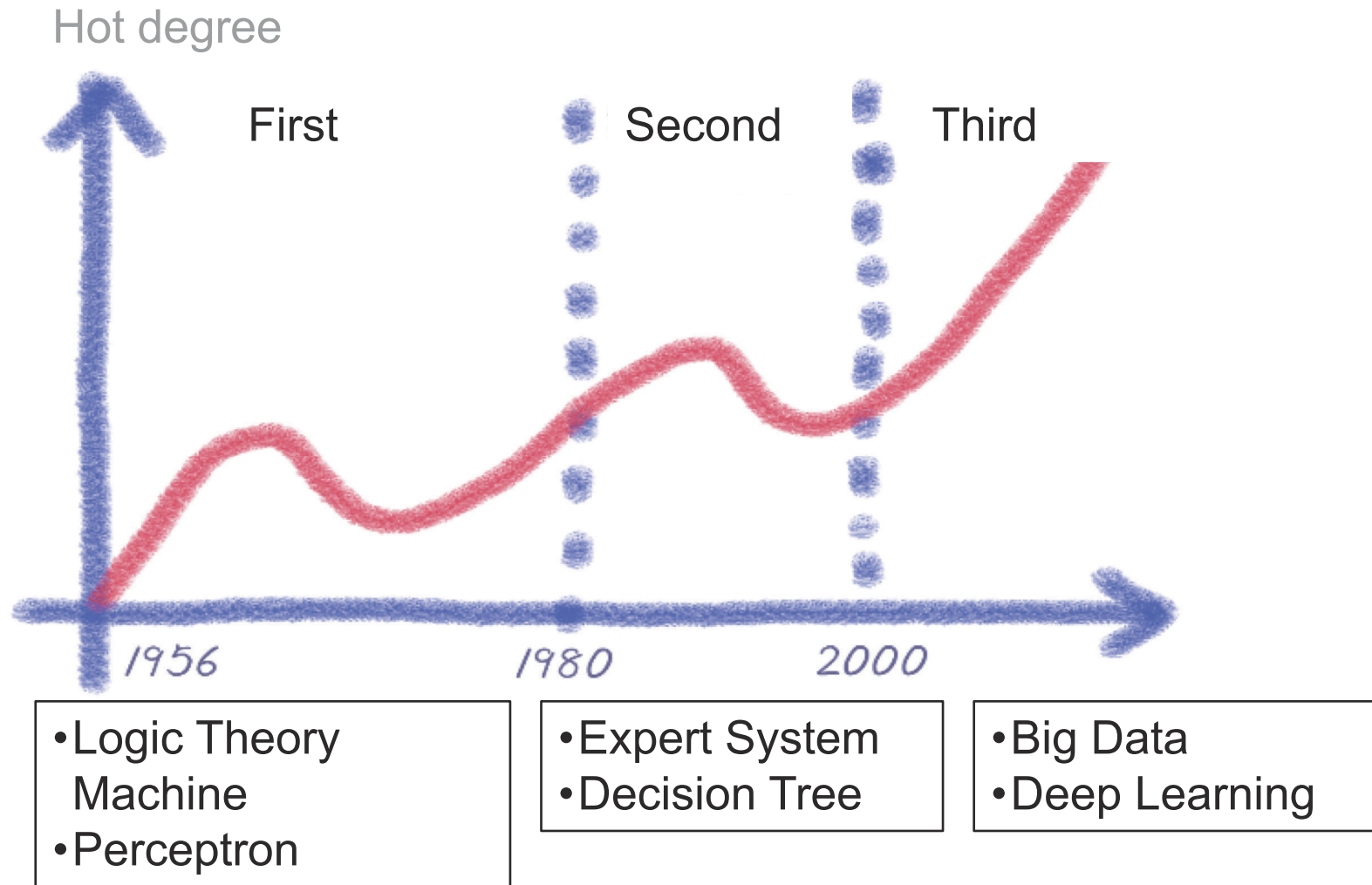


# 科技知识平台AMiner应用案例

—AI技术系列报告

# AI History



# 智能型知识挖掘

传统知识分析

以文档为中心的关键词分析

难以满足产业快速发展的需求

语义挖掘

智能型知识挖掘

构建知识图谱

隐含关联

挖掘知识推理网络

形成知识驱动的智能挖掘技术体系

# 新挑战

1 知识以碎片化的形式分散在异构、多源数据中，**知识获取**难度大；



美国NSF和Army (仅2015年就超过8千万美元) - Science of Science

2 知识网络结构复杂、交互行为动态多样，挖掘知识**隐含关联关系**十分困难；



欧盟第8框架H2020 (6项关于科技情报分析, 共21项学科领域项目)

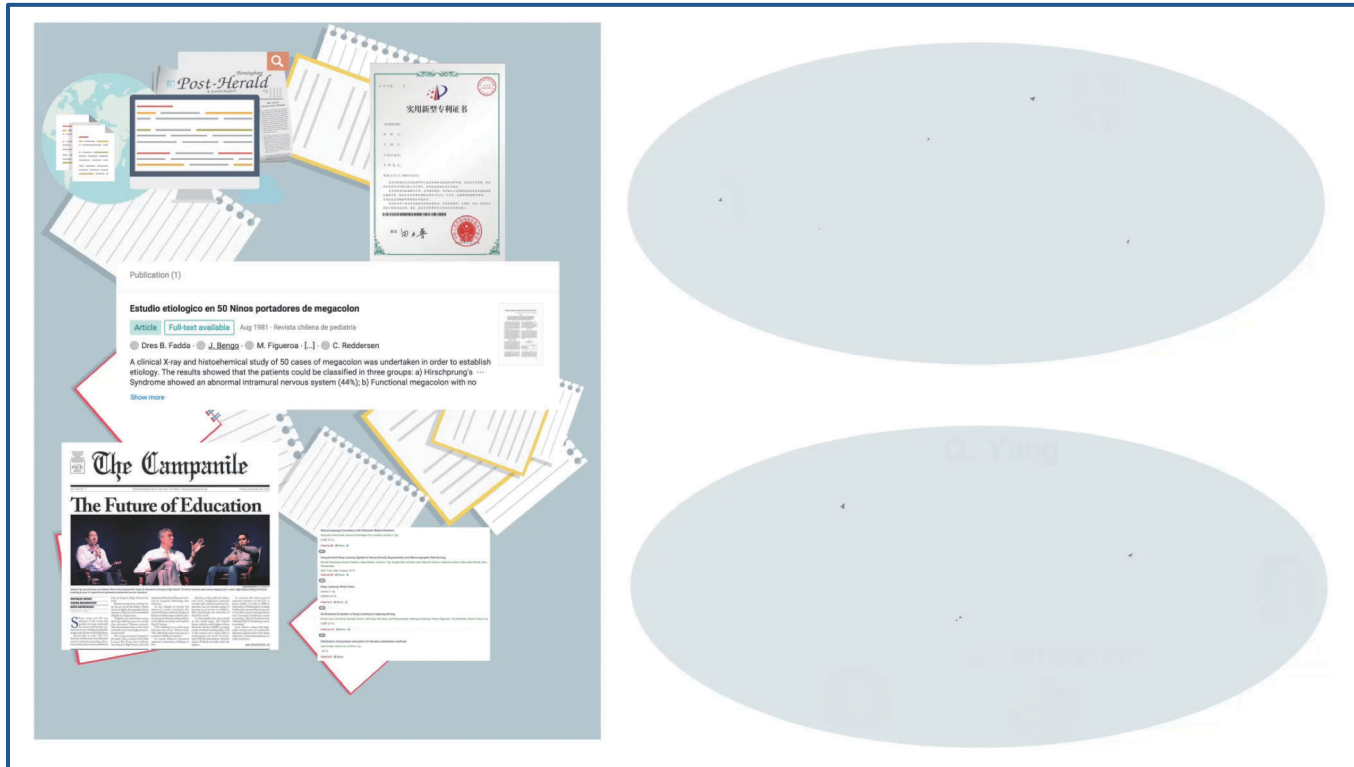
3 知识网络规模大、数据多维，**匹配效率低**是智能型知识计算的瓶颈。

**Science nature**

今年截止目前为止的Nature和Science杂志中已有23篇关于科技大数据挖掘的论文

多个顶尖学术研究机构 (MIT, 哈佛, 康奈尔, 芝加哥等) 建立相关团队

# AMiner.cn: 大规模科技知识图谱



# AMiner.cn

MINING DEEP KNOWLEDGE FROM SCIENTIFIC NETWORKS

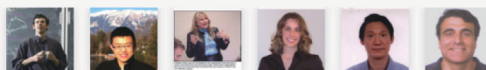
[Advanced](#)

