

2018

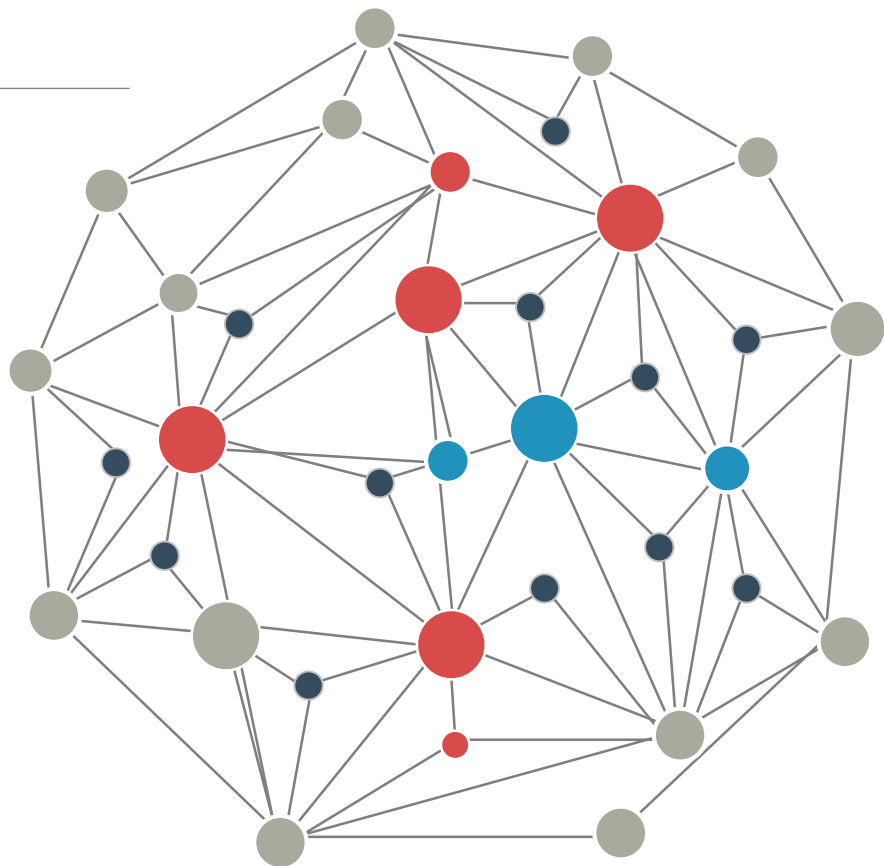
“社交媒体对学术成果传播的提升”论坛

社交网络上的学术信息组织与检索

报告人：吴丹

武汉大学信息管理学院

2018年11月1日



CONTENTS



1. 引言

1.1 学术社交网络概述

1.2 学术社交网络类型

1.3 学术社交网络对于学术传播的重要性

1.4 社交网络上的学术信息获取存在的问题

2. 学术社交网络中的社会化标签组织与利用

2.1 学术社交网络中的社会化标签的规范化

2.2 学术社交网络中的社会化标签的利用

3. 学术社交网络上的信息检索

3.1 学术社交网络中的学习式搜索

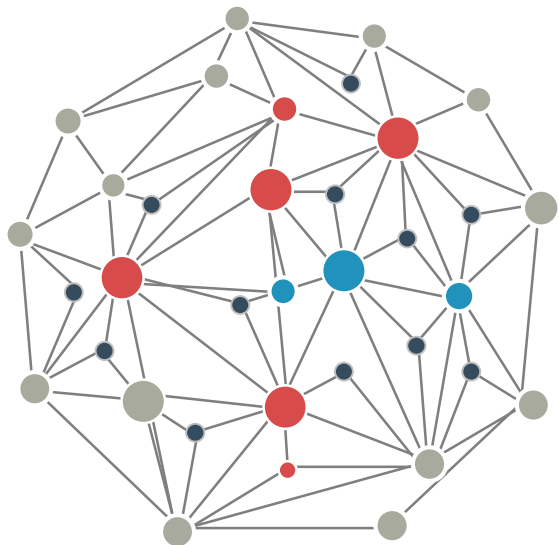
3.2 学术社交网络中的同行发现

4. 展望



武汉大学

Wuhan University



/01

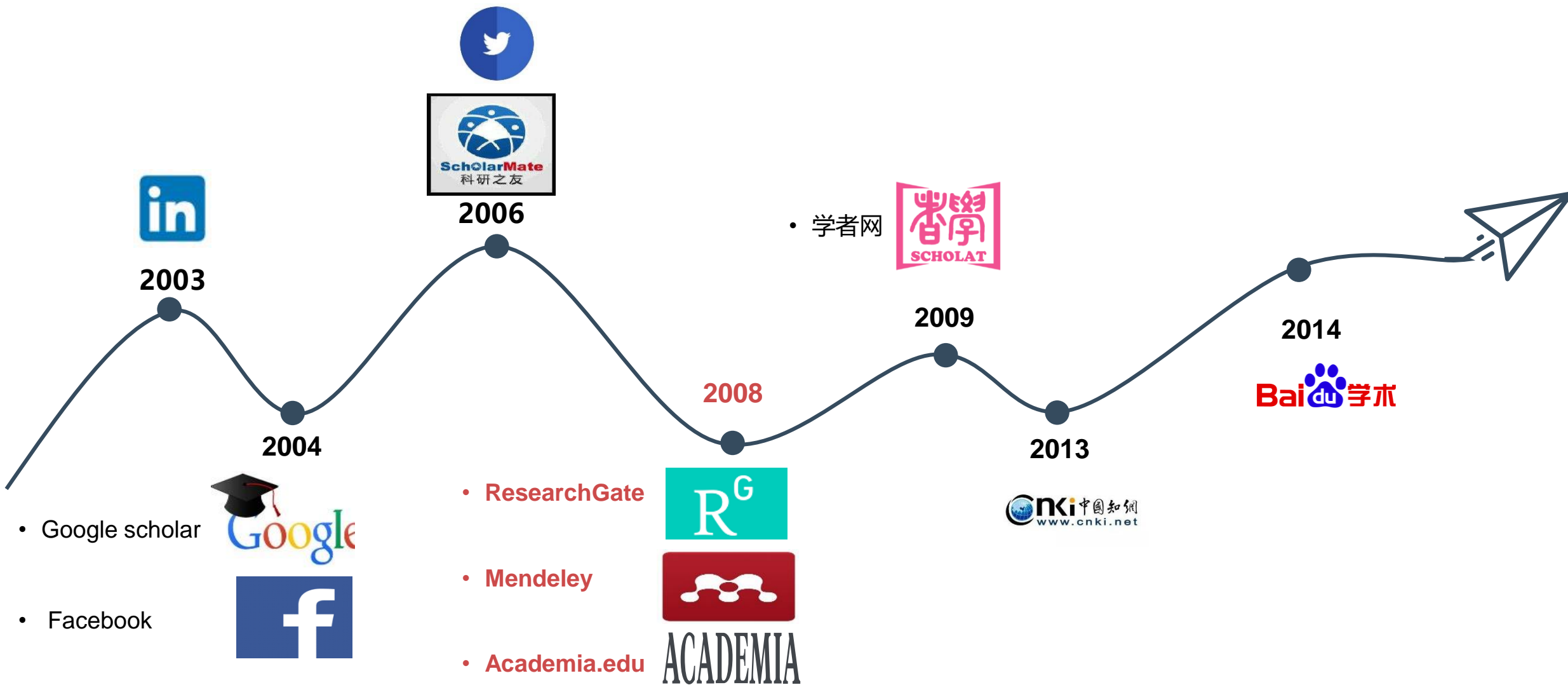
引言



武汉大学

Wuhan University

1.1 学术社交网络概述



1.1 学术社交网络概述

学术社交网络的定义:

- 学术社交网络 (Academic Social Networking, ASN) 是为学者们提供的支持是以研究为导向的在线活动以及社会网络建立为目的的在线服务——Oh J S等
- 以促进知识交流和扩散为目的, 能够帮助研究者建立和维护他们的人际关系网络, 同时能够支持他们在研究过程中从事各项活动的服务或者平台——王日芬等

参考文献:

[1]Oh J S, Jeng W. Groups in Academic Social Networking Services—An Exploration of Their Potential as a Platform for Multi-Disciplinary Collaboration. 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

[2]贾新露,王日芬. 学术社交网络的概念、特点及研究热点[J]. 图书馆学研究,2016(05):7-13.



武汉大学

Wuhan University

1.2 学术社交网络类型

图书分享网络

主要指用于图书/著作的发布、交流和分享的网络社交平台，如LibraryThing、豆瓣等

科研成果社交网络

主要是指用于科研人员或者科研团队之间进行学术成果存储、发现、交流、信息共享、科研项目查询、学术内容分析与可视化等的社交平台。

可分为三类：

- (1) **科研分析工具**，如Profiles、VIVO等
- (2) **科研社交网络平台**，如ResearchGate、CiteULike、Bibsonomy、Mendeley等
- (3) **科研数据库**，如谷歌学术、百度学术学者、CNKI学者圈、Emerald Crescendo等

在线学术社区

主要是指科研人员以博客、论坛、问答等形式进行自我展示和学术交流的学术社区平台，如新浪博客、Facebook、Twitter、知乎、LinkedIn、科学网等。



武汉大学

Wuhan University

1.3 学术社交网络对于学术传播的重要性



- (1) 有利于学者进行学术在线分享，促进同行学术交流
- (2) 有利于学者自我推广，增强学者认知度
- (3) 方便学者进行学术成果管理
- (4) 有利于促进同行学者发现，在网络上构建虚拟学术团队



- (1) 促进学者生产原生内容，促进学术知识的积累与传播
- (2) 能够为学术成果分析软件、文献管理软件或者替代计量提供知识发现的指标和内容，促进科研知识的深度挖掘。
- (3) 促进学术成果的开放存储，增强学术成果的可见度



可以通过学术社交网络认识来自不同专业背景的专家学者，促进跨学科交流，寻找合作伙伴，建立跨学科科研团队，合作攻关科研项目。



武汉大学

Wuhan University

1.4 社交网络上的学术信息获取存在的问题

(1) 学术信息组织

▪ **传统学术数据库**：结构化、主题词受控、组织完整固定

Web of Science

Web of Science interface showing search options and filters. The search bar contains the example query "oil spill* mediterranean". A dropdown menu is open, showing search filters: 主题 (selected), 标题, 作者, 作者识别号, 团体作者, 编者, 出版物名称.

▪ **学术社交网站**：非结构化、个性化（如社会标签）、非受控词、协同组织

ScholarMate interface showing search options and filters. The search bar contains the example query "oil spill* mediterranean". A dropdown menu is open, showing search filters: 主题 (selected), 标题, 作者, 作者识别号, 团体作者, 编者, 出版物名称.

