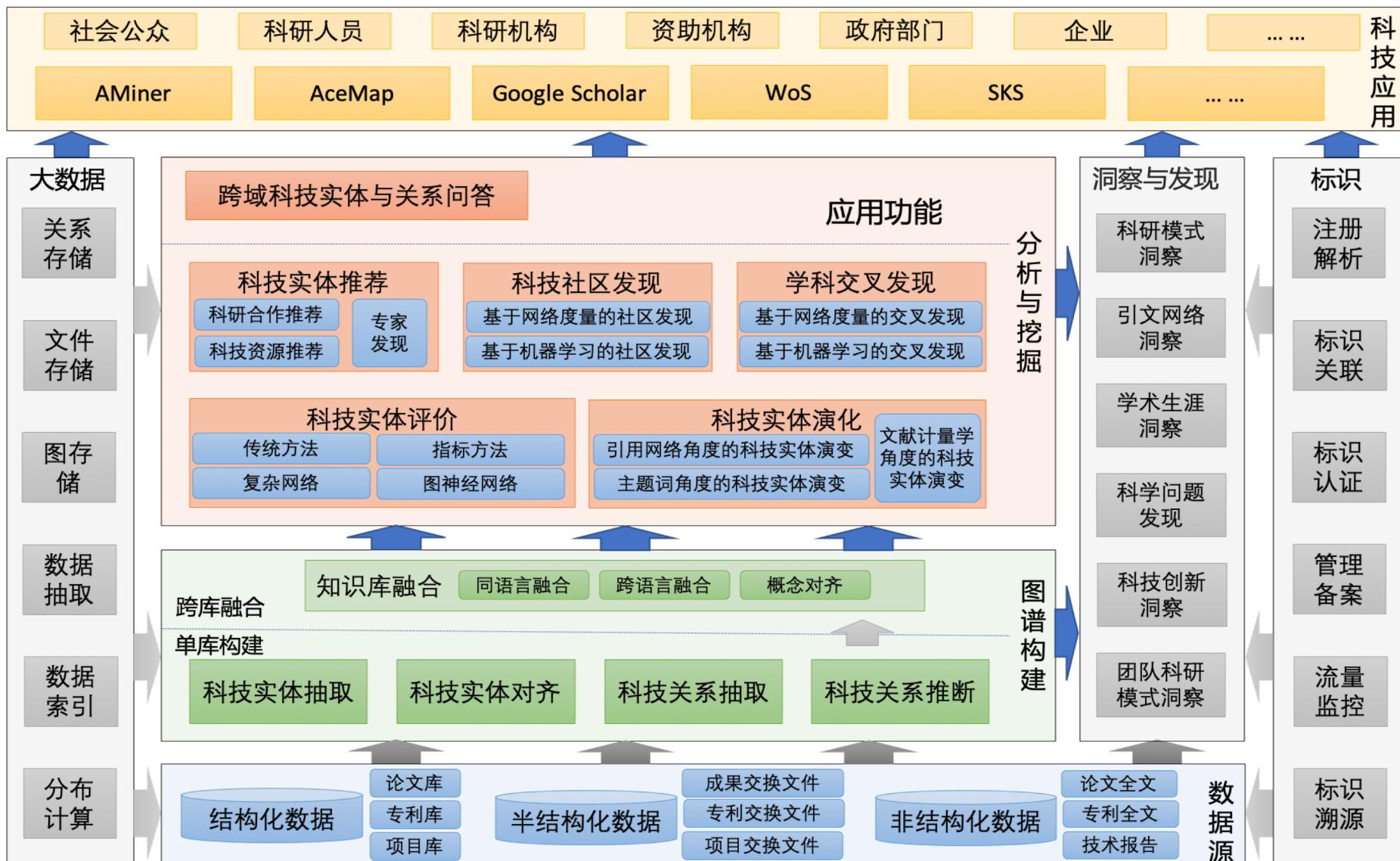
基于网络度量指标+计量学的科 研人员影响力评价



基于网络度量指标+计量学的 科研项目影响力评价

- 王卫军, 崔文娟, 杜一*, 周园春, 基于词嵌入的国家自然科学基金学科交叉知识点分析方法, *情报学报*, 2020.(in submission)
- 姚畅, 王晓帆, 杜一, 张兆田, 李建军, 郝艳妮, 国家自然科学基金大数据知识管理服务平台总体方案及关键技术研究, *中国科学基金*, 2019.

科技领域知识图谱：研究架构



请批评指正！