## Hot Topics

### Data Mining



### Social Network



### Machine Learning



### Deep Learning



### Computer Vision



### Database



136,722,706  
RESEARCHERS

101,390,721  
PUBLICATIONS

7,854,301  
CONCEPTS

133,196,029  
CITATIONS

# Open Services

## — 语义搜索

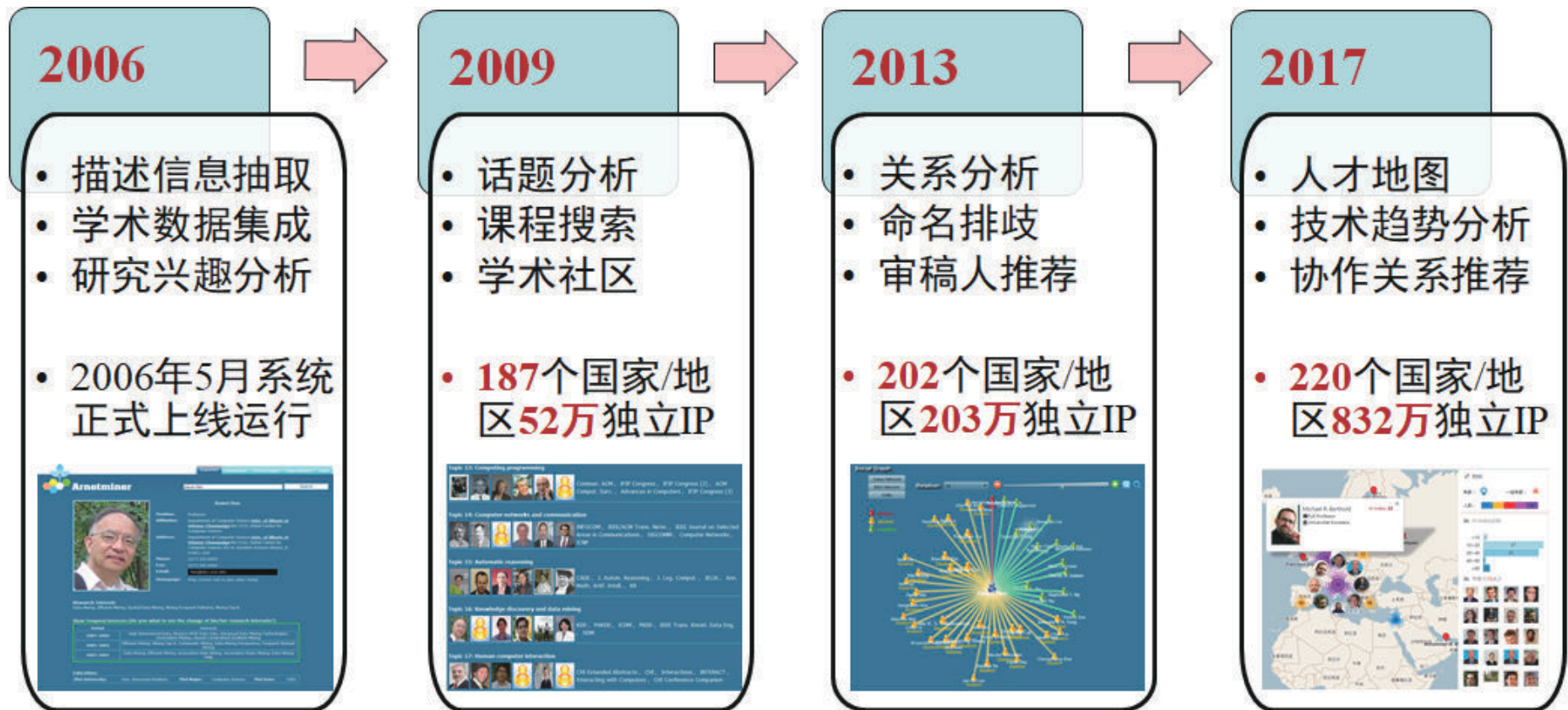
机器学习  
专家

The screenshot shows the Aminer search results for 'machine learning'. The interface includes a search bar at the top with the query 'machine learning'. Below the search bar, there are tabs for 'Expert' and 'Paper'. The 'Expert' tab is active, showing a 'Knowledge Graph' section with a list of related concepts such as 'Computer Science', 'Artificial Intelligence', 'Learning', 'Deep Learning', 'Supervised Learning', etc. A blue box labeled 'Knowledge Graph' highlights this section. Below the 'Expansion' section, there are 'Search Results' filters for 'h-index', 'Gender', 'Language', and 'Location'. A blue box labeled 'Demographics: gender, language, location, etc.' highlights these filters. The 'Relevance' section shows various ranking metrics like 'h-index', 'A-index', 'Activity', etc. A blue box labeled 'Ranking Metrics' highlights this section. The main content area displays profiles of experts, including Thomas G. Dietterich, Guang-Bin Huang, and Bernhard Schölkopf. A blue box labeled 'Rich Semantics' highlights the 'Similar' tags for these experts. On the right side, there is a 'Machine Learning' section with a 'Popularity Over Time' bar chart and a 'Description' box. A blue box labeled 'Machine Learning' highlights this section. The URL 'https://aminer.cn' is visible at the bottom left.