1.4 社交网络上的学术信息获取存在的问题

(2) 学术信息检索

- 学术社交网络上信息检索需求更加复杂多样
(如社交需求、自我展示需求、成果传播需求、学术团队建设需求等)
- 信息组织非结构化，非字段匹配，无法精确检索
- 往往为探寻式检索
- 多语言检索时的术语翻译障碍



学术快讯 Academic news

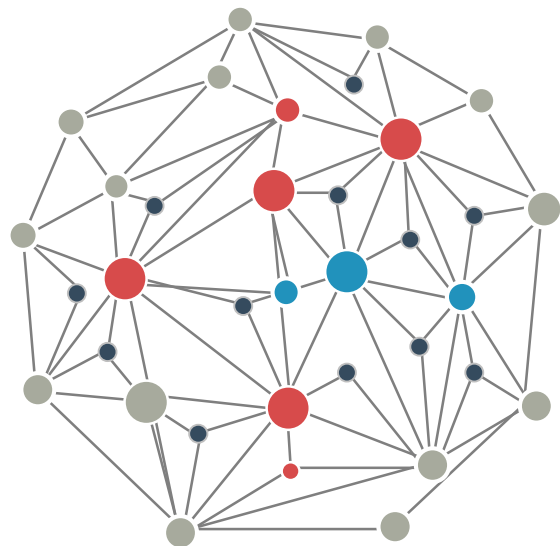
招生 • 招聘 Recruitment

会议 • 征稿 Call for papers



武汉大学

Wuhan University



/02 学术社交网络中的社会化标签组织与利用

- 2.1 学术社交网络中的社会化标签的规范化
 - 2.2 学术社交网络中的社会化标签的利用
-



2.1 学术社交网络中的社会化标签的规范

中英文学术图书标签与主题词表对比实验

实验一

- LibraryThing
- 豆瓣标签

VS

- 国会图书馆主题词表(LCSH)
- 中国分类主题词表(CCT)

中英文学术博客标签与主题词表对比实验

实验二

- CSSCI论文的博客引文
- 博客搜索引擎

VS

- 国会图书馆主题词表(LCSH)
- 中国分类主题词表(CCT)

学术论文标注与学术论文数据库标引对比实验

实验三

- 标注系统CiteULike
- Bibsonomy

VS

- EBSCO数据库中的图书馆、
信息科学与技术摘要数据库



武汉大学

Wuhan University

2.1.1 图书的社会化标签分析

(1) 研究主题：中英文学术图书标签与主题词表对比实验

(2) 研究问题：

- **R1**：当社会标签与传统的受控词汇如国会图书馆主题词表（LCSH）被应用于图情领域中的图书标注时，它们将呈现怎样的关系？
- **R2**：社会标签与传统的受控词汇如国会图书馆主题词表（LCSH）应用在图情领域中的图书标注，英语标注和汉语标注之间有何不同？
- **R3**：在哪些方面可以使得社会标签和专家创建的主题词之间相互补充？



武汉大学

Wuhan University

2.1.1 图书的社会化标签分析

(3) 数据采集:

◆ 获取中英文社会标签

- 英文图书标注系统LibraryThing (<http://www.librarything.com>)
- 中文图书标注系统 豆瓣 (<http://www.douban.com>)。各收集500本书



豆瓣图书标签

分类浏览 / 所有热门标签

文学 ·····

小说(5298252)	外国文学(1937844)	文学(1611318)	随笔(1132086)
中国文学(1049945)	经典(944029)	日本文学(826644)	散文(686198)
村上春树(437578)	诗歌(328662)	童话(292853)	儿童文学(238582)
古典文学(234851)	王小波(223395)	名著(222805)	杂文(216391)
余华(205153)	张鑫玲(190301)	当代文学(149029)	钱钟书(105148)
外国名著(95686)	鲁迅(90101)	诗词(81175)	茨威格(63079)
米兰·昆德拉(53417)	杜拉斯(44135)	港台(7292)	

流行 ·····

漫画(1277390)	推理(938118)	绘本(898125)	青春(655992)
东野圭吾(564932)	科幻(528843)	言情(506635)	悬疑(497968)
奇幻(323915)	武侠(315619)	日本漫画(295640)	韩寒(263675)
推理小说(263256)	耽美(258073)	亦舒(236304)	网络小说(211996)
三毛(208039)	安妮宝贝(173473)	阿加莎·克里斯蒂(155557)	郭敬明(154037)
穿越(153755)	金庸(151437)	科幻小说(147529)	轻小说(144440)
青春文学(120334)	魔幻(116896)	几米(115526)	幾米(98912)
张小娴(97828)	J.K.罗琳(86065)	古龙(76118)	高木直子(72811)
沧月(65929)	校园(61718)	落落(58556)	张悦然(57819)

文化 ·····

豆瓣是国内最大的社交网络和标注网站标签，
其标签体系被称为“全球最精准图书”



武汉大学

Wuhan University

2.1.1 图书的社会化标签分析

◆ 获取相应的中英文主题标引词

- 美国国会图书馆的OPAC检索英文书籍
- 中国国家图书馆的OPAC检索中文书籍

LC Control No.: 99010033
LCCN Permalink: <http://lcn.loc.gov/99010033>
000 01056cam a2200253 a 450
001 4520227
005 20000222113414.0
008 990104s1999 nyua b 001 0 eng
035 __ |9 (DLC) 99010033
906 __ |a 7 |b cbc |c orignew |d 1 |e ocip |f 19 |g y-gencatlg
925 0_ |a acquire |b 2 shelf copies |x policy default
955 __ |a pb04 to ja00 01-04-99; lk09 01-06-99; lj07 01-
010 __ |a 99010033
020 __ |a 020139829X
040 __ |a DLC |c DLC |d DLC
050 00 |a Z667 |b .B34 1999
082 00 |a 025.04 |2 21
100 1_ |a Baeza-Yates, R. |q (Ricardo)
245 10 |a Modern information retrieval / |c Ricardo Baeza-
260 __ |a New York : |b ACM Press ; |a Harlow, England : |
300 __ |a xx, 513 p. : |b ill. ; |c 24 cm.
504 __ |a Includes bibliographical references (p. 455-499)
650 0_ |a Information storage and retrieval systems.
700 1_ |a Ribeiro, Berthier de Araújo Neto, |d 1960-

图a USMARC记录的一个示例

FMT	BK
LDR	01717nam 2200457 450
001	012005032484
010	a 7-111-15878-4 d CNY45.00
100	a 20050421d2005 ekmy0chiy50 ea
1011	a chi c eng
102	a CN b 110000
105	a ak a 001yy
106	a r
2001	a 现代信息检索 A xian dai xin xi jian suo d = Modern information
210	a 北京 c 机械工业出版社 d 2005
215	a xviii, 388页 c 图 d 26cm
2252	a 计算机科学丛书 A ji suan ji ke xue cong shu
410 0	1 2001 a 计算机科学丛书
50010	a Modern information retrieval m Chinese
6060	a 情报检索 A qing bao jian suo
690	a G252.7 v 4
690	a G354.4 v 4
701 1	a 巴埃萨-耶茨 A ba ai sa - ye ci g (Baeza-Yates, Ricardo) 4 著
701 1	a 里贝罗-内托 A li bei luo - nei tuo g (Ribeiro-Neto, Berthier) 4 著

图b CNMARC记录的一个示例



武汉大学

Wuhan University

2.1.1 图书的社会化标签分析

(4) 数据分析方法:

◆ 数据预处理

- 客观标签: 描述一本书的内容或主题的术语 (如 “信息检索”、“网页设计” 等)
- 主观标签: 通常是描述个人主观情感 (如 “喜欢”、“收藏夹” 等)

◆ 将标引词进一步分为与内容相关的标引词 (内容标引词) 和与形式相关的标引词 (形式标引词)

◆ USMARC的6060字段和CNMARC的650字段的所有术语视为内容标引词, 而将除此之外的子字段的所有术语视为形式标引词。

◆ 对于与内容标引词不匹配的客观标签, 直接与LCSH和 CCT进行比较, 任何与主题词表直接匹配的标签也会被记录。



2.1.1 图书的社会化标签分析

(5) 研究发现:

- 标签中使用的词汇类似主题词表中的词汇的使用。
- 但很多标签没有在主题词表中的明确定义的语义关系，用户构建的词汇库可以看做是传统受控词汇的补充。
- 两者的结合将有益于用户词汇和专家词汇。

	同义词(SN)		上位类术语(BT)		下位类术语(NT)		相关术语(RT)		不相关术语(NR)	
	SN的数量	书的数量	BT的数量	书的数量	NT的数量	书的数量	RT的数量	书的数量	NR的数量	书的数量
英文图书	93 (3.0%)	67 (13.4%)	159 (5.1%)	120 (24%)	77 (2.5%)	46 (9.2%)	1209 (38.9%)	286 (57.2%)	1570 (48.5%)	347 (69.4%)
中文图书	263 (13.6%)	202 (40.4%)	212 (10.9%)	164 (32.8%)	197 (10.2%)	139 (27.8%)	620 (32.0%)	339 (67.8%)	645 (33.3%)	289 (57.8%)

Dan Wu, Daqing He, Jin Qiu, Ruonan Lin, Yang Liu. Comparing Social Tags with Subject Headings on Annotating Books: A Study Comparing the Information Science Domain in English and Chinese. Journal of Information Science, 2013, Vol.39, No.2, pp.169-187

2.1.2 学术博客上的社会化标签分析

(1) **研究主题：**以图情领域为例，进行了中英文学术博客标签与主题词表对比实验

(2) **研究问题：**

- R1: 学术博客标注的社会标签与主题词表中的主题词的重合率有多大?
- R2: 中文和英文学术博客社会标签有什么相同和相异之处?

(3) **学术博客获取方式：**

- 博客引文
CSSCI中文期刊论文的博客引文
- 博客搜索引擎
Google Blog (<http://technorati.com/>)
Technorati (<http://technorati.com/>)



+



武汉大学

Wuhan University

2.1.2 学术博客上的社会化标签分析

(4) 实验步骤



武汉大学

Wuhan University

2.1.2 学术博客上的社会化标签分析

(5) 结果分析

◆ 中英文学术博客标签对比分析

相同点:

- ①社会标签客观性较强
- ②社会标签与词表的重合度较低,
规范性均有待提升

差异性:

- ①社会标签数量和出现频次不同
- ②社会标签的重视程度不同



武汉大学

Wuhan University

2.1.3 学术论文的社会化标签分析

(1) **研究主题:** 学术论文社会标注与学术论文的作者关键字、学术论文数据库标引词的对比实验

(2) **研究问题:**

- **R1:** 学术论文标注的社会标签与学术数据库的标引词的重合率有多大?
- **R2:** 学术论文标注的社会标签与作者的关键词的重合率有多大?

ci

Vf

W

Cl

Expi

Be

Anr

Sec

Hig

Gu

Orc

ABSTRACT

We study how the use of online learning systems stimulate cognitive activities, by conducting an experiment with the use of eye tracking technology to monitor eye fixations of 60 final year students engaging in online interactive tutorials at the start of their Final Year Project module. Our findings show that the students' visual scanning behaviours fall into three different types of eye fixation patterns, and the data corresponding to the different types relates to the performance of the students in other related academic modules. We conclude that this method of studying eye fixation patterns can identify different types of learners with respect to cognitive activities and academic potentials, allowing educators to understand their students better. **数据库标引词** using online learning environments to stimulate cognitive activities.

Categories and Subject Descriptors

- Applied computing-Interactive learning environments
- Social and professional topics-Information technology education

General Terms

Experimentation, Measurement

Keywords

Eye Tracking, Human-Computer Interaction, Instructional Design, Cognitive Activity, Online Learning **作者关键词**

1. INTRODUCTION

The continuous advancement of teaching and learning with the increasing use of technology within the vast selections of modules offered in higher education has been evident over the years. Most examples of the advancement are contributed by the increasing use of online learning systems in the learning environment. The rapid

posted to [quantization static output feedback](#) by [corgan](#) on 2018-10-28 11:19:04 ★★

■ Abstract

INSTANT RE
12 MILLION SCHOLAR

oppers.

. 4. (April 2018), pp. 419-433

2774446

2.1.3 学术论文的社会化标签分析

(3) 实验工具选择:

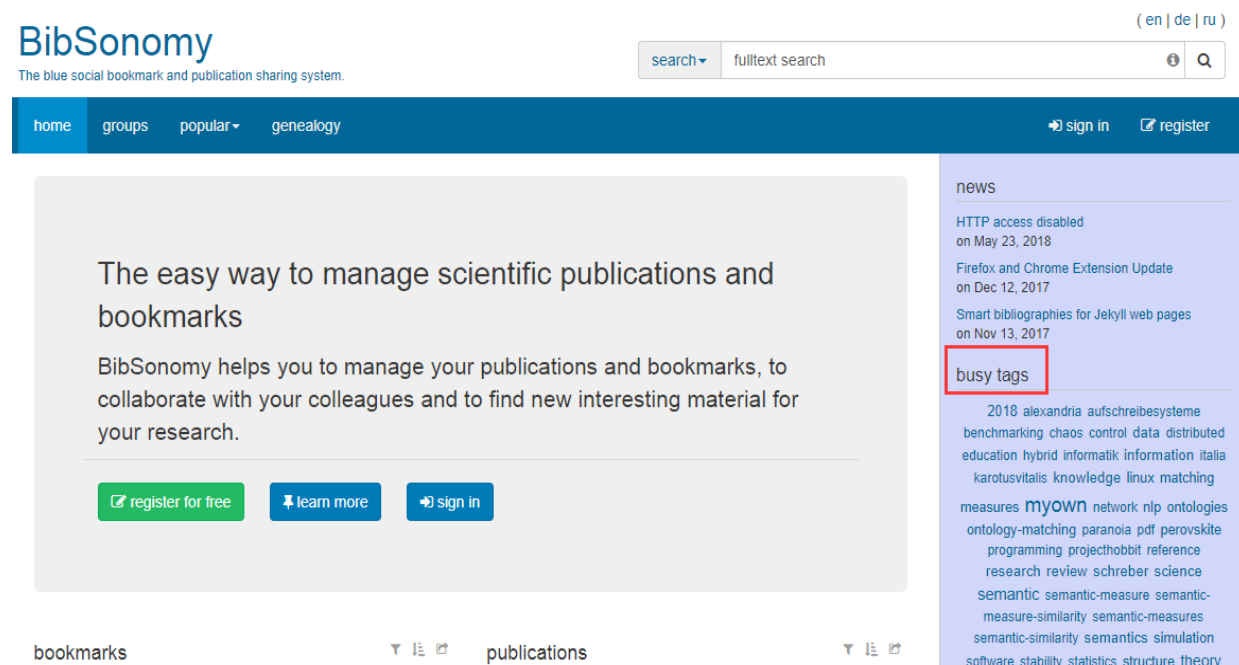
- 英文学术论文标签系统: CiteULike (为主)、Bibsonomy (补充)。
- 学术论文数据库为: EBSCO数据库中的图书馆、信息科学与技术摘要数据库 (Library, Information Science & Technology Abstracts)



citeulike is a free service for managing and discovering scholarly references

8,741,189 articles - 313 added today.

