



清华大学

Micro-blogging, Wikipedia, Folksonomy, What's next?

Miao Fan
@CSLT(THU)

24th Sept. 2012



Preface

This is a naïve idea (Not Mature)

Discussion Needed!

Contents

1

Introduction

2

Related Work & Issues

3

New Research Paradigm

4

Research Proposal

5

References

1

Introduction

马少平 THU (设置备注) LV10

聊天 <http://weibo.com/thumsp>

北京

博客: <http://blog.sina.com.cn/thumsp>

已关注 | 取消 发私信 求关注 @他 更多

新浪认证 清华大学计算机系教授 马少平 申请认证

476 5477 10484
关注 粉丝 微博

关注他的人同时关注了

白硕SH 加关注	梁斌penn 加关注	姜太文 加关注
富士通研 加关注	jijames 加关注	独孤跑者 加关注

这些人也关注他 (15)

王海勋...	文曦us...	王斌_1...
布凡THU	夏云庆...	刘洋THU

有点麻烦, 认不出来了怎么办

@姜太文: 为了准备恢复打篮球, 找出了几年前买的隐形开始适应, 虽然有不适应, 但还是帅了不少啊 😎 不过可不要影响待会儿的培训。

10分钟前 来自iPhone客户端 转发(2) | 评论(2)

3分钟前 来自新浪微博 | 举报 转发(1) | 收藏 | 评论

What is Micro-blogging?

- Like twitter, which combines SMS, Social Network & Broadcast medium
- 140 characters (Chinese or English) – **Short text**
- Record Daily life, Opinions, Retweets etc.
- Follow or be followed

1

Introduction

WIKIPEDIA



English
The Free Encyclopedia
4 000 000+ articles

日本語
フリー百科事典
815 000+ 記事

Español
La enciclopedia libre
900 000+ artículos

Deutsch
Die freie Enzyklopädie
1 430 000+ Artikel

Русский
Свободная энциклопедия
875 000+ статей

Français
L'encyclopédie libre
1 270 000+ articles

Italiano
L'enciclopedia libera
939 000+ voci

Polski
Wolna encyklopedia
908 000+ haseł

中文
自由的百科全书
500 000+ 条目

Search · Suchen · Rechercher · Zoeken · Ricerca · Szukaj · Buscar ·
 ПОИСК · 検索 · Busca · Sök · 搜索 · Tìm kiếm · ПОШУК · Cerca · Søk ·
 Haku · Hledání · Keresés · 찾기 · Cari · Ara · جستجو · Căutare · بحث ·
 Hľadaf · Søg · Serču · Претрага · Paieška · Poišči · Cari · חיפוש ·
 Търсене · Izdeu · Bilatu · Suk · Bilnga · Traži · खोज

Q 中文

What is Wikipedia?

- Online encyclopedia
- Freely create and revise articles
- A precious knowledge base
- Multilingual
- Category-concept organized

豆瓣读书

书名、作者、ISBN

搜索

分类浏览 阅读

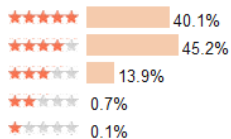
简爱



原名: Jane Eyre
 作者: [英] 夏洛蒂·勃朗特
 出版社: 世界图书出版公司
 出版年: 2003-11
 页数: 436
 定价: 18.00元
 装帧: 平装
 丛书: 名著典藏
 ISBN: 9787506261579

★★★★★ 8.4

(74885人评价)



想读 在读 读过 评价: ☆☆☆☆☆

写笔记 写书评 加入购书单 加入豆列

推荐

内容简介 ·····

《简爱》是英国女作家夏洛蒂·勃朗特的代表作品。女主人公简爱是一个追追求平等与自主的知识女性形象，小说以其感人的对于一位“灰姑娘式”人物奋斗史的刻划取胜，《简爱》也是女性文学的代表作品。

豆瓣成员常用的标签(共2013个) ·····

简爱(10450) 外国文学(8835) 经典(4923) 小说(4676) 爱情(4088) 名著(4065) 英国(3199) 外国名著(2388)

丛书信息

名著典藏 (共78册), 这套丛书还有《培根散文集》,《德伯家的苔丝》,《消失的地平线》,《傲慢与偏见》,《富兰克林自传》等。

What is Folksonomy?

- Freely annotate tags
- Flat way to re-organize resources
- Reflect personal interest to some extent



Introduction

What do they (Micro-blogging, Wikipedia, Folksonomy) have in common?

- Users contribute content (UCC) in various forms.
- Web 2.0 era emerged

But now, there still exists gaps between every two of them.

Micro-blogging:

- Personal interests mining, opinion mining (sentiment), personalized search etc.

Wikipedia:

- Semantic relatedness analysis

Folksonomy:

- Tag recommendation, personalized recommendation.



The question is : Why they are isolated?

	Micro-blog	Wikipedia	Folksonomy
What are its advantages?	Rapid Information Dissemination	World knowledge (concepts, categories, fine edited articles)	Semantic Tags, resources re-organized form
What are its disadvantages? (Short comings)	Few semantics, few world knowledge	How to use to knowledge?!	Few user contribution

How about in this way?