<https://aminer.cn>

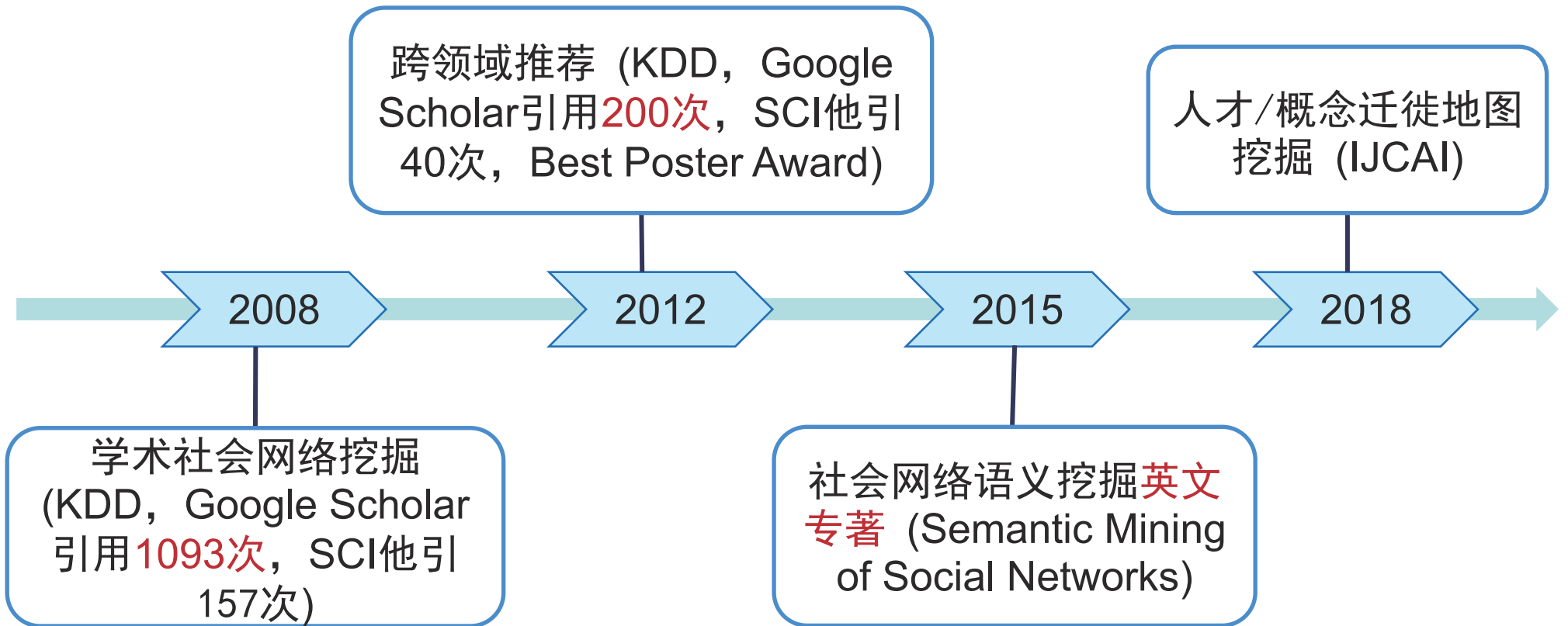
# AMiner—知识驱动的科技大数据挖掘平台

□ 研发了知识驱动的科技情报挖掘平台AMiner，建立超过**2亿篇**学术论文/专利和**1.36亿**科研人员的科技智库





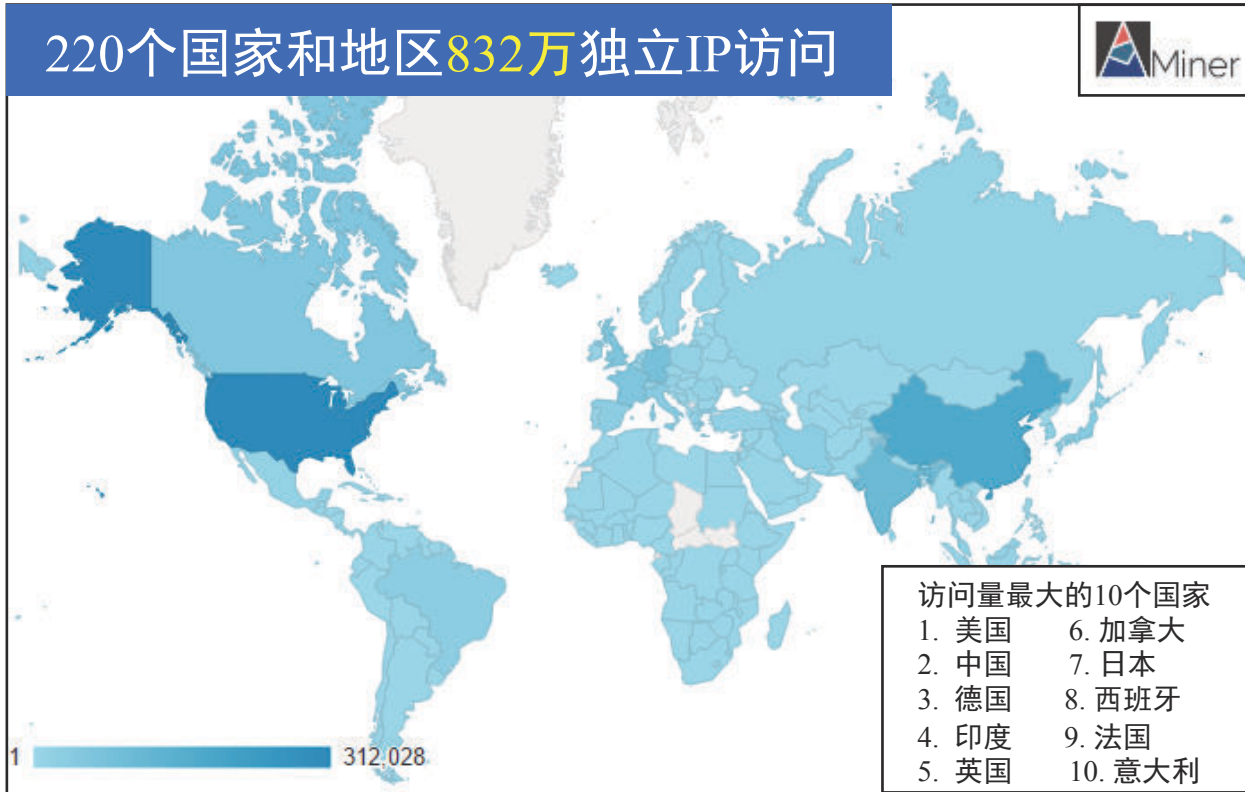
# 代表性论文



# AMiner访问量及用户分布

Google Analytics

220个国家和地区832万独立IP访问



爱尔兰研究机构  
DERI资深研究员  
P. Buitelaar



## Exploring Your Research: Sprinkling some Saffron on Semantic Web Dog Food

Fergal Monaghan, Georgeta Bordea, Krystian Samp, and Paul Buitelaar

fron cannot be positioned relative to all of these here given the limited space, we now compare Saffron with the most related work: ArnetMiner<sup>8</sup> [6], a well-known state of the art “academic researcher social network search” tool.

ArnetMiner has an emphasis on classification and consists of two main parts. In the first part, probabilistic topic models such as Latent Dirichlet Allocation (LDA) [1] are extended and a unified topic model for papers, authors and conferences is proposed. It seems only the content of the papers is analysed, and structured data, such as social connections, is not considered. They cluster all the words into 100 topics, which is a rather small number considering their size.

AMiner is a well-known state of the art tool.....

在线运行超过十年

科研数据下载230万次，年均数据访问量超过1100万次

# 核心技术

— How to populate a **semantic**-based profile database for researchers?