- Easily store references you find online
- Discover new articles and resources
- Automated article recommendations **NEW**
- Share references with your peers
- Find out who's reading what you're reading
- Store and search your PDFs

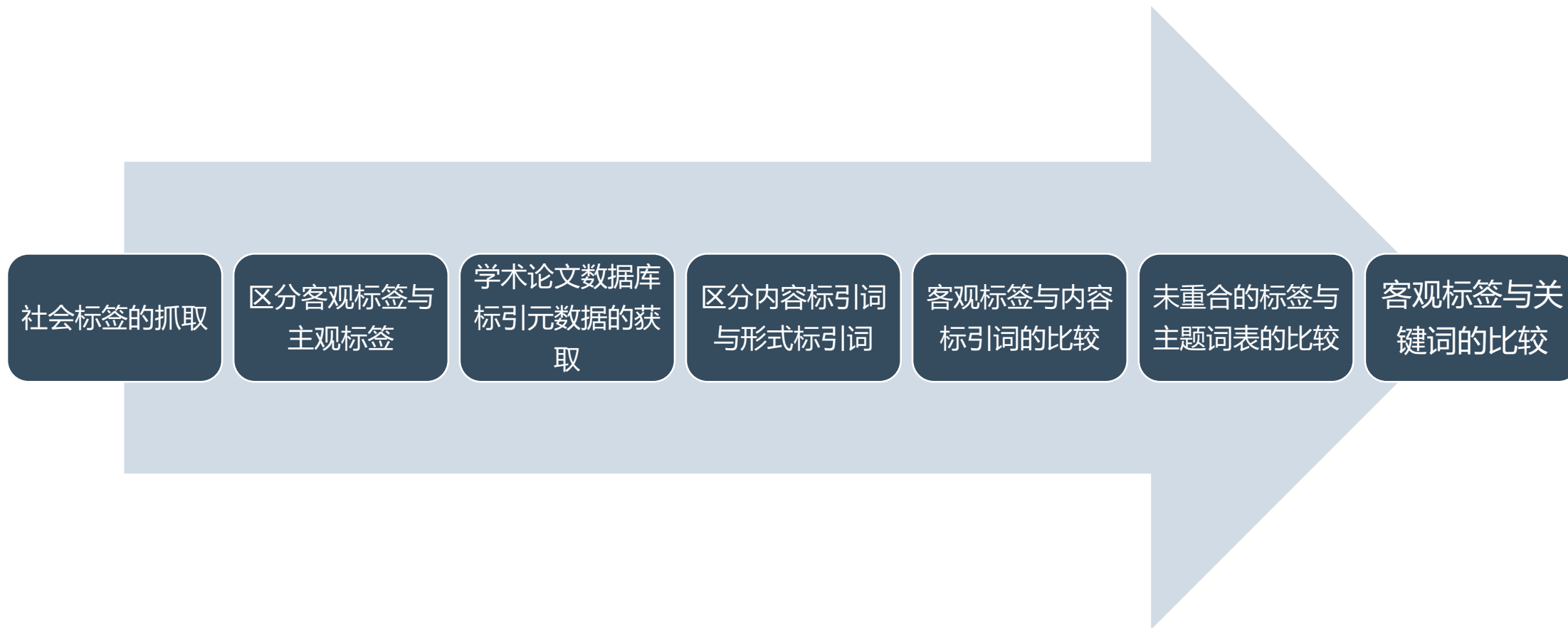


武汉大学

Wuhan University

2.1.3 学术论文的社会化标签分析

(4) 实验步骤



武汉大学

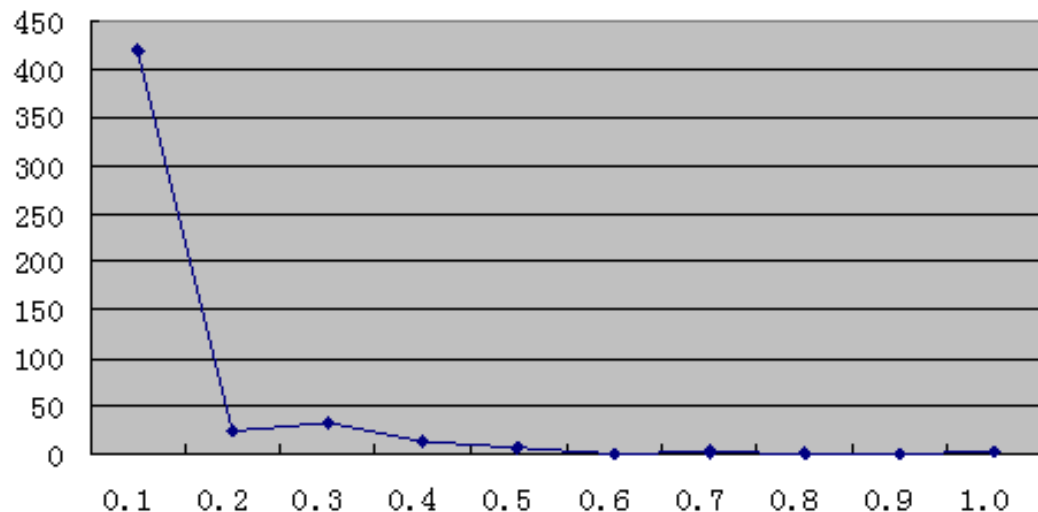
Wuhan University

2.1.3 学术论文的社会化标签分析

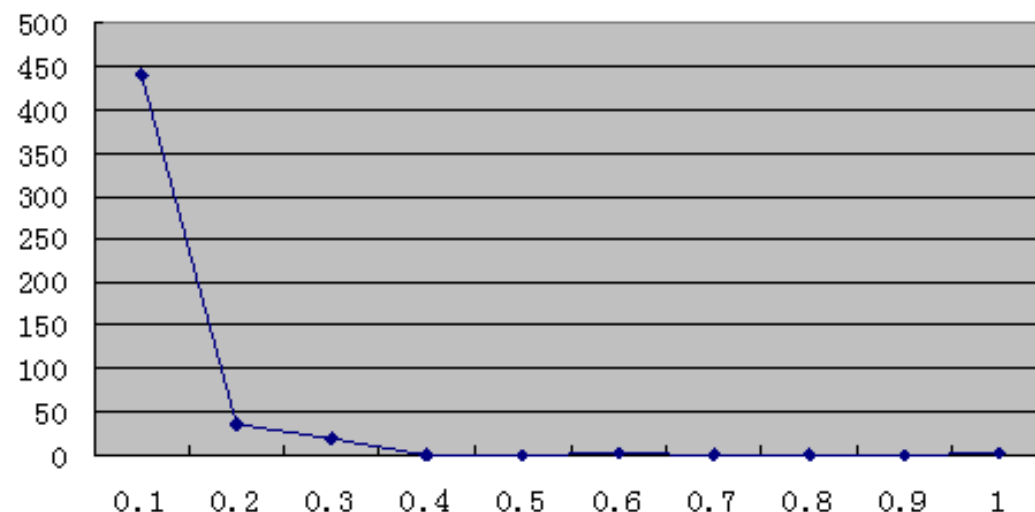
(5) 研究发现:

- ◆ 学术论文标注的社会标签与学术数据库的标引词的重合率

计算Dice系数和Jaccard指数, 值区间 (0, 1), 相关度值越大, 则二者联系越紧密。



(a) 学术论文的客观标签与内容标引词重合的Dice系数分布

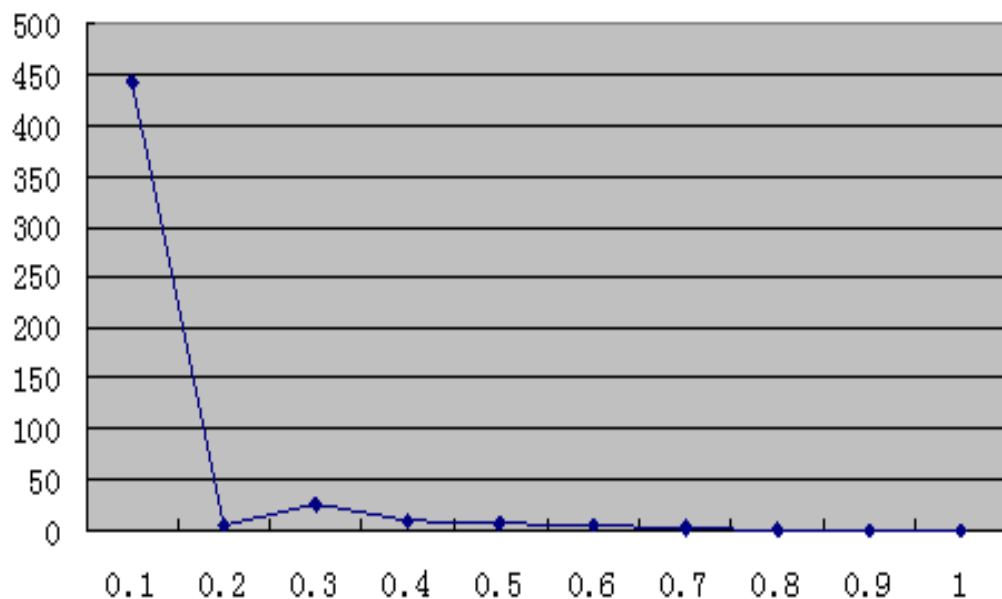


(b) 学术论文的客观标签与内容标引词重合的Jaccard指数分布

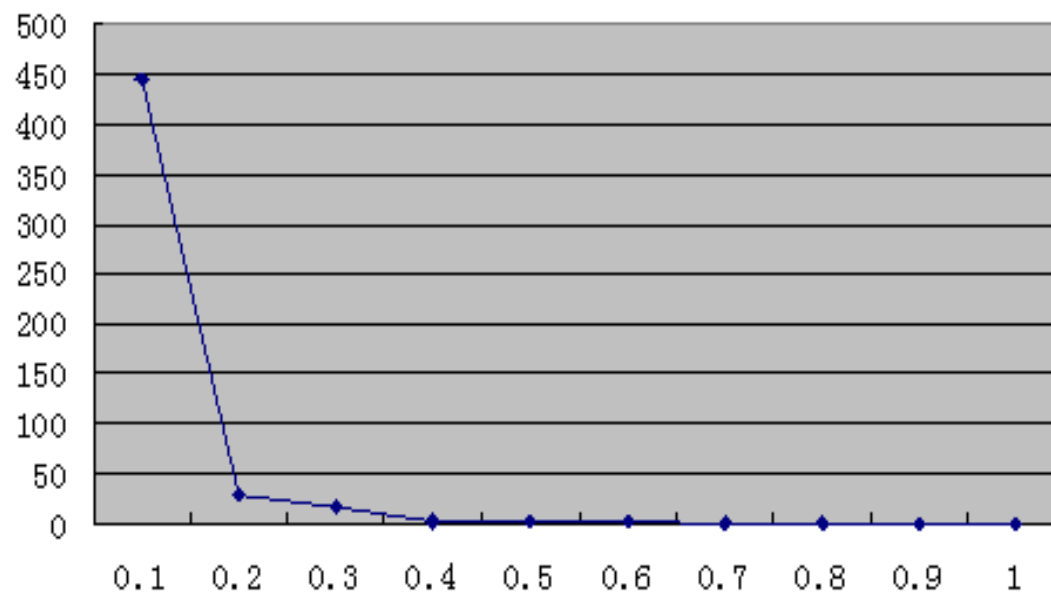
2.1.3 学术论文的社会化标签分析

(5) 研究发现:

- ◆ 学术论文标注的社会标签与作者的关键词的重合率低



(a) 学术论文的客观标签与关键词重合的Dice系数分布



(b) 学术论文的客观标签与关键词重合的Jaccard指数分布

2.1.4 关于学术社交网络标签组织的思考

- ◆ 标签与主题词的重合度比较低，虽然标签的噪音比较大，但是更能代表用户在信息检索中的需求
- ◆ 标签的语义关系应该被充分挖掘，对词表构成补充，应用到学术社交网络的信息检索
- ◆ 中英文标签有较大差异，对不同语种的标签进行映射，应用到多语言学术信息检索

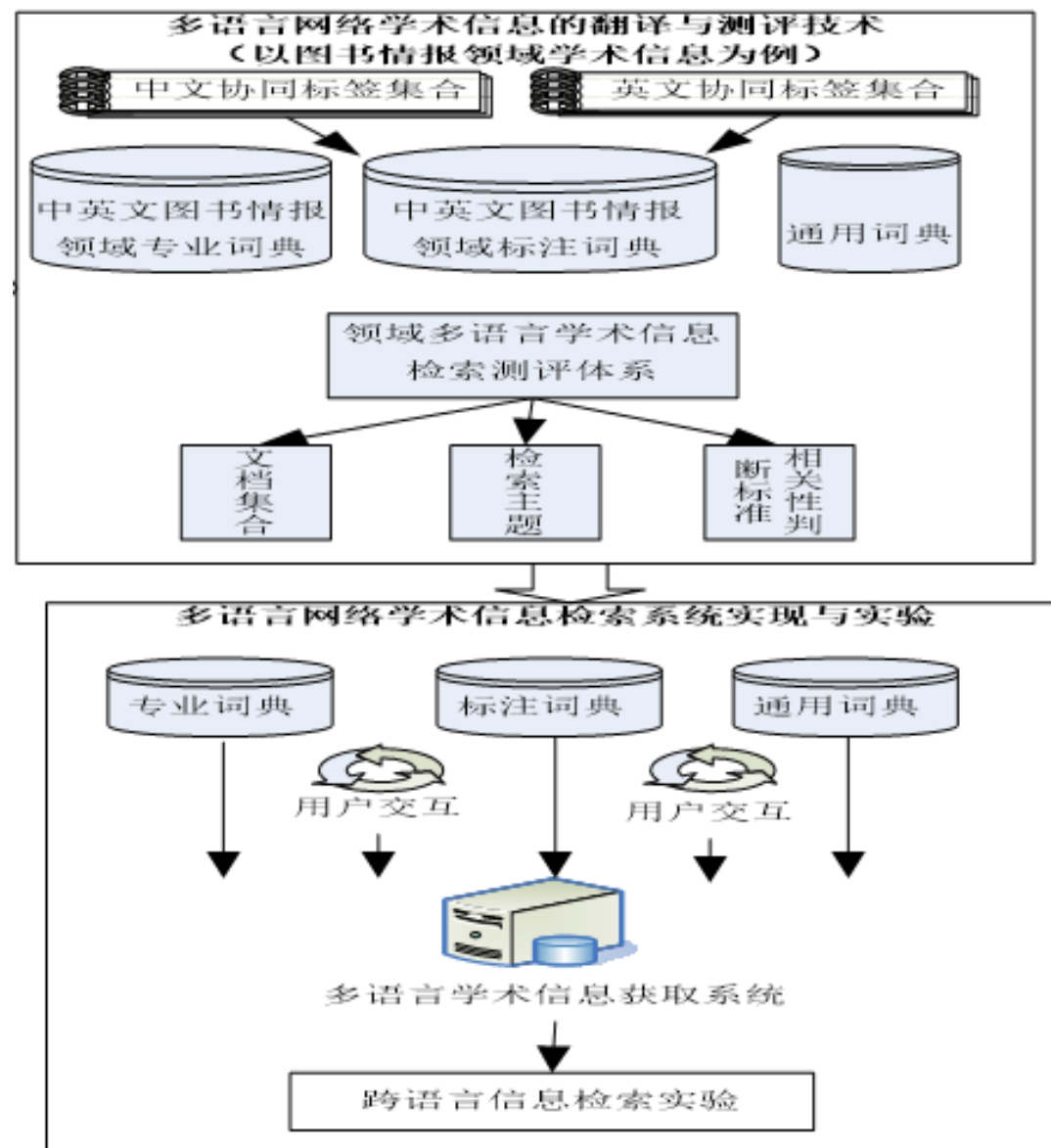
2.2 学术社交网络中的社会化标签的利用

(1) 研究主题: 标签应用于跨语言网络学术信息检索

(2) 研究问题:

- 利用标签构建图书情报领域中英文专业词典
- 将多语言翻译资源应用于图书情报领域跨语言学术信息检索

(3) 技术路线



2.2.1 图书情报领域中英文专业词典构建

- ❖ 选择中国知网的“CNKI翻译助手”构建专业词典
- ❖ 爬取了CNKI翻译助手中与图书情报领域相关的几个类目下的词条及翻译——图书情报与数字图书馆、档案及博物馆、出版等，共计19552条记录
- ❖ CNKI翻译助手的改造
 - 词典预处理
 - 候选翻译词根还原
 - 翻译概率的计算
 - 词典后处理

最终图书情报领域汉英专业机读词典（带翻译概率），中文词条共有19552条



CNKI 翻译助手

全文文献 工具书 数字 学术定义 翻译助手 学术趋势 更多

搜索

学术翻译必备词汇:

[综合]	现代(modern)	对照(group)	小学(primary)	更多...
[中国近...]	政府(government)	时期(during)	社会(society)	更多...
[初等教育]	小学(schools)	英语(english)	教学(students)	更多...

更多学术翻译必备词汇

近期中文热门查询词汇:

[综合]	程序	生物质	语言	儿童	更多...
[有机化工]	聚氨酯	聚丙烯	环氧树脂	阻燃	更多...
[农作物]	不育	亲本	单株	配合力	更多...

更多中文热门查询词汇

近期英文热门查询词汇:

[综合]	combustion	capital	numerical	constitution	更多...
[环境科...]	leachate	landfill	coagulation	influent	更多...
[行政法...]	reconsideration	license	omission	alien	更多...

更多英文热门查询词汇

缩略语:

· dv	· ACAD	· DCS
· FEM	· IAEA	· ZPS
· WA	· OOM	· SR

更多缩略语查询

```
1 高层管理者 senior manag//0.470590 top manag//0.294120 the top manag//0.235290 record19300.xml
2 医学图书情报学 medic//0.643203 book//0.137745 librari//0.083199 the book//0.012339 a medic//0.
3 闲暇时间 leisur time//0.686490 spare time//0.124320 free time//0.097300 more leisur time//0.09
4 素质建设 qualiti build//0.346158 qualiti develop//0.207693 literaci construct//0.138465 constr
5 核心计划 core plan//0.500000 core program//0.500000 record12241.xml
6 现刊记到 regist of the present journal//1.000000 record13408.xml
7 成果信息 achiev inform//0.950000 achiev data//0.050000 record9362.xml
8 瓦尔登湖 lake//0.896220 jezioro//0.020000 lk//0.020000 lakeminnow//0.020000 warden//0.020000
9 大学图书馆建设 univ librari construct//1.000000 record7196.xml
10 一稿多投 multipl contribut with on manuscript//0.240003 multi contribut of on paper//0.180000 r
11 地方高校图书馆 local univ librari//0.587754 local colleg librari//0.183672 librari in local ur
12 辅助决策 assist deci make//0.202882 assist deci//0.154098 aid deci//0.144403 deci sup port//0.
```

2.2.2 图书情报领域中英文标注词典构建



得到图书情报领域汉英标注词典（带翻译概率，并全部由社会标签构成） 中文词条806条

```
1 4G技术 4g//1.000000
2 低功耗IPv6无线个人区域网络 6lowpan//0.500000 ipv6 over low power wireless person area network//0.500000
3 英美编目规则第二版 aacr2//0.500000 anglo american catalogu rule second edit//0.500000
4 图书馆馆长日 aaldai//0.500000 ask a librarian dai//0.500000
5 学术服务台 academ desk//1.000000
6 学术图书馆 academ librari//1.000000
7 学术搜索引擎优化 academ search engin optim//1.000000
8 学术搜索引擎垃圾 academ search engin spam//1.000000
9 学术图书馆员 academ librarianship//0.500000 academ librarian//0.500000
10 大学与研究图书馆协会 associ of colleg and research librari//0.500000 acrl//0.500000
11 DNS域名动态目录 activ directori dn domain name//1.000000
12 自动数据采集 adc//0.500000 automat data captur//0.500000
```



武汉大学

Wuhan University

2.2.3 图书情报领域多语言学术信息的检索测评体系构建

❖ (1) 研究思路:

- 参照TREC测评体系的构建模式，以图书情报领域为例，构建一个该领域的学术信息文档集合
- 设计一套该领域的检索主题，通过该领域的专家针对每个检索主题对文档集合检索出的结果进行相关性判断，建立一套图书情报领域标准的测试集与评价标准

❖ (2) 学术信息文档集合

- 选取ISI的Journal Citation Report中社会科学版下的“信息学与图书馆学（Information Science & Library Science）”类目下的全部66种期刊
- 下载从1956-2010年的共207320篇文章的题录信息，每篇文章题录信息包括51个字段



武汉大学

Wuhan University

2.2.3 图书情报领域多语言学术信息的检索测评体系构建

❖ (3) 检索主题设计

- 50个图书情报领域的检索主题。
- 检索主题从中文版设计开始，再翻译成英文。
- 设计者在设计检索主题时，到SSCI数据库里试检索一下，取那些检索结果在40-200篇左右文献的检索主题为合适的

Field	Value
Number:	1
Title:	library 2.0
Narrative:	Library 2.0 is a loosely defined model for a modernized form of library service. It reflects a transition within the library world that services are delivered to users. The focus is on user-centered change and participation in the creation of content and community. I want to know the mechanism how Library 2.0 encourages feedback and participation in the design and implementation of library services, and the advantage and disadvantage debate surrounding it.
Related terms:	Learning 2.0, Web 2.0, User-generated content, social network
Related authors:	MILOSAVLJEVIC, B., SURLA, D., CHAWNER, B., AHARONY, N., BATTENFELD, RL., BOOTH, C
Related author affiliations:	UNIV NOVI SAD, VICTORIA UNIV WELLINGTON, LONG ISL UNIV, UNIV COLL DUBLIN, UNIV LOUGHBOROUGH, UNIV POLITECN VALENCIA
Related source journals:	ELECTRONIC LIBRARY, PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS, PROFESIONAL DE LA INFORMACION, LIBRARY HI TECH, LIBRARY JOURNAL, JOURNAL OF ACADEMIC LIBRARIANSHIP
Years to be covered:	1990-2010
Any related information:	1. Library 2.0, Casey ME, Savasthuk LC, LIBRARY JOURNAL, Volume: 131, Issue: 14, Pages: 40-42, Published: 2006.2. Electronic journal forum - Web 2.0, library 2.0, and the hyperlinked library, Stephens M, SERIALS REVIEW, Volume: 33, Issue: 4, Pages: 253-256, Published: 2007.3. 9TH ANNUAL CONFERENCE OF THE LIBRARY-AND-INFORMATION-ASSOCIATION-OF-SOUTH-AFRICA (LIASA).