好! // @林鸿飞: 每年CCIR, 安排一个徒步走的Session, 边走边议。 // @马少平 THU: 很赞啊, 有想参加的冲动。林院长, 把这个发展成一个IR徒步走吧? 我肯定带队参加。

@林鸿飞- V: 今天是我们实验室传统项目-海边徒步走, 已经进行了七年。前几年的路线都是沿着大连最美丽的观景线-滨海路, 今年在资深驴友@老赛2050的策划和带领下, 尝试开辟新的路线, 希望带来不一样的感受。据报有阵雨, 不过现在看天气很适宜户外活动。我们在路上...

今天08:50 来自新浪微博 转发(8) | 评论(11)

今天11:04 来自新浪微博 | 举报 转发 | 收藏 | 评论



维基百科
自由的百科全书

首页
分类索引
特色内容
新闻动态
最近更改
随机条目

帮助
帮助
维基社群
方针与指引
互助客栈
询问处
字词转换
IRC即时聊天
联系我们
关于维基百科

条目 讨论 大陆简体 ▾

信息检索

维基百科，自由的百科全书
(重定向自**信息检索**)

信息检索（**英语：**Information Retrieval）是指搜索信息的信息需求提供查询的方法以及查询的过程。

然而这里会出现一个使人混乱的情况，就是数据检索、文件检索。信息检索像是刚开始存在的、跨学科的，其创建于**计算机科学**。供检索**书目**、**期刊**和其他的**文件**，信息检索系统通常与对象和储于数据库中，但是会在系统中使用**文件代理者**来代替文件本。

信息检索的类型

按照检索手段，分为：传统信息检索（手工检索）和现代信息；按照检索内容，分为：书目检索、数据检索、事实检索、全文。

信息检索的主要技术指标

他的标签(5)

搜索引擎 信息检索

智能信息处理 音乐

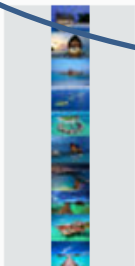
爬山散步

3

New Research Paradigm

力求在博士期间有机会去一次，此生无憾矣

@旅途絮语：马尔代夫剩余寿命紧紧只有12年了，美国发布最新探测报告称马尔代夫将在20年内被淹没，原因正是气候变暖，这天堂般的旅游胜地你会错过吗？
(转)



9月6日09:00 来自精彩微客

转发(25257) | 评论(3224)


标签 马尔代夫 天堂

9月6日16:53 来自新浪微博

删除 | 置顶 | 转发 | 收藏 | 评论

WIKIPEDIA

<p>English <i>The Free Encyclopedia</i> 4 000 000+ articles</p>	<p>日本語 フリー百科事典 815 000+ 記事</p>
<p>Español <i>La enciclopedia libre</i> 900 000+ artículos</p>	<p>Deutsch <i>Die freie Enzyklopädie</i> 1 430 000+ Artikel</p>
<p>Русский <i>Свободная энциклопедия</i> 875 000+ статей</p>	<p>Français <i>L'encyclopédie libre</i> 1 270 000+ articles</p>
<p>Italiano <i>L'enciclopedia libers</i> 939 000+ voci</p>	<p>Polski <i>Wolna encyklopedia</i> 908 000+ hasel</p>
<p>Português <i>A enciclopédia livre</i> 742 000+ artigos</p>	<p>中文 自由的百科全书 500 000+ 条目</p>



Search · Suchen · Rechercher · Zoeken · Ricerca · Szukaj · Buscar ·
 ПОИСК · 搜索 · Busca · Sök · 搜索 · Tìm kiếm · ПОШУК · Cerca · Søk ·
 Haku · Hledání · Keresés · 契ル · Cari · Ara · جستجو · Căutare · بحث ·
 H'adaf' · Søg · Serču · Претрага · Paieška · Poišči · Cari · חיפוש ·
 Търсене · Izdeu · Bilatu · Suk · Bilnga · Traži · ৳

What if? By exploiting Wikipedia knowledge to micro-blog (short text) mining, such as tag-form interests annotation?

Therefore, Micro-blogging & Wikipedia & Folksonomy Mutually reinforce each other.

- By Leveraging the adequate semantic resources and world knowledge in Wiki.
- To enhance the short text mining on Micro-blogging platform.
- Then automatically generate the personalized tags.

- As a matter of the result, re-organize the resources on micro-blogging platform.

Micro-blog mining by leveraging Wikipedia

App form:

- Personalized Tag recommendation on Micro-blogging platform
- Tag-based micro-text management
- Personalized resource recommendation (Ad. Or multi-media)

- [1] H. Kwak, C. Lee, H. Park and S. Moon. What is twitter, a social network or a new media? In WWW'2010, pages 591-600.
- [2] Z. Liu, X. Chen and M. Sun. Mining the interests of Chinese microbloggers via keyword extraction. *Front. Comput. Sci.*, 2012, 6(1): 76-87.
- [3] W. Wu, B. Zhang and M. Ostendorf. Automatic generation of personalized annotation tags for twitter users. In NAACL'10, pages 689-692.
- [4] E.Meiji, W. Weerkamp, M. de Rijike. Adding semantics to microblog posts. In Proceedings of the fifth ACM international conference on Web Search and data mining, pages 563-572, ACM, 2012.
- [5] X. Li, L. Guo, Y. Zhao. Tag-based social interest discovery. In Proceedings of the 17th international conference on World Wide Web, pages 675-684. ACM, 2008.



Q & A



Thank You!