Edward Feigenbaum  
专家系统之父  
图灵奖

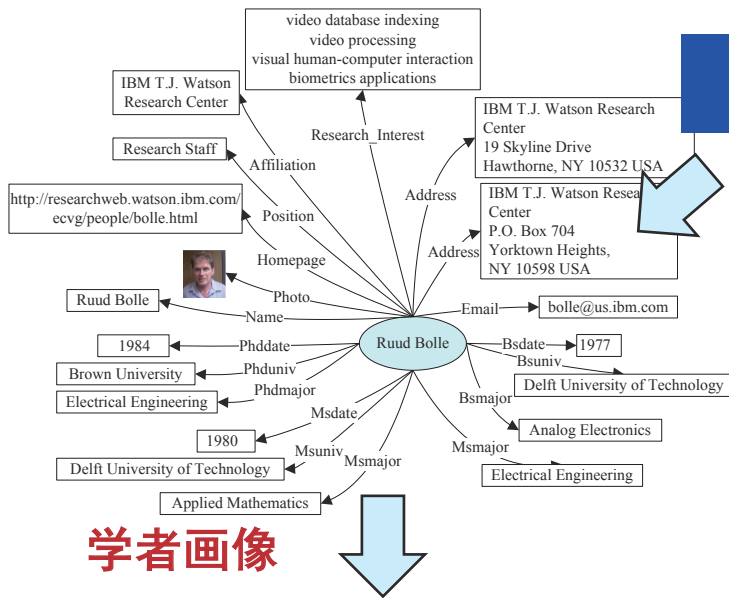


Tim Berners Lee  
WWW创始人  
图灵奖



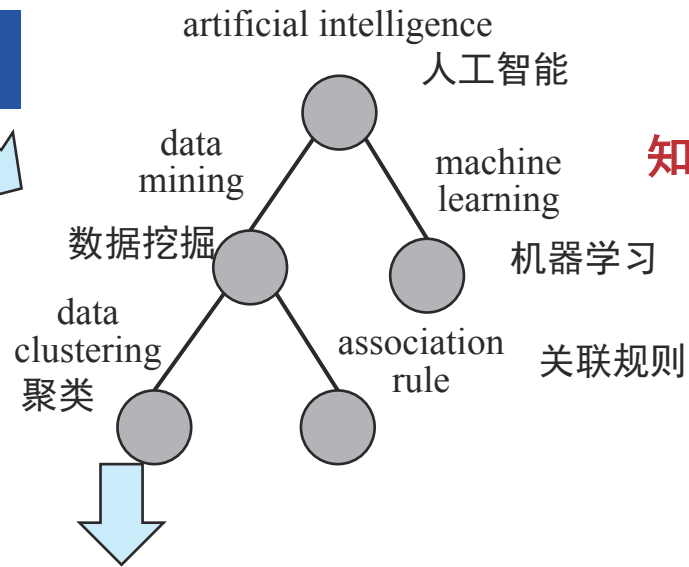
\*人工智能的两个重要阶段：大规模知识库 + 智能服务

# Architecture—以AI为例

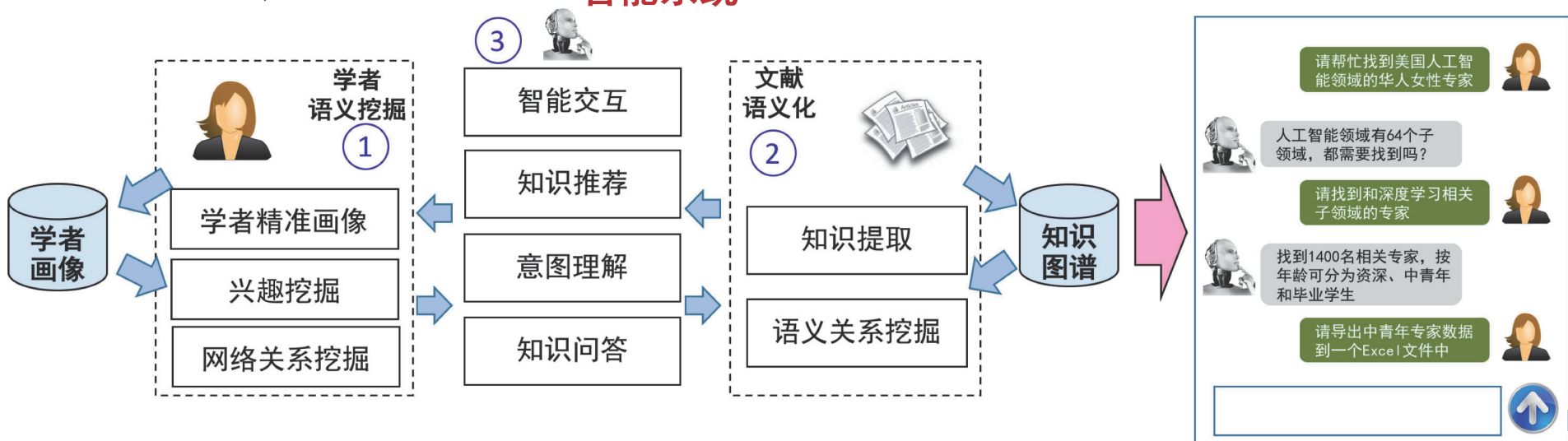


学者画像

智能系统



知识图谱



# Technology Overview

Data->Semantics  
Search->Intelligence

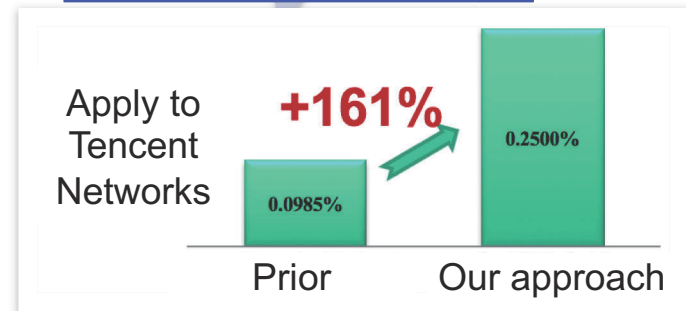
智能服务

语义集成

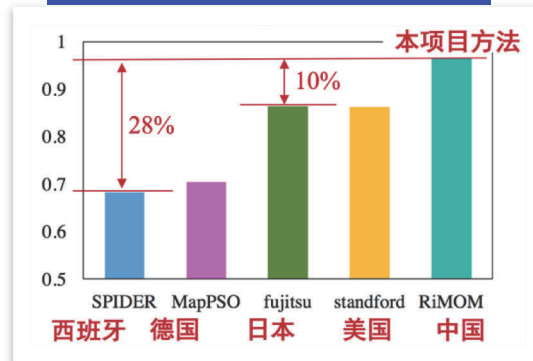
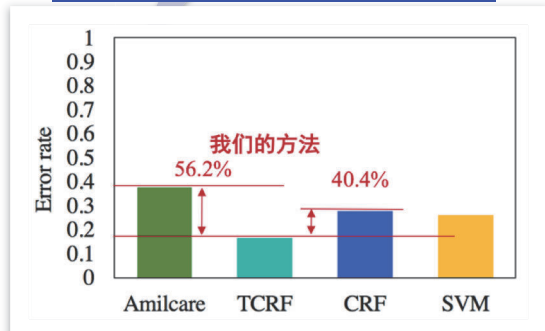
Recommendation  
accuracy +161%

数据语义化

15 champions in  
the past 7 years



Error rate reduced  
40-56%



Reported by UN



UNITED NATIONS GLOBAL PULSE

Harnessing big data for development and humanitarian action



ABOUT  
PROJECTS  
LABS  
NEWS  
CHALLENGES  
PRIVACY

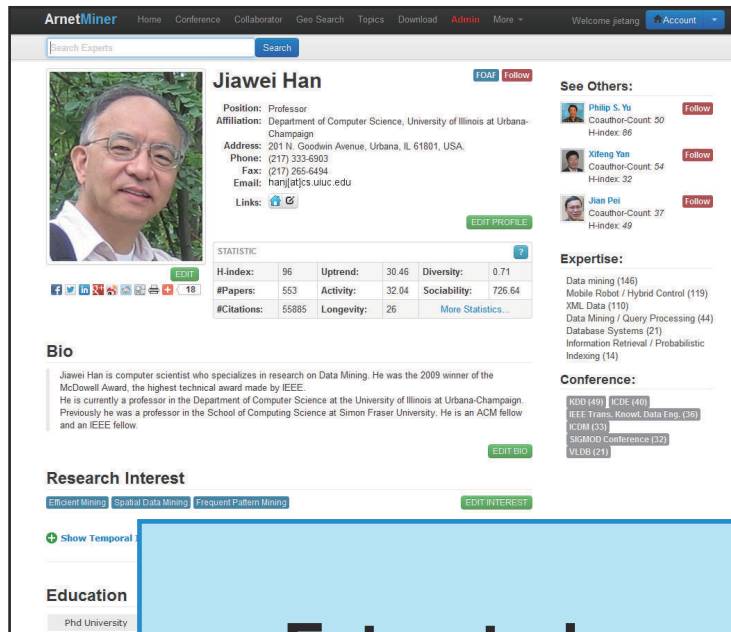
PULSE LAB DIARIES

Research Bites: "Inferring User Demographics and Social Strategies in Mobile Social Networks"

Aug 24, 2014

#citation>10,000, published ~200 papers on major journals and conferences

# Researcher Profile Database



**Jiawei Han** FOAF Follow

**Position:** Professor  
**Affiliation:** Department of Computer Science, University of Illinois at Urbana-Champaign  
**Address:** 201 N. Goodwin Avenue, Urbana, IL 61801, USA.  
**Phone:** (217) 333-6903  
**Fax:** (217) 265-6494  
**Email:** hanj[at]cs.uiuc.edu  
**Links:** [Social Media Icons]

**STATISTIC**

H-index:	96	Uptrend:	30.46	Diversity:	0.71
#Papers:	553	Activity:	32.04	Sociability:	726.64
#Citations:	55885	Longevity:	26	<a href="#">More Statistics...</a>	

**Bio**


Jiawei Han is computer scientist who specializes in research on Data Mining. He was the 2009 winner of the McDowell Award, the highest technical award made by IEEE. He is currently a professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. Previously he was a professor in the School of Computing Science at Simon Fraser University. He is an ACM fellow and an IEEE fellow.

**Research Interest**

[Efficient Mining](#) [Spatial Data Mining](#) [Frequent Pattern Mining](#)

**Education**

Phd University



**M. I. Jordan** FOAF Follow

**ALIAS:** Michael I. Jordan Michael Jordan Michael Irwin Jordan

**Position:** Professor  
**Affiliation:** Department of EECS Department of Statistics University of California, Berkeley  
**Address:** University of California, Berkeley EECS Department 731 Soda Hall #1776 Berkeley, CA 94720-1776  
**Phone:** (510) 642-3806  
**Fax:** (510) 642-5775  
**Email:** jordan@stat.berkeley.edu  
**Links:** [Social Media Icons]

**STATISTIC**

H-index:	75	Uptrend:	7.2	Diversity:	0.03
#Papers:	242	Activity:	11.12	Sociability:	331.69
#Citations:	44312	Longevity:	23	<a href="#">More Statistics...</a>	

**Expertise:**

- Data mining (146)
- Mobile Robot / Hybrid Control (119)
- XML Data (110)
- Data Mining / Query Processing (44)
- Database Systems (21)
- Information Retrieval / Probabilistic Indexing (14)

**Conference:**

- RDD (49)
- ICDE (40)
- IEEE Trans. Knowl. Data Eng. (36)
- ICDM (33)
- SIGMOD Conference (32)
- VLDB (21)