2.2.4 多语言网络学术信息检索系统实现与实验研究

(1) 系统构建:

图书情报领域的多语言信息获取系统

Multilingual Information in LIS:

Knowledge, Translation, Evaluation and

Access (简称**Milk-Tea**)



SCHOOL OF INFORMATION MANAGEMENT
WUHAN UNIVERSITY, P.R.CHINA

*Multilingual Information in LIS: Knowledge,
Translation, Evaluation and Access*

Milk Tea

WUHAN UNIVE

User ID: Password: Login



武汉大学

Wuhan University

2.2.4 多语言网络学术信息检索系统实现与实验研究

(2) 研究目的:

- 比较三种翻译资源效果

将“通用词典”作为基准 (Baseline), “专业词典+通用词典”和“专业词典+标注词典+通用词典”作为比较的对象, 检验我们构建的两个领域语义词典在图书情报领域多语言网络学术信息检索中的作用。

- 检验不同长度查询式对多语言学术信息检索的影响。

The screenshot displays a search interface with a top navigation bar containing 'Search', 'History', and 'Logout'. Below the navigation bar, there are sections for '检索条件:' (Search Criteria) and '查询语句:' (Query Statement).

检索条件:

- 词典: 专业+标注+通用
- 专业词典阈值: 1.0
- 标注词典阈值: 1.0
- 通用词典阈值: 0.5
- 检索结果数: 200

查询语句:

```
#combine( #wsyn(0.767921 #1(digit librari) 0.13698 #1(the digit librari) 0.025 #1(digit librari intellectu properli) 0.025 #1(school librari digit librari) 0.014517 #1(a digit librari) 0.011178 #1(digit librari develop) 0.008739 #1(digit librari system) 0.0027 #1( numer librari) 0.001539 #1(figur librari) 0.001287 #1(digit on) 0.0009 #1(d librari) 0.000639 #1(intellig digit librari) 0.000639 #1(peke univ librari) 0.000639 #1(electron librari) 0.000513 #1(the number librari) 0.000387 #1(digit base librari) 0.000261 #1(digit librari utili) 0.000261 #1(bit librari) 0.000261 #1(digit book) 0.000261 #1(arithmet figur librari) 0.000126 #1(comput librari) 0.000126 #1(digit librari analyz) 0.000126 #1(net librari).ti)
```

检索结果: 执行此检索共花费3.766秒

1: [170040.xml](#)

Title: Digital libraries '94: The First Annual Conference on the Theory and Practice of Digital Libraries, June 19-21, 1994, College Station, Texas - Schnase, JL 数字图书馆 94: 理论与实践的数字图书馆, 1994年6月19日至21日, 德克萨斯大学站-Schnase, JL 第一次年度会议

Abstract: None

Author: Liebscher, P

Not Judged (selected) Not Relevant Somewhat Relevant Highly Relevant

2: [054217.xml](#)

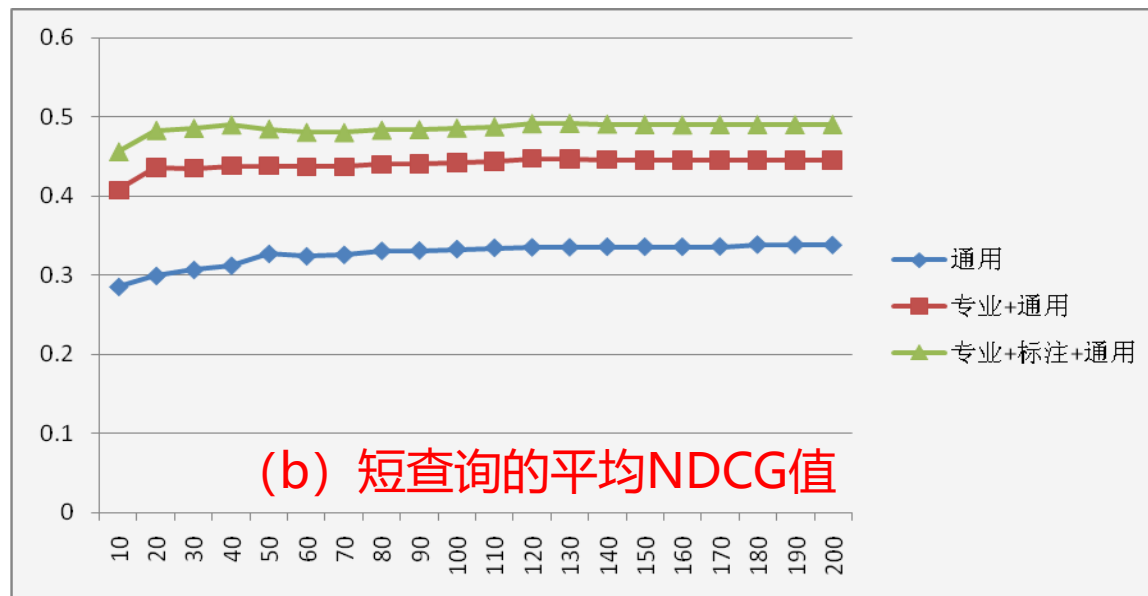
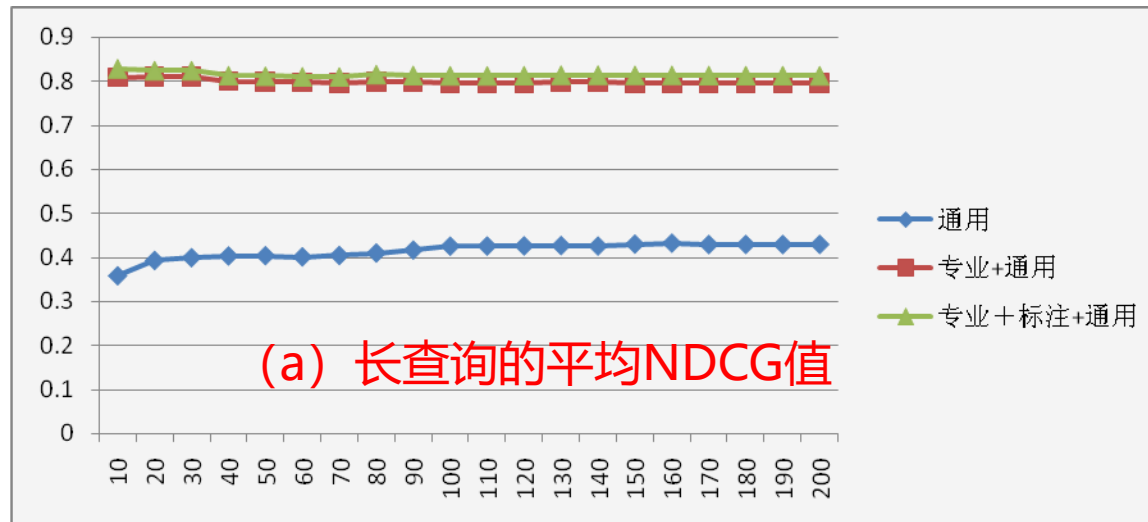
Title: Resource sharing digital libraries: a case study of Taiwan's InfoSpring Digital Library Project 资源共享数字库: 台湾的InfoSpring数字图书馆工程为例

Abstract: The InfoSpring Digital Library Project conducted by National Chiao Tung University (NCTU) is a project to establish mirror sites in Taiwan for well-known reference databases and full-text electronic journals. The role of NCTU in this project is to provide the hardware (including servers and storage space) and network infrastructure, system maintenance and data update. By means of resource sharing, universities and industries in Taiwan can access the databases loaded in the InfoSpring Digital Library under certain subscription agreements. Realizing the importance of resource sharing, libraries in Taiwan have organized a consortium to decide which databases to be hosted in Taiwan and negotiate competitive prices with information providers. This paper presents the motivation, current status, experiences, and future directions of the InfoSpring Project and introduces the Taiwanese InfoSpring 数字库项目进行的国家交通大学(交大)是一个项目, 建立镜像站点在台湾知名参考数据库和全文电子期刊。交大在此项目中的作用是提供(包括服务器和存储空间)的硬件和网络基础设施、系统维护和数据更新。资源共享, 大学和台湾产业可以访问 InfoSpring 数字库中加载某些订阅协议下的数据库。实现资源共享的重要性, 在台湾图书馆举办了财团来决定哪些寄存在台湾和谈判具有竞争力的价格, 与供应商信息的数据库。本文给出的动机、现状、经验和未来方向的 InfoSpring 项目, 并介绍了台湾图书馆财团的IT服务。(C) 2000 Elsevier 科学有限公司保

2.2.4 多语言网络学术信息检索系统实现与实验研究

(3) 实验分析:

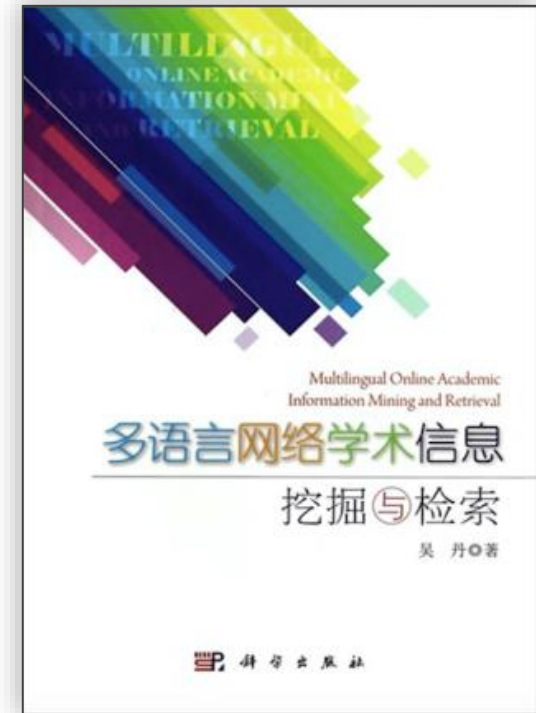
- 采用NDCG作为检索评价指标
- 短查询为检索表单中的“问题标题”；长查询为检索表单中的“问题标题”和“相关检索词”。
- 对50个检索主题全部按照短查询与长查询送入每个翻译资源所形成的系统，最终共获得2（2种长度的查询式）*3（3种翻译资源）*50（50个检索主题）=300个检索结果列表。系统设置为每次检索返回200条结果。
- 按照检索结果列表的前10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200个检索结果统计了50个检索主题的平均NDCG值。



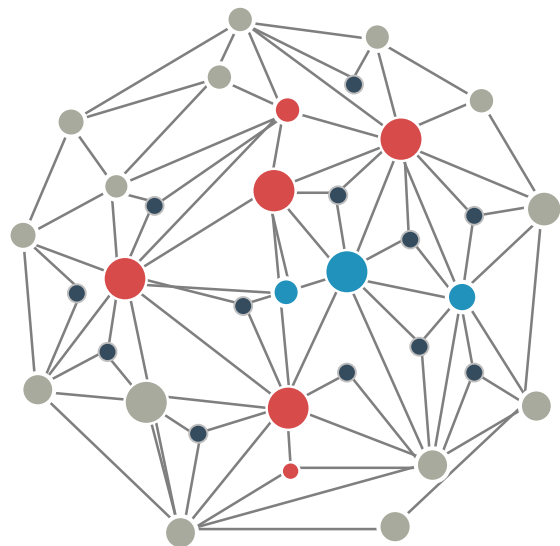
2.2.4 多语言网络学术信息检索系统实现与实验研究

(4) 实验结果:

- 短查询 OR 长查询: 无论在检索结果列表的什么位置, “专业词典+标注词典+通用词典” 以及 “专业词典+通用词典” 都比 “通用词典” 取得了显著性差异
- 相比短查询和长查询, 当查询式较短的时候, 系统所获得的 NDCG值较高, 而短查询是更接近于用户的真实查询。因此, Milk-Tea系统能够支持真实用户的检索, 其性能较好。
- 在多语言学术信息检索时, 标注词典是对专业词典的有益补充



吴丹著. 多语言网络学术信息挖掘与检索. 北京: 科学出版社, 2014.10



/03 学术社交网络上的信息检索

3.1 学术社交网络中的学习式搜索

3.2 学术社交网络中的同行发现



武汉大学

Wuhan University

3 学术社交网络上的信息检索

◆ 社交网络逐渐成为学术信息检索的方式之一

数据库导航

数据库名称: 查找

字顺排序: 中文库: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
外文库: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

学科推荐库: 法学 类型: 百科全书

常用数据库: 中国知网 万方 维普 读秀 新东方 CSSCI 超星电子书 北大法意网 武大博硕士论文库
WOS核心合集(含SCIE、SSCI、A&HCI、CPCI数据库) EI Elsevier Wiley Taylor & Francis 更多