**H. Garcia** FOAF Follow

**ALIAS:** H. Garcia Molina H. Garcia Molina Hector Garcia Molina Hector Garcia Molina

**Position:** Professor  
**Affiliation:** Departments of Computer Science and Electrical Engineering.

**STATISTIC**

H-index:	75	Uptrend:	7.2	Diversity:	0.03
#Papers:	242	Activity:	11.12	Sociability:	331.69
#Citations:	44312	Longevity:	23	<a href="#">More Statistics...</a>	

**See Others:**

- Andreas Paepcke** Coauthor-Count: 32 H-index: 37
- Jennifer Widom** Coauthor-Count: 24 H-index: 79
- v. Barbara** Coauthor-Count: 26 H-index: 0

**Expertise:**

- Data (115)
- Database Systems (60)
- Time Systems / Automated Test Data (30)
- Indexing (23)
- Mobile Robot / Hybrid Control (22)
- Library / Information Access

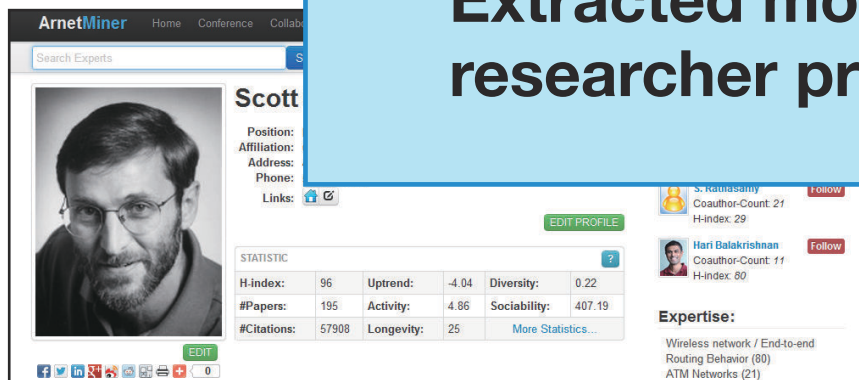
**Conference:**

- VLDB (27)
- SIGMOD Conference (25)
- ICDCS (23)
- IEEE Data Eng. Bull. (13)
- PVLDB (11)

**Research Interest**

[Database Systems](#) [Data Management](#) [Data Warehousing](#)

Extracted more than 1,000,000 researcher profiles from the Web



**Scott**

**Position:** Professor  
**Affiliation:** Department of Computer Science, University of Illinois at Urbana-Champaign  
**Address:** 201 N. Goodwin Avenue, Urbana, IL 61801, USA.  
**Phone:** (217) 333-6903  
**Email:** [Redacted]  
**Links:** [Social Media Icons]

**STATISTIC**

H-index:	96	Uptrend:	-4.04	Diversity:	0.22
#Papers:	195	Activity:	4.86	Sociability:	407.19
#Citations:	57908	Longevity:	25	<a href="#">More Statistics...</a>	

**Expertise:**

- Wireless network / End-to-end Routing Behavior (80)
- ATM Networks (21)

[1] J. Tang, L. Yao, D. Zhang, and J. Zhang. A Combination Approach to Web User Profiling. ACM Transactions on Knowledge Discovery from Data (TKDD), (vol. 5 no. 1), Article 2 (December 2010), 44 pages.

# 人才大数据 — 人才地图



学术资源地图

Data Mining

搜索

登录

专家 地图

搜索范围: ALL IEEE Fellow(2013-2016) ACM Fellow 英国皇家科学院 - Research Fellows Directory 英国皇家科学院 - Fellows Directory 中国科学院院士  
美国科学院外国专家

过滤条件: ALL x

h指数: >=60 (68) 50-59 (43) 40-49 (75) 30-39 (137) 20-29 (222) 10-19 (357) <10 (98)

语言: Chinese (259) English (196) Greek (39) French (29) German (28) Japanese (26) Indian (24) Korean (14) Italian (1)

国家: USA (229) China (111) United Kingdom (19)



学术资源地图

Data Mining

搜索

登录

专家 地图

综合排序 H-index 学术活跃度



韩家炜 (Jiawei Han)

H-index: 152 A-index: 430.90  
Abel Bliss Professor  
Department of Comp  
(217) 333-6903  
hanj@illinois.edu



俞士纶 (Philip S. Yu)

H-index: 144 A-index: 457.58  
Distinguished Profess  
Department of Comp  
(312) 996-0498  
psyu@uic.edu



Mohammed J. Zaki

H-index: 61 A-index: 22.81  
Professor  
Department of Comp  
(518) 996-6340  
zaki@cs.rpi.edu

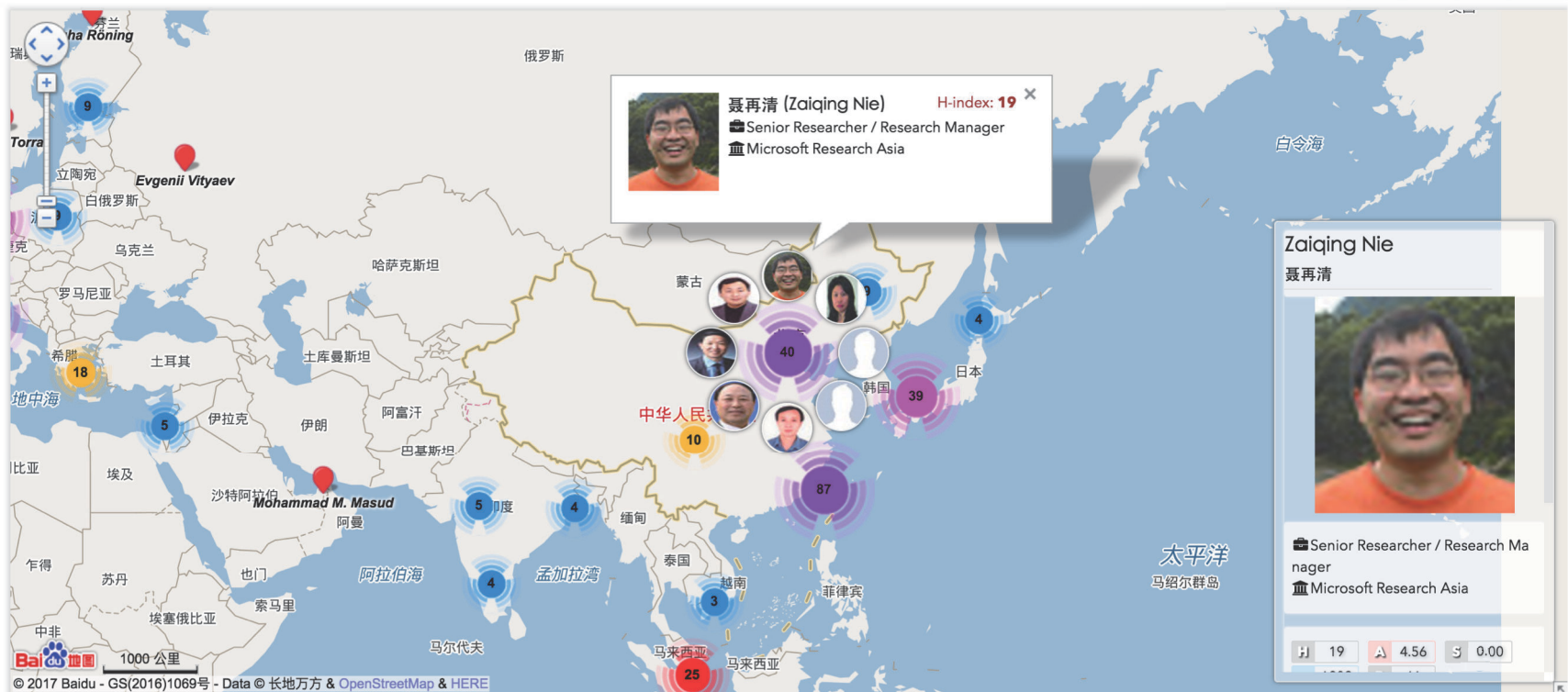


Jian Pei

H-index: 75 A-index: 61.60  
Professor  
School of Computing  
1-778-782-6851, 1-778  
jpei@cs.edu.ca, jpei

专家地图:

按照层级显示: 自动 大区 国家 城市 机构



# 人才大数据 - CCF专家系统



CCF 专家库



- 首页
- 学会活动
- 活动统计
- 设置
  - 用户管理
  - 角色管理
  - 协办单位
  - 贡献类别
  - 活动类型
  - 机构列表

人工智能

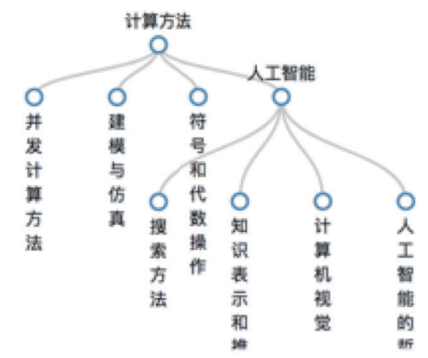
搜索范围: **CCF会员** 会士(F) 杰出会员(D) 高级会员(S) 全球专家

过滤条件: **CCF会员** ×

h指数: >=60 (2) 50-59 (2) 40-49 (2) 30-39 (13) 20-29 (11) 10-19 (46) <10 (123)

语言: Chinese (187)

国家: China (164) Singapore (1)



导出

相关度 | 学术成就 | 学术活跃度 | 领域新星 | 学会贡献

**刘大有 (Dayou Liu)**  
0.00 | 23 | 19.57  
吉林大学计算机科学系/计算机科学与技术学院  
liudy@jlu.edu.cn

研究兴趣: 人工智能, 定性空间推理, 遗传算法, 复杂网络, 约束满足问题, 数据挖掘, 计算机应用, 时空推理

**欧阳彤彤 (Dantong Ouyang)**  
0.00 | 10 | 1.73  
吉林大学计算机科学与技术学院  
ouyd@jlu.edu.cn

研究兴趣: 基于模型的诊断, 人工智能, 基于模型诊断, 启发式搜索, 离散事件系统, 混合键合图, 约束满足问题, 极小碰集

**宗成庆 (Chengqing Zong)**  
0.00 | 18 | 37.97  
Professor  
中科院自动化研究所  
+86-10-8254 4688  
cqzong@nlpr.ia.ac.cn

研究兴趣: 机器翻译, 人工智能, 自然语言处理, 条件随机场, 中文信息处理, 机器翻译评测, 计算机应用, 命名实体

**刘群 (Qun Liu)**  
0.00 | 31 | 55.10

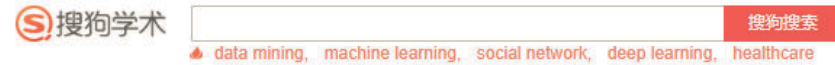
研究兴趣:



# AMiner应用于搜狗搜索

建立常用搜索同义词  
库及上万节点的知识  
图谱

## 搜狗学术搜索提供者



### 2016 Most Influential Scholars (Computer Science)



Timothy L. Harris

h-index: 40 | 论文数: 75 | 引用数: 16931

Lead Researcher

Oracle Labs group in Cambridge, UK

Transactional Memory | Software Transactional Memory



Peter Druschel

h-index: 68 | 论文数: 196 | 引用数: 38498

Professor

Distributed Systems Group Max Planck Institute for Softwa...

Operating System | Overlay Network | P2p | ...



Andrew Warfield

h-index: 24 | 论文数: 56 | 引用数: 14700

Associate Professor

Department of Computer Science University of British Col...

Virtual Machine | Virtualization | High Availability | ...



Michael Burrows

h-index: 38 | 论文数: 90 | 引用数: 20023

Computer Scientist

Google

Distributed System | Access Control | Rocky Shore



Keir Fraser

h-index: 16 | 论文数: 24 | 引用数: 14297

Co-Founder and Chief Architect

Coho Data

Operating System | Virtual Machine Monitor | ...

### Academic Rankings

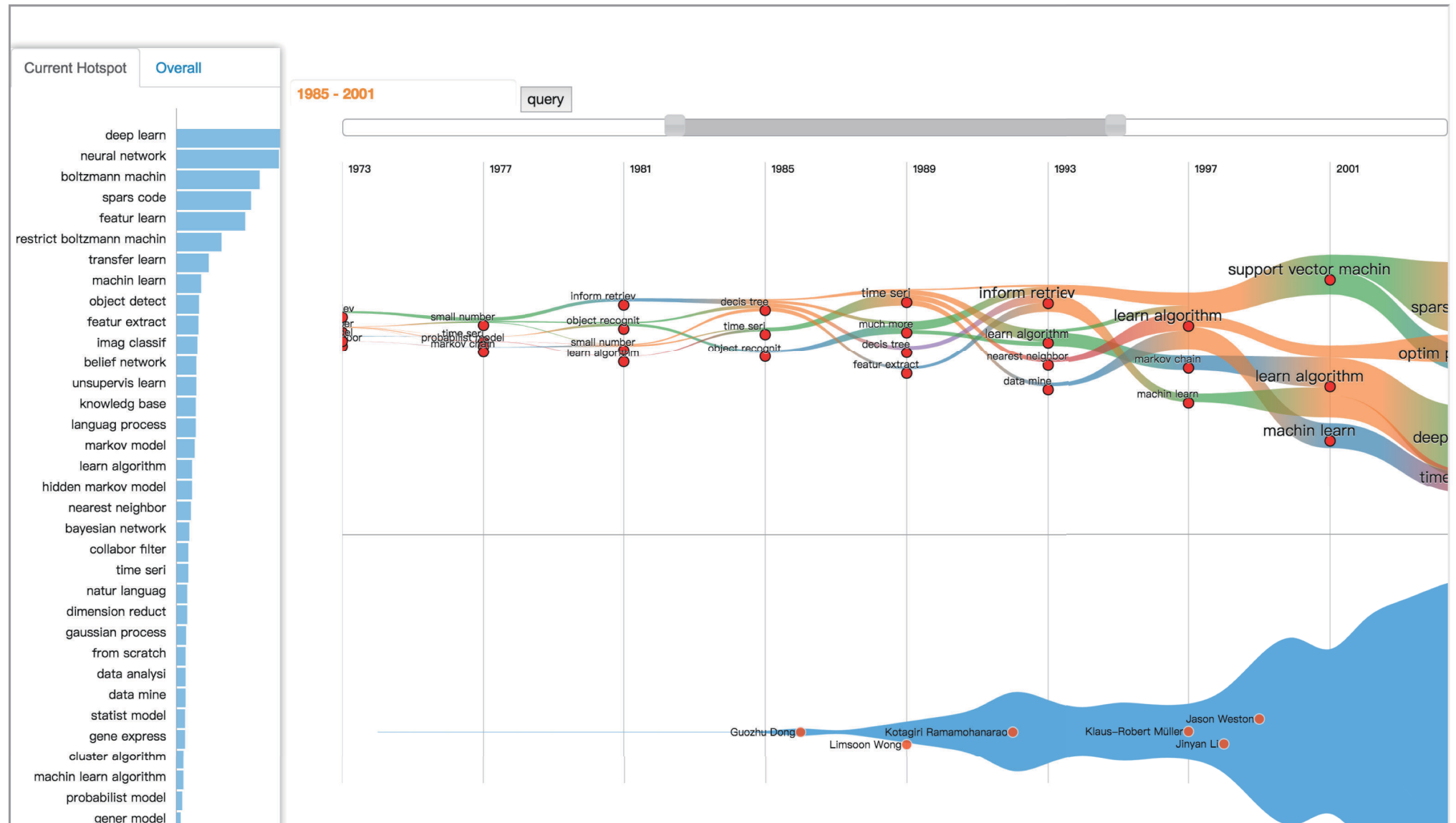
- Researcher Rank**  
Rank researchers by various metrics.
- Organization Rank**  
Rank organizations by different metrics.
- Conference Rank**  
Rank conferences in Computer Science by Impact Factor.
- Best Papers & Top Cited Papers**  
Best papers and top cited papers in Computer Science.

### AI LABS

- Gender Prediction**  
Predict gender with people name and affiliation/location.
- Open Data**  
Citation network, topic experts, disambiguated names, etc.
- Open Seminar**  
Recent academic seminars, e.g., all seminars from Tsinghua.
- Open API**  
APIs to access our data and to build advanced functions.