中文数据库	外文数据库
<input type="checkbox"/> AiritiLibrary台湾学术文献数据库	<input type="checkbox"/> 17th and 18th Century Burney Collection Newspapers
<input type="checkbox"/> Apabi电子教学参考书	<input type="checkbox"/> AAS美洲回溯典藏文献数据库 NEW!
<input type="checkbox"/> CADAL (中英文图书数字化国际合作计划)	<input type="checkbox"/> ACM (Association for Computing Machinery)
<input type="checkbox"/> CALIS联合目录(原: 高校联合目录)	<input type="checkbox"/> ACS (American Chemical Society)
<input type="checkbox"/> CALIS学位论文中心服务系统	<input type="checkbox"/> AGU (American Geophysical Union) 电子期刊和图书
<input type="checkbox"/> EPRS全球统计数据/分析平台	<input type="checkbox"/> AIAA (The American Institute of Aeronautics and Astronautics)
<input type="checkbox"/> 丝路图情	电子期刊、会议录和电子图书
<input type="checkbox"/> KUKE数字音乐图书馆	<input type="checkbox"/> AIP (American Institute of Physics)
<input type="checkbox"/> NoteExpress文献管理软件	<input type="checkbox"/> Annual Reviews
<input type="checkbox"/> SPSS Statistics 统计分析软件	<input type="checkbox"/> APA (American Psychological Association)
<input type="checkbox"/> 百度教育云平台 NEW!	<input type="checkbox"/> APS (American Physical Society)
<input type="checkbox"/> 北大法宝	<input type="checkbox"/> APS (American Physiological Society)
<input type="checkbox"/> 北大法意网	<input type="checkbox"/> ASC (Academic Search Complete)
<input type="checkbox"/> 博看网畅销报刊阅读平台	<input type="checkbox"/> ASCE (The American Society of Civil Engineers)
<input type="checkbox"/> 测绘科技文摘数据库	<input type="checkbox"/> ASM (American Society for Microbiology)
<input type="checkbox"/> 超星法源卓越法律人才学习平台	<input type="checkbox"/> ASME (American Society of Mechanical Engineers)



领英 in 搜索

武汉大学 Wuhan University

Mendeley上建立自己的学术社交网络
ResearchGate上查找某位专家学者的学术成果

已关注 查看校友

LinkedIn搜索专业匹配的职位或者学术人脉资源

关于我们

Wuhan University is a comprehensive and key national university directly under the administration of the Ministry of Education. It is also one of the "211 Project" and "985 Project" universities with full support in the construction and development from the central and local government of China.

Since its establishment, Wuhan University has cultivated more than 300 thousand professional talents in various occupations, among whom there are over 100 members of the Chinese Academy of Science and the Chinese Academy of Engineering. They have made great contribution to the national construction and social advancement. The remarkable achievements of Wuhan University have won itself an extensive international reputation. In 1999, the world renowned journal Science listed Wuhan University as one of the most prominent institutions of higher education in China.

更多

82,478+ 位校友的就业去向

居住地区	就职公司
65,875 中国	902 华为
7,846 美国	456 腾讯
6,791 中国 北京	235 Baidu Inc.

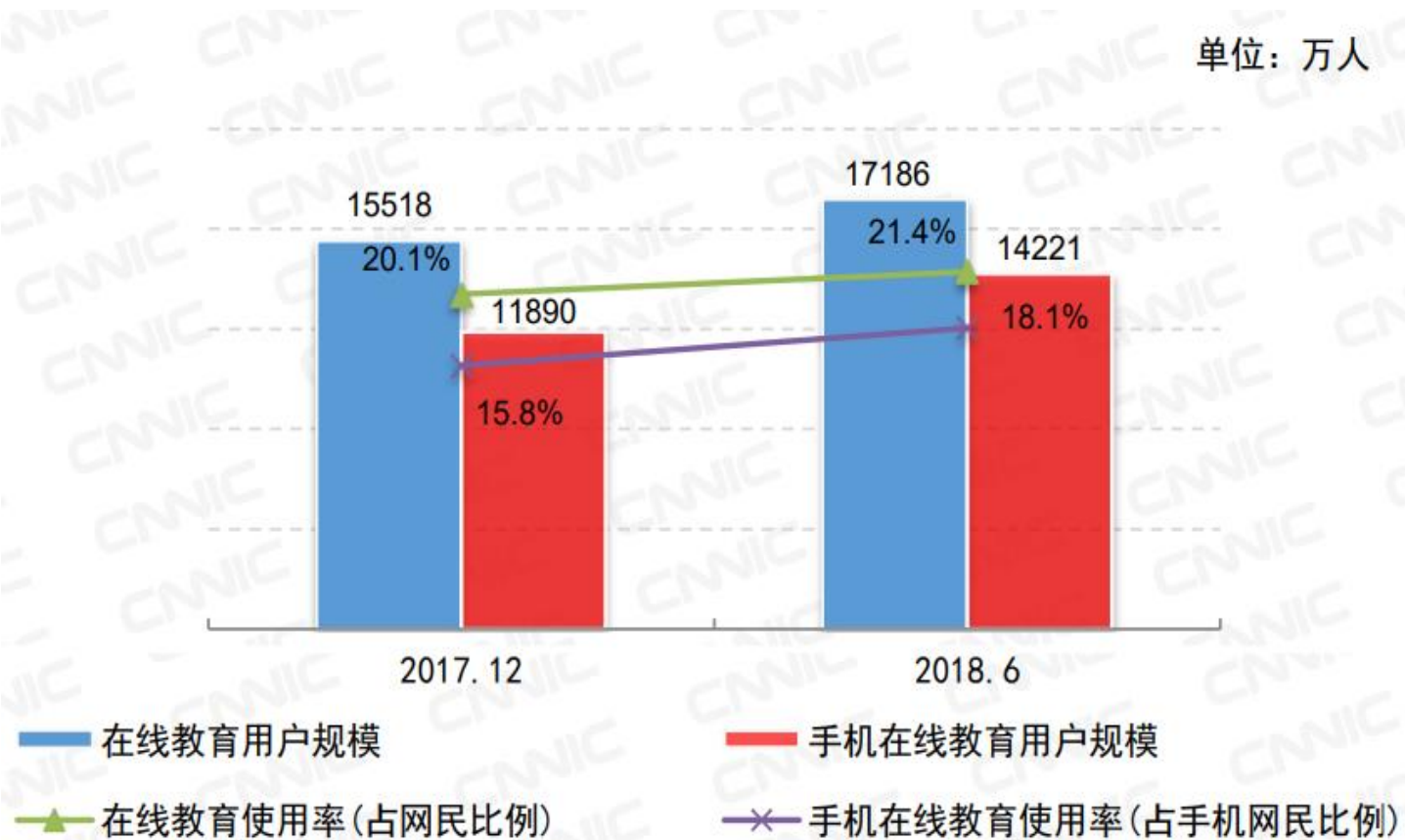
3 学术社交网络上的信息检索

◆ 学术社交网络信息检索研究的主要内容：

- 社交网络信息检索的使用意愿
- 社交网络检索的信息类型
- 社交网络中信息检索系统的优化设计
- 社交网络中信息检索的个性化推荐研究

- 学术社交网络用户搜索行为研究
 - 学习式搜索
 - 同行用户搜寻行为

3.1 学术社交网络中的学习式搜索



2018年



在线学习人数增长
10.7%

2017年

2017.12-2018.6在线教育/手机在线教育用户规模及使用率

数据来源：CNNIC第42次《中国互联网络发展状况统计报告》



武汉大学

Wuhan University

3.1 学术社交网络中的学习式搜索

◆ 学习式搜索概念：

- 用户利用信息检索系统获取信息和学习知识的方式称之为“学习式搜索” (search as learning)，例如科研人员在开展研究之前通过对该领域的文献的调研发现当前研究现状和研究热点。
- 学习式搜索更加注重用户的学习结果，是以学习目标导向的搜索行为，其结果不仅仅注重检索结果，更加注重用户在搜索过程中的知识构建。

◆ 学术社交网络中的学习式搜索：

即用户利用学术社交网络平台开展的学习式搜索，如ResearchGate、Mendeley、谷歌学术学者、百度学术学者等。



武汉大学

Wuhan University

3.1 学术社交网络中的学习式搜索

(1) 研究问题:

RQ1: 用户在学术社交网络中进行学习式搜索的动机是什么?

RQ2: 学术社交网络中学习式搜索的过程是如何进行的?

RQ3: 学术社交网络中学习式搜索是如何改变用户的知识结构的?



武汉大学

Wuhan University

3.1 学术社交网络中的学习式搜索

(2) 研究内容

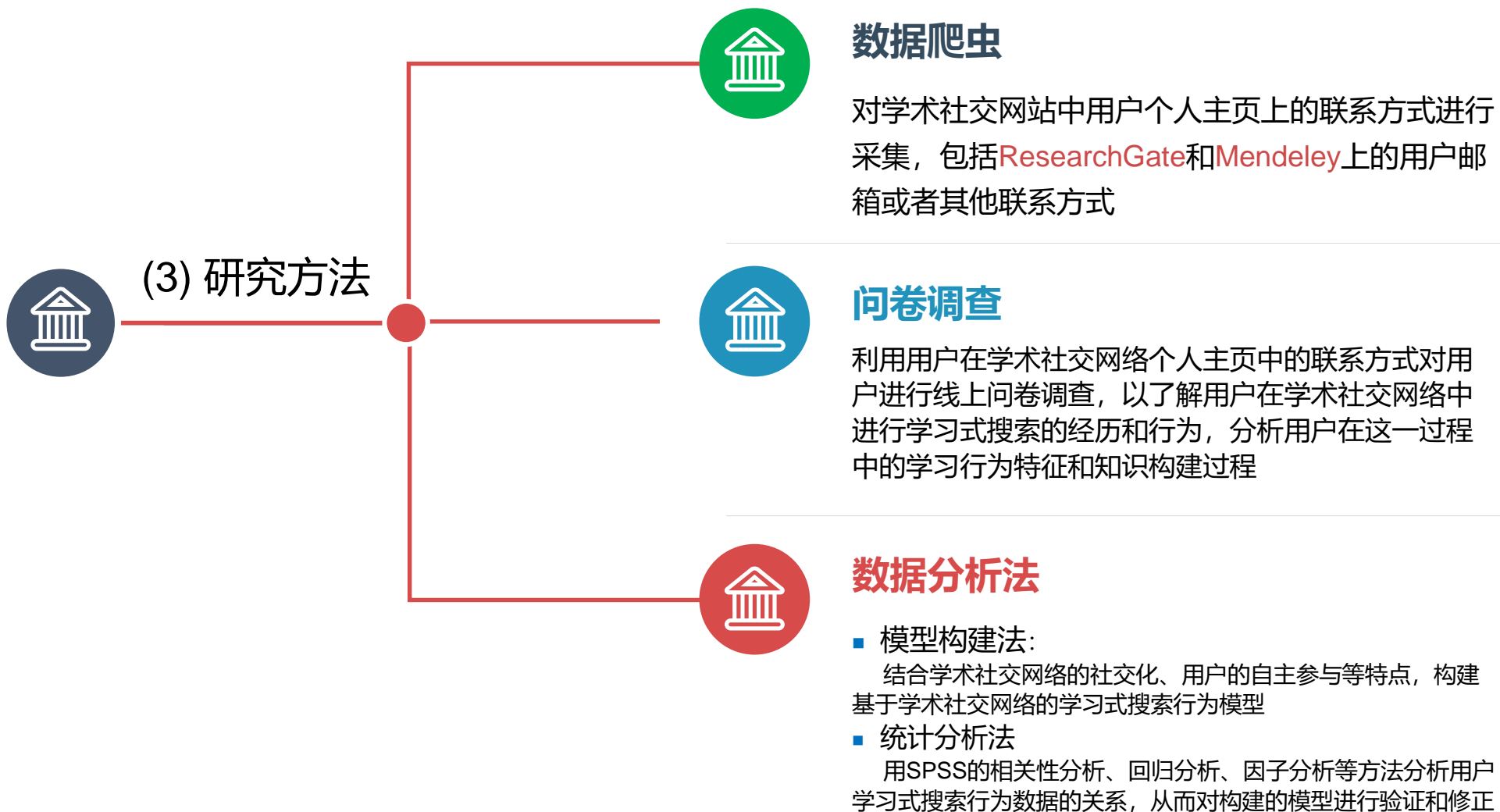
I 学术社交网络学习式搜索行为模型

II 学习式搜索中用户知识结构的变化

- 增强现有的知识 (accretion)
- 调整现有知识 (tuning)
- 重构知识结构 (re-structuring)