# 知识图谱 — 技术趋势预测

技术趋势预测:





# 知识图谱 — 技术趋势预测


## 知识图谱

- L0 Design and analysis of algorithms (算法的设计与分析) 11
- L1 Data structures design and analysis (数据结构的设计与分析) 5
- L2 Pattern matching (模式匹配) 20
- **L3 data mining**
- L3 pattern matching
- L3 lempel-ziv compressor
- L3 geometry and discrete structures
- L3 indexing data structure
- L0 Visualization (可视化) 6
- L1 Visualization application domains (可视化应用程序域) 4
- L2 Scientific visualization (科学可视化) 20  
ALIAS: scientific visualisation,
- L3 reporting system
- **L3 data mining**
- L3 transformation and loading
- L3 bioinformatics
- L3 extraction
- L0 Probability and statistics (概率和统计) 9

INFO EXPERTS PUBLICATIONS MORE >

 **刘冰 (Bing Liu)**  
73 41056 272  
Professor  
Department of Computer Science, University of Illinois at Chicago (UIC)  
研究兴趣:  
Data Mining Opinion Mining Web Pages Web Mining  
Association Rule Proposed Technique Knowledge Discovery  
Web Page

 **Wynne Hsu**  
47 17021 196  
Provost's Chair; Professor  
Department of Computer Science, School of Computing, National University of Singapore  
研究兴趣:  
Data Mining Indexation Association Rule Mining  
Association Rule Data Structure Experiment Result  
Feature Extraction Data Cleaning

 **Tok Wang Ling**  
35 4741 256  
Professor  
Department of Computer Science School of Computing National University of Singapore  
研究兴趣:  
Xml Document Xml Xml Database Xml Data  
Data Model Entity Relationship Keyword Search

# Open Report—技术发展报告

[https://www.aminer.cn/research\\_report/articlelist](https://www.aminer.cn/research_report/articlelist)



# THU AI-TR人工智能系列报告

THU AI TR

THU AI TR

## 人工智能之数据挖掘 Research Report of Data Mining

2019年 第1期



清华大学人工智能研究院  
北京智源人工智能研究院  
清华-工程院知识智能联合研究中心  
2019年1月

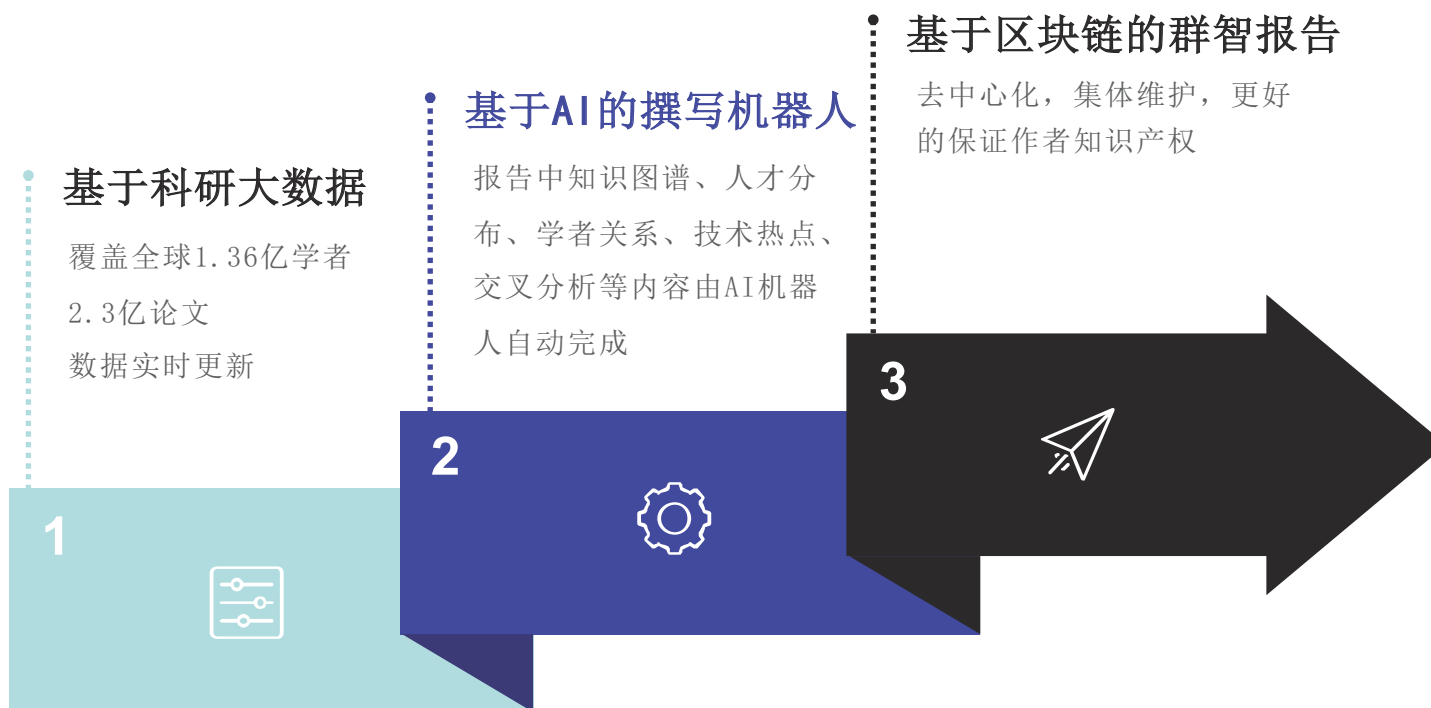
## 人工智能之知识图谱 Research Report of Knowledge Graph

2019年 第2期



清华大学人工智能研究院  
北京智源人工智能研究院  
清华-工程院知识智能联合研究中心  
2019年1月

# 3大技术优势





Edward Feigenbaum  
专家系统之父  
图灵奖

# AMiner Future: 大规模常识知识图谱

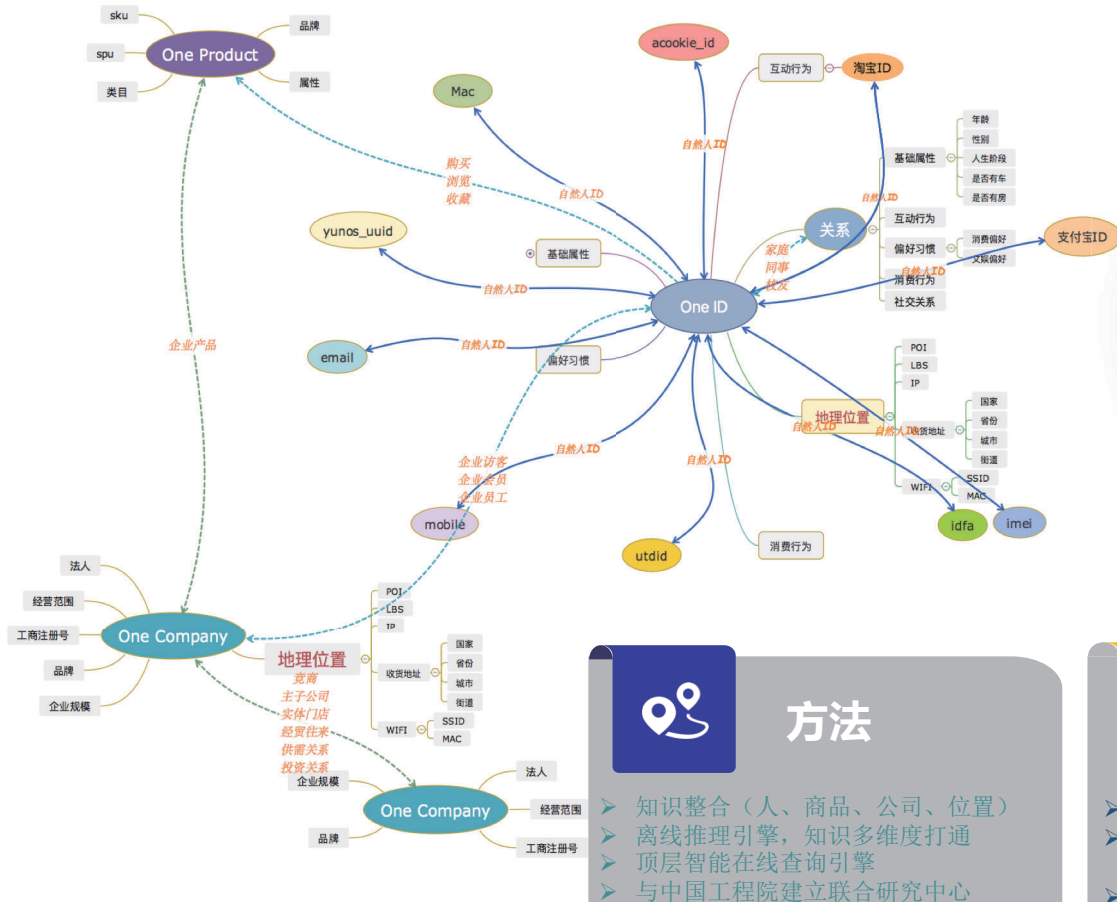


Tim Berners Lee  
WWW创始人  
图灵奖



\*人工智能的两个重要阶段：大规模知识库 + 智能服务

# 数据到知识的全面升级



多维度跨域知识打通融合，做大规模用户参与者的知识获取与推理统一平台

## 常识知识图谱愿景

**方法**

- 知识整合（人、商品、公司、位置）
- 离线推理引擎，知识多维度打通
- 顶层智能在线查询引擎
- 与中国工程院建立联合研究中心

**应用**

- 异构知识网络的交互多度查询
- 相比于现在的单度查询，优势在于一次性多度查询
- 例：某流行商品的消费者所在的行业类型（或朋友圈消费层次）

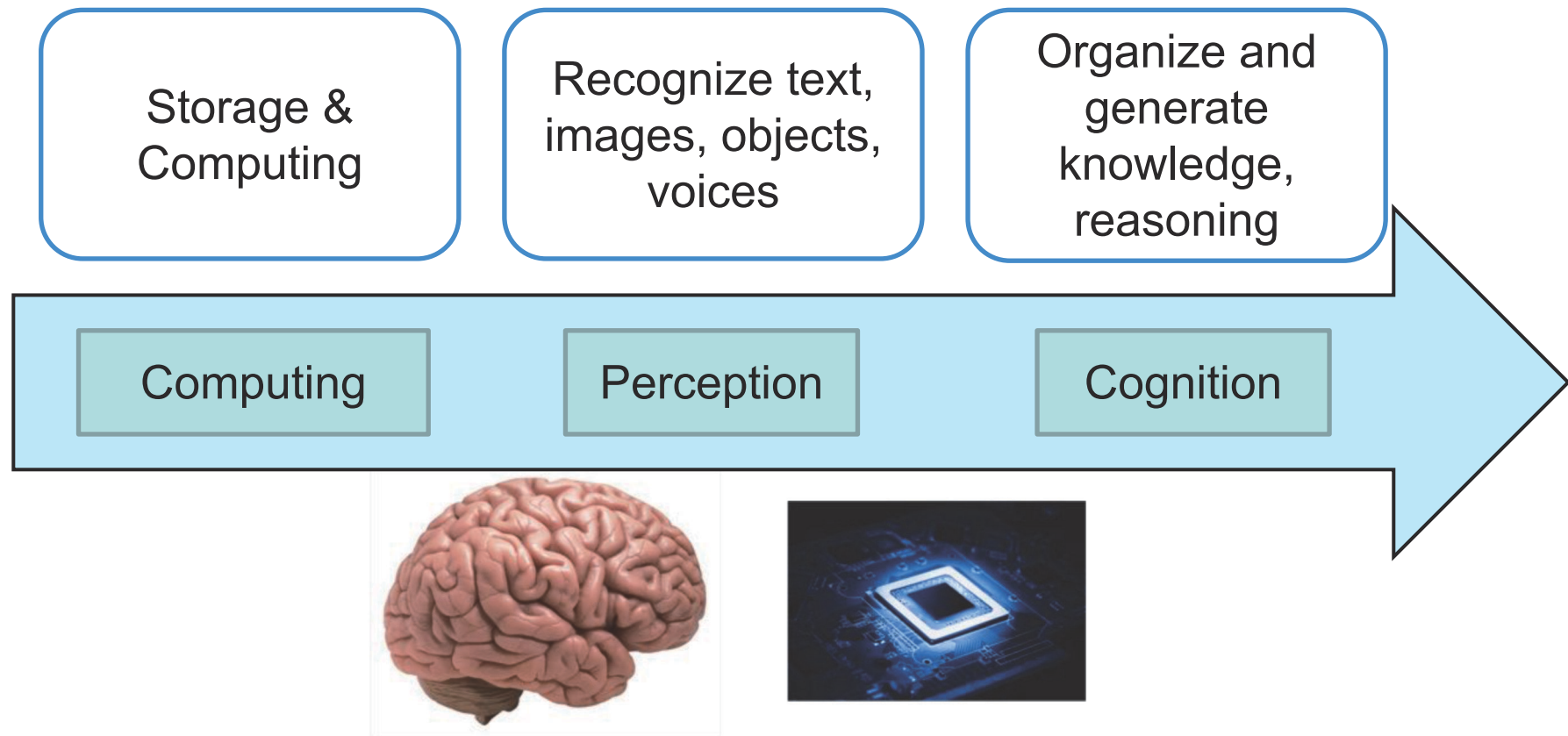
**挑战**

- 需要搜索推荐、知识推理等个性化推荐场景配合，效果提升等相关数据实时回流与反馈。



# AI Trend

- From perceptron to cognition



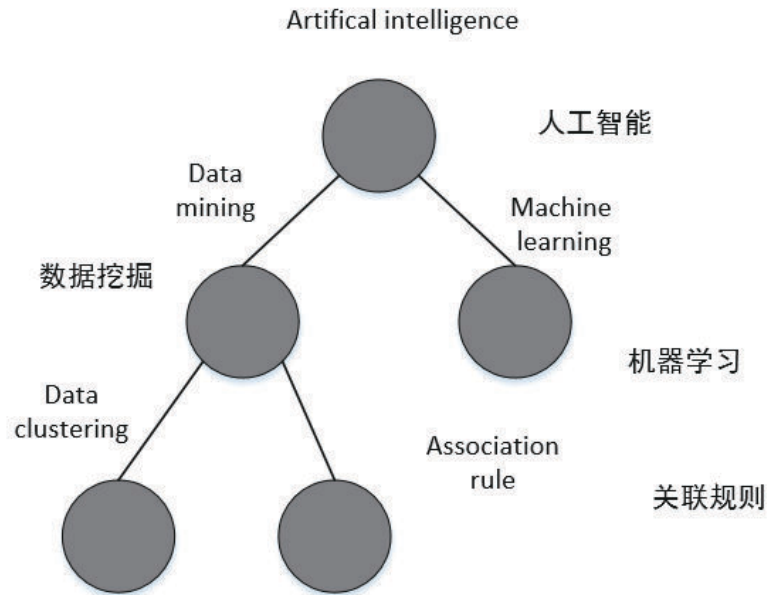
Stochastic vs Deterministic  
*Uncertainty!*

# Data → knowledge → Cognition

## Knowledge Graph:

A graph of concepts

To **organize** knowledge

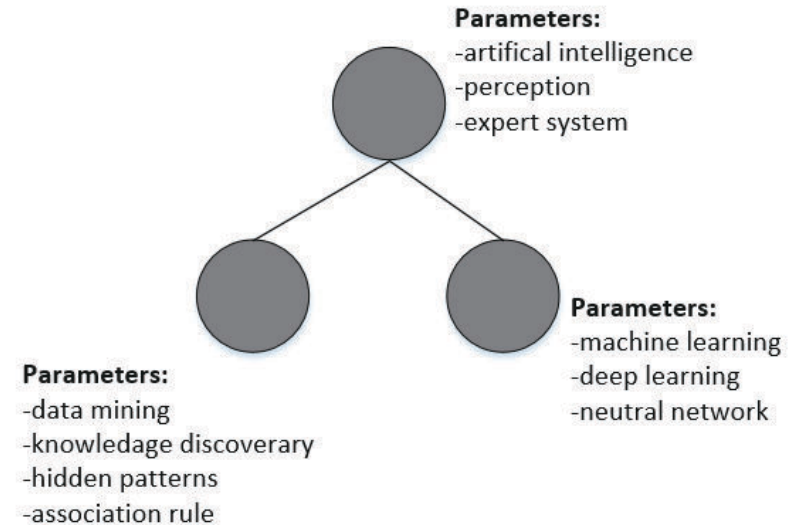


Allow computers to process data **automatically**

## Cognitive Graph:

A graph of generators

To **generate** knowledge



Allow computers to process data **autonomously**



# Thanks

<http://aminer.cn>