III 学习式搜索过程

3.1 学术社交网络中的学习式搜索



3.1 学术社交网络中的学习式搜索

(4) 研究步骤



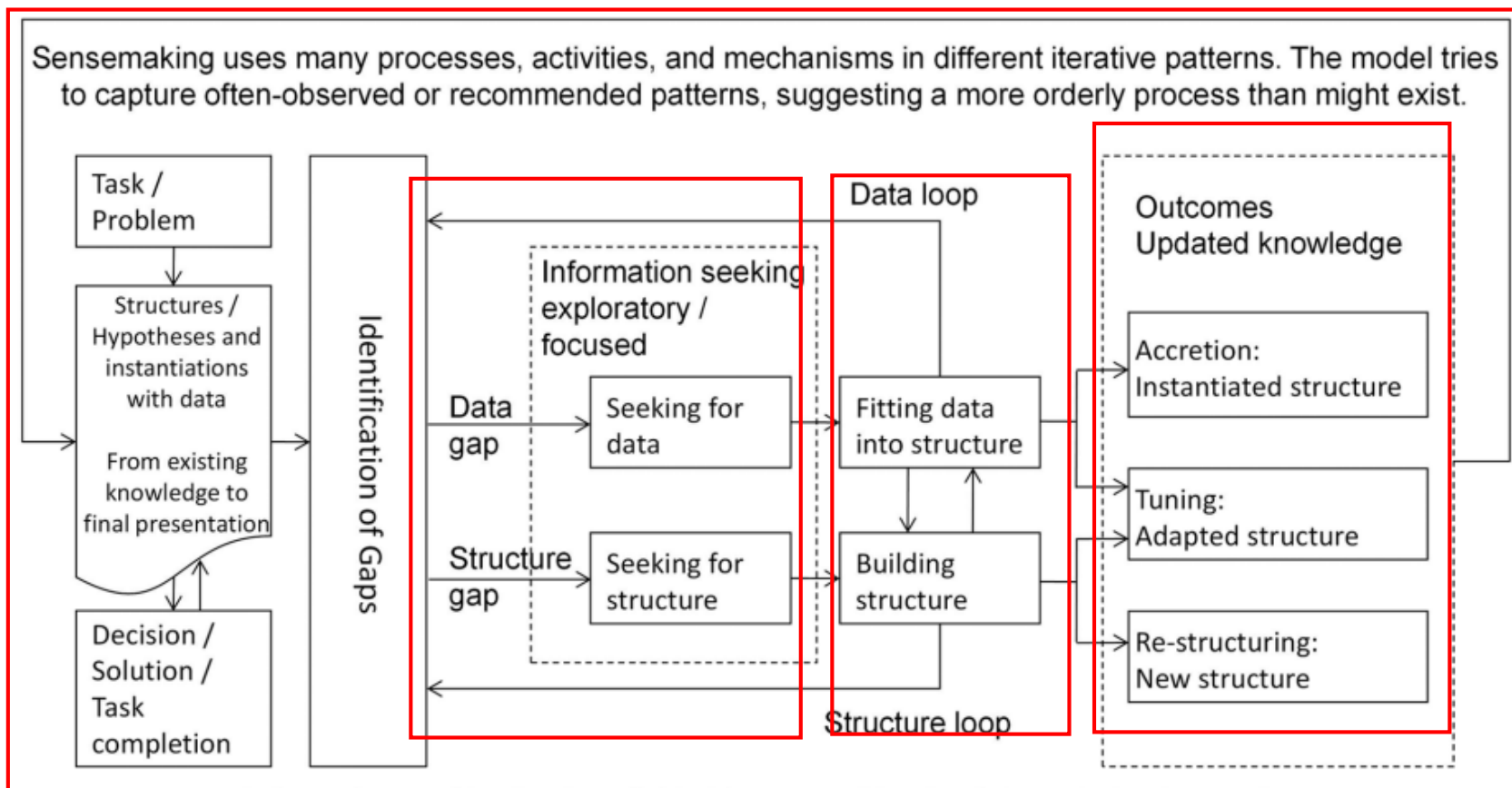
武汉大学

Wuhan University

3.1 学术社交网络中的学习式搜索

用户在搜索过程中的知识变化模型

(Zhang P and Soergel D. Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. Journal of the Association for Information Science and Technology 2014; 65(9): 1733–1756)



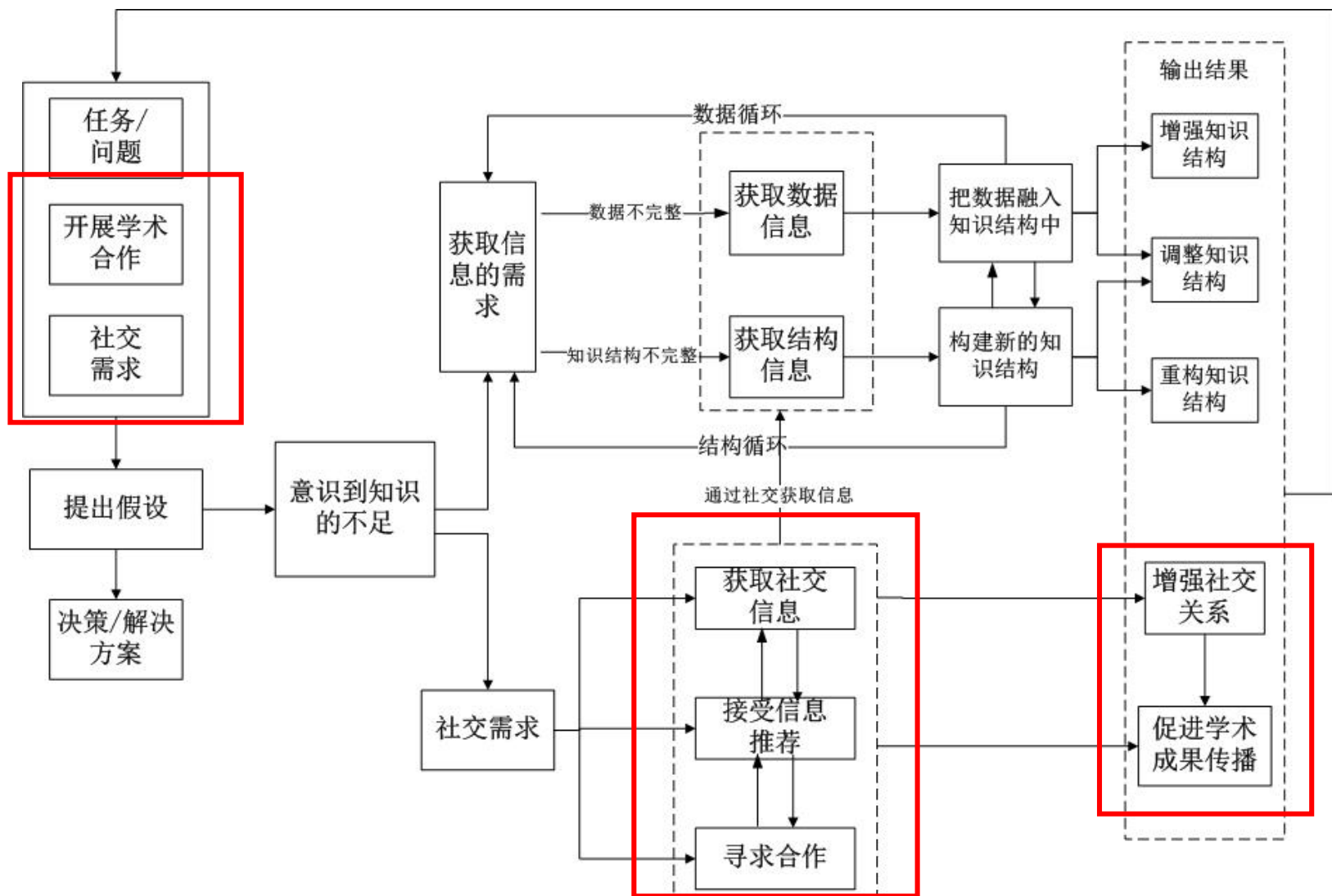
Information seeking is often divided into searching for data and structure and searching for / extracting data and pieces of structure.

(1) 搜索模型

- (2) **意义构建理论** (Sensemaking process model) (1985)
- (3) **学习理论** (Sensemaking Model)
- Kuhlthau's information search process (ISP) model (2004)
- (4) **认知理论** (Theory of Knowledge) (1978)
- Kuhlthau's information search process (ISP) model (2004)
- Kuhlthau's information search process (ISP) model (2004)
- Riegel, Rumelhart & Norman Intelligence Process Model (1981)
- Soergel & Soergel (1984)
- Soergel & Soergel (1984)
- Piore and Gaden National Model (1987)
- Generative Learning Theory of Sensemaking (2006)
- and using (PSU) Model (2013)
- Qu & Furnas Structural Information Seeking Model (2007)

3.1 学术社交网络中的学习式搜索

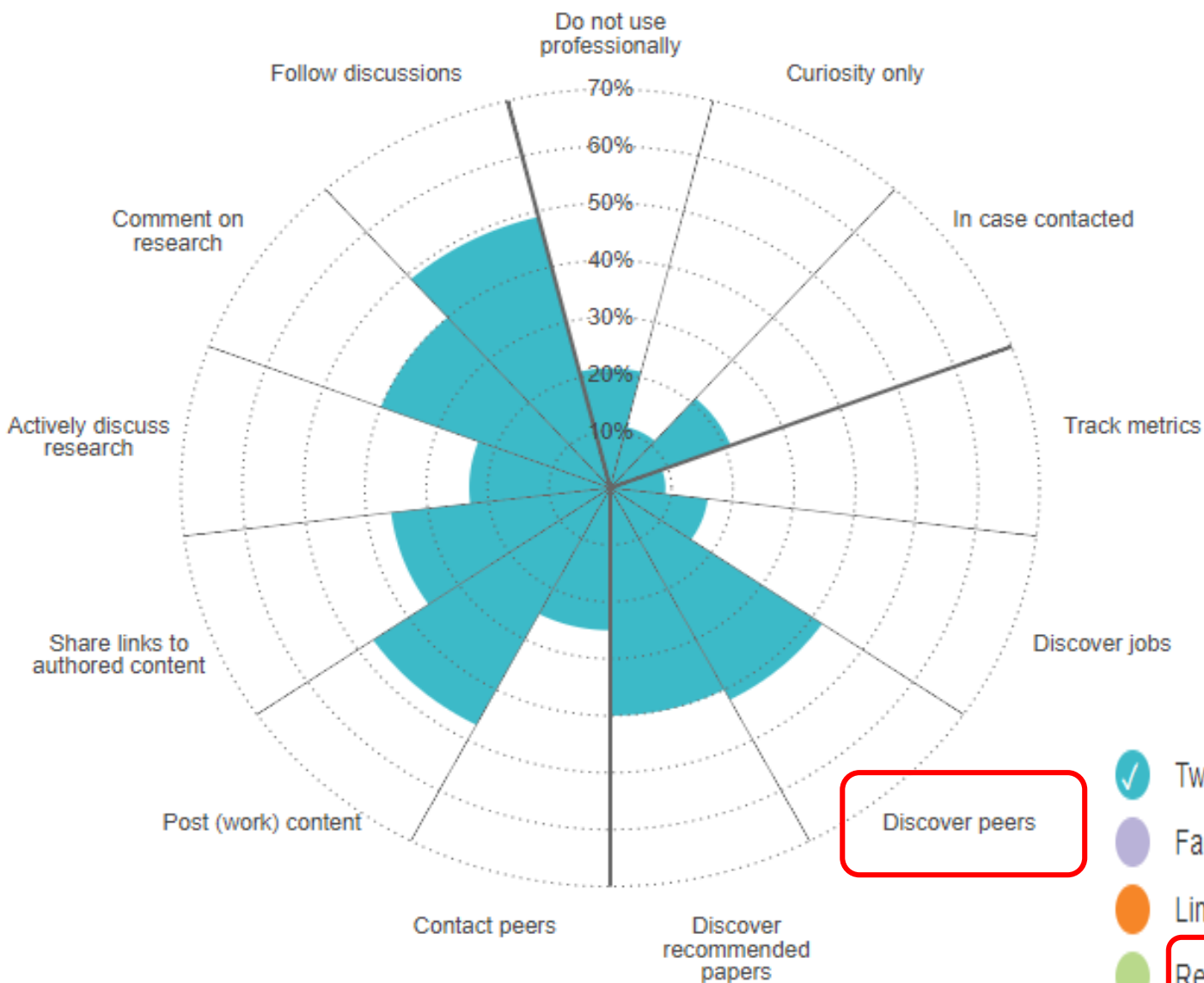
学术社交网络学习式搜索行为模型



参考文献:

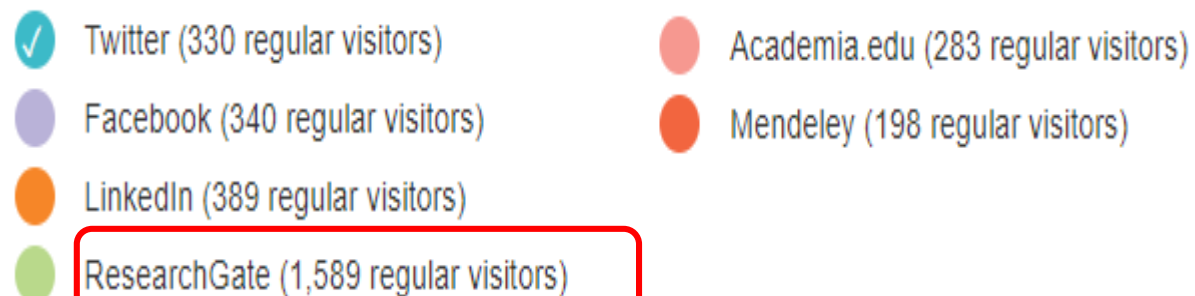
- [1] Corvello V, Genovese A, Verteramo S. Knowledge sharing among users of scientific social networking platforms[C]// Frontiers in Artificial Intelligence & Applications. 2014:369-380.
- [2] 刘晓娟, 余梦霞, 黄勇, 等. 基于 Research Gate 的学术交流行为实证研究--以北京师范大学为例[J]. 情报工程, 2016, 2(3):26-36.
- [3] 周庆山, 杨志维. 学术社交网络用户行为研究进展[J]. 图书情报工作, 2017, 61(16):38-47.

3.2 学术社交网络中的同行发现

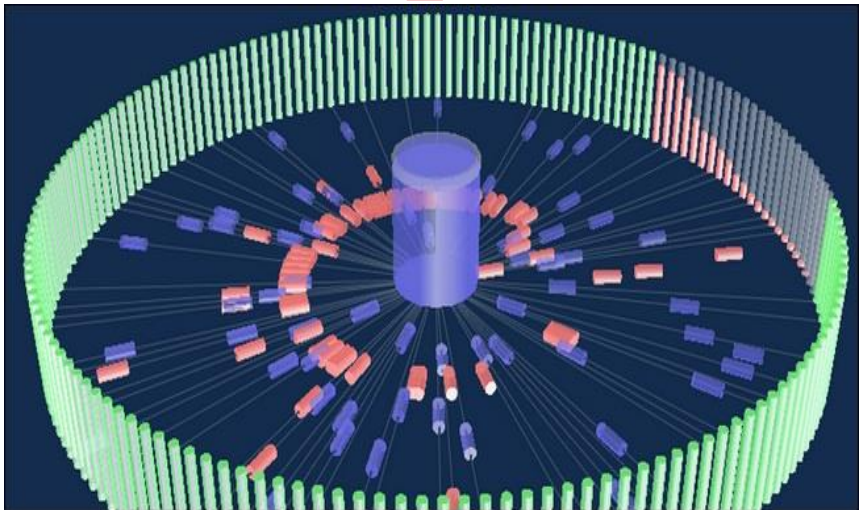
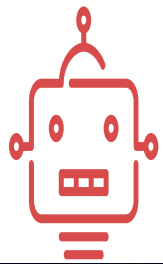


Nature社交网络使用行为调查

- 95个国家
- 3500多分回函
- “发现同行”是第二大动机，仅次于保持联系
- ResearchGate是使用最多的社交媒体



3.2 学术社交网络中的同行发现



(1) 研究问题:

RQ1: 学术社交平台用户通过哪些途径在平台中搜寻感兴趣的学者?

RQ2: 学术社交平台用户在搜寻同行时的路径以及特征如何?

RQ3: 不同路径的搜寻效果是否有差异? 哪些因素影响搜寻效果?

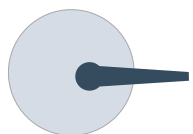


武汉大学

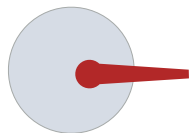
Wuhan University

3.2 学术社交网络中的同行发现

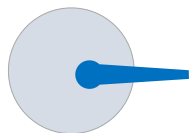
(2) 研究内容



- 分析学术社交网络平台上“同行发现”的行为



- 分析其“同行发现”途径或模式



- “同行发现”效率和学术社交平台的交互设计建议

3.2 学术社交网络中的同行发现

(3) 研究方法



数据收集方法

• 问卷调查

通过站内信的方式向ResearchGate用户发放问卷,问卷内容将包括人口属性、学者搜寻行为频次、常用搜寻途径、搜寻难易度、搜寻效果等

• 用户实验

通过录屏的方式记录实验过程,所要收集的数据类型包括用户人口属性、搜寻途径、用户访问页面、访问路径、搜寻总时长

• 本结构化访谈

旨在了解用户在实验中的情绪变化、实验难易度、搜寻途径选择原因、搜寻效果判定原因等



数据分析方法

• 定量分析

进行用户搜寻点击流分析,分析用户搜寻学者时的路径差异,通过聚类方法总结用户搜寻学者时的不同行为模式,通过相关性分析和回归分析探索影响搜寻效果的因素。

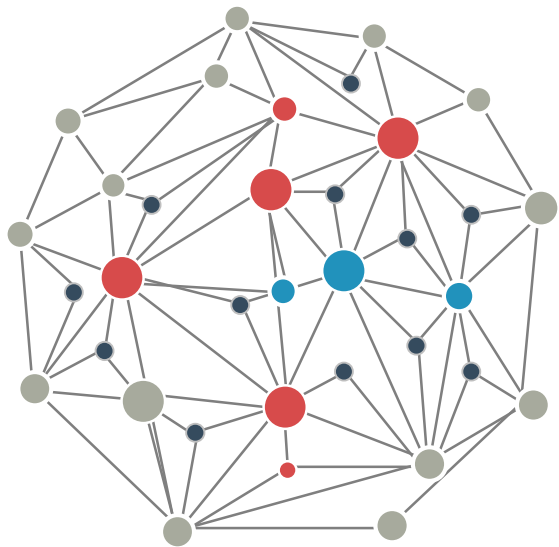
• 定性分析

将访谈录音作为文字,并且访谈内容进行简单编码和分析,了解用户搜寻行为背后的原因以及情感变化



武汉大学

Wuhan University



/04

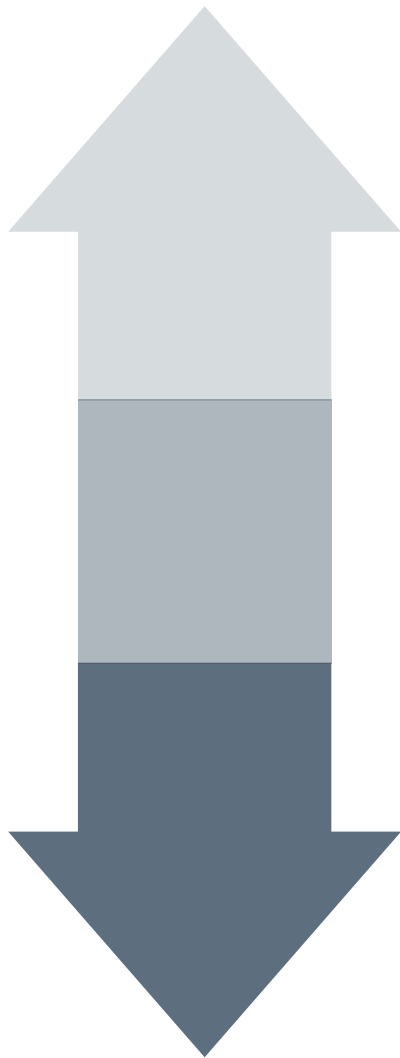
展望



武汉大学

Wuhan University

4.1 学术社交网络研究趋势



- 学术社交网络基础研究（如社交平台使用情况、信息传播特征）
- 基于替代计量学的学术社交网络研究（如论文影响力评价与预测）
- 学术社交网络使用行为研究（如资源推荐、同行发现、学习式搜索）

4.2 不足与建议

- **学术社交网络平台的性能优化问题**

研究多以国外社交平台为主，国内学术社交网络平台在推荐系统等性能方面较弱，因此相关平台的研究较为缺乏，应进一步增加国内学术社交网络平台的功能（如信息检索、信息推荐功能等）

- **学术社交网络特性的理论构建有待加强**

目前研究主要集中于平台使用情况、平台系统优化等基础或者技术层面，缺乏学术社交网络用户行为理论研究（如知识共享行为的影响因素等）

- **学术社交网络研究本土化**

目前研究多以国外学术社交网络平台为主，应加强国内本土学术社交网络在平台采用，使用行为、传播特征等方面的研究

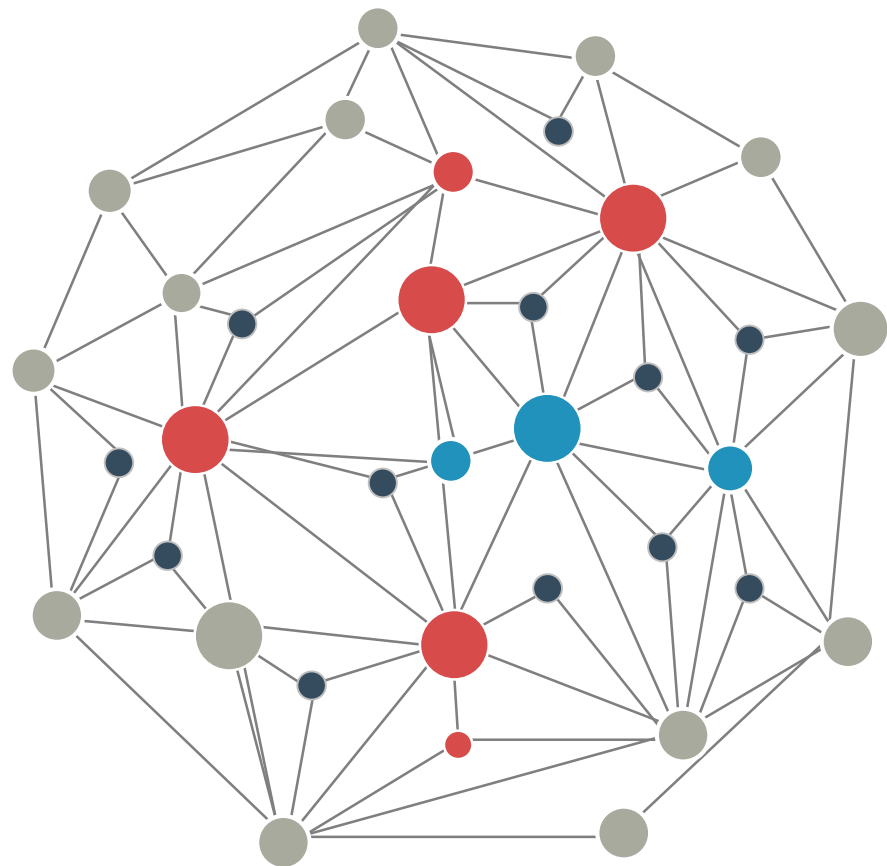


THANKS

欢迎提问与交流分享

吴丹

woodan@whu.edu.cn



武汉大学

Wuhan University