

第3节. 搜索引擎性能评价

艾清遥

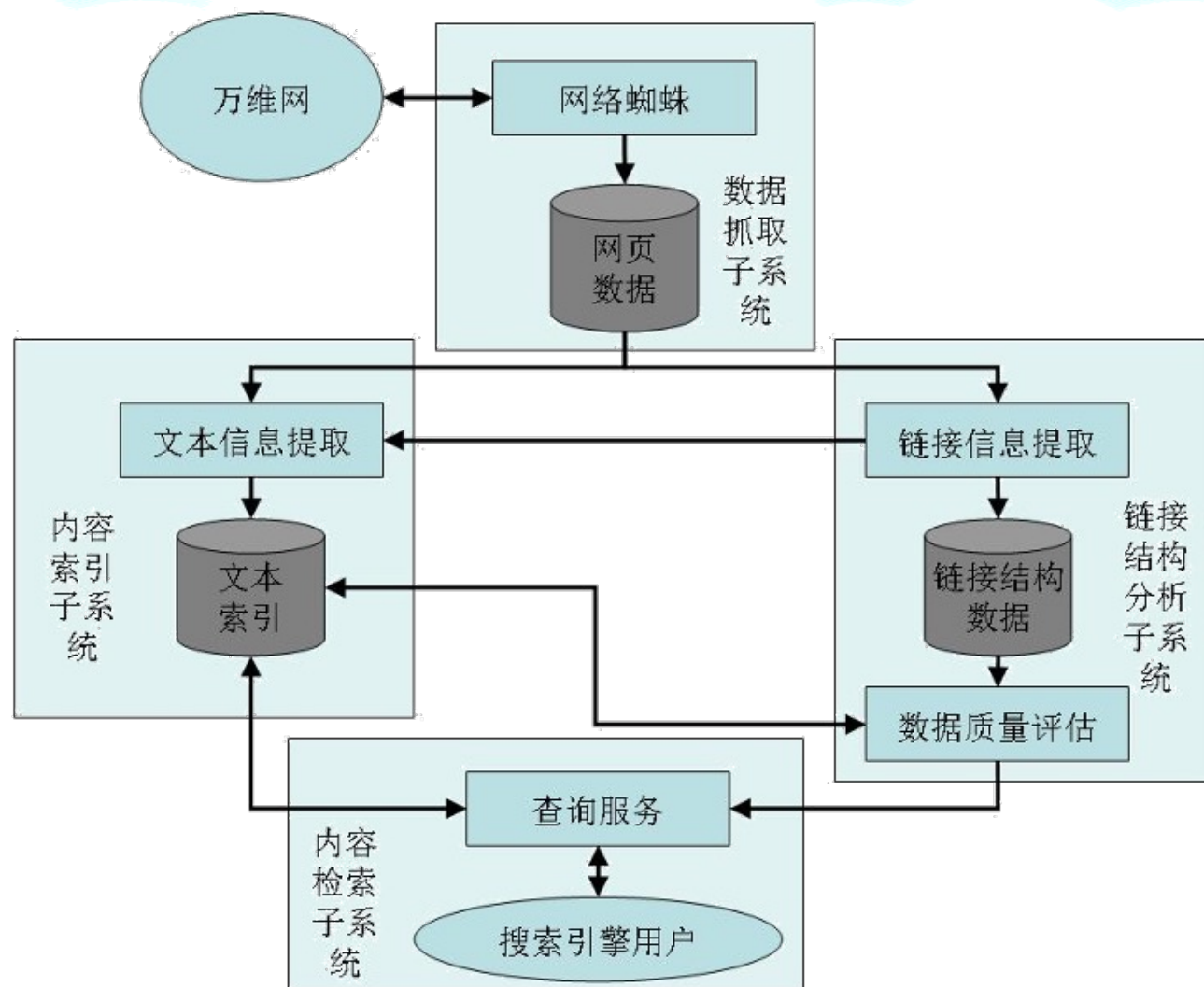
清华大学计算机系

清华大学互联网司法研究院

2023年3月7日



上节内容回顾: 搜索引擎体系结构



引言. 搜索引擎性能评价的目的

谷歌网站搜索速度提升 雅虎末路

雅虎搜索引擎重大升级 自称比Google功能更全

张忆芬评价：百度谷歌都不如雅虎搜索？

雅虎称其数据库超谷歌为全球最大

谷歌每天处理2万TB数据 与雅虎微软相比优势明显

张朝阳：搜狗3.0已超百度、雅虎

【来源：中国证券网.上海证券报】

□本报记者 张韬

昨天记者从搜狐获悉,搜狗3.0版本将于2007年1月1日正式上线。搜狐公司董事局主席兼首席执行官张朝阳对搜狗充满信心,甚至放言:“3.0是换代产品,已经在技术上超过了百度和雅虎!”

2007年正式上线的搜狗3.0采用新的技术架构——自主研发的服务器集群并行抓取技术,中文网站收录量从2006年8月的50亿猛增到100亿,已覆盖中文网页数据量的50%以上,每天的更新速度达到5亿网页。搜狗3.0在搜索结果的排名上采用搜狗网页评级体系,不仅考察网

页之间的链接关系,同时考察了链接质量、链接之间的相关性等特性,网页评级越高,该网页在搜

站的承诺提供了强有力的支持



引言. 搜索引擎性能评价的目的

- 对搜索引擎用户而言：
 - 挑选最有利于获取信息的渠道
 - 对搜索引擎广告商而言：
 - 挑选最有效的广告投放手段
 - 对搜索技术研究人员而言：
 - 评价在信息检索系统的研发中一直处于核心的地位，以致于算法与其效果评价方式是合二为一的
- (Saracevic, Salton Award, SIGIR 1997)

<https://sigir.org/awards/gerard-salton-awards/>



引言. 搜索引擎性能评价的目的

- 对搜索技术研究人员而言:
- 从较差查询样例中学习

内外合作伙伴, 建立长期战略合作关系, 实现优势互补, ...
www.chinaunicom.com.cn 2012-03-22 - [即刻快照](#)

[中国联通网上营业厅](#) www.10010.com [iphone 4s,小米手机,3g...](#)
中国联合通信有限公司
www.10010.com 2012-04-01 - [即刻快照](#)

[中国联通-600050-上交所-东方财富网](#)

最新价格: **4.27** -0.01 (-0.23%) 2012-04-19 12:12



4.30
4.30
4.29
4.29
4.28
4.27
4.27
4.26
4.26

9:30 10:30 11:30/13:00 14:00 15:00

10000
7500
5000
2500

0.53% 今开: 4.30
0.40% 昨收: 4.28
0.27% 最高: 4.30
0.13% 最低: 4.27
0.00% 市盈率: 64.11
0.13% 成交量: 254341手
0.27% 上证指数
0.40% 2377.57(-0.14%)
0.53% 深证成指
10005.05(-0.64%)

[内乡风景园林果苗基地](#)
内乡风景园林果苗基地
www.ylgsm.com 2012-03-09 - [即刻快照](#)

内乡风景园林果苗基地
Nei Xiang Landscape Nursery Stock Base

网站首页 基地简介 新闻中心 产品展示 栽培技术 招商合作 在线留言 联系我们

咨询电话: 15036283268

本苗圃基地所供大、小苗木, 花草规格齐全, 欢迎各界人士。
欢迎实地考察订货。
宗旨: 顾客至上, 诚信为本, 我基地有的给您最好。
没有的帮您想办法, 让质量、价格、速度成为我们永久的保证。

产品分类 **公司动态**

园林
大叶女贞、梅花、杜鹃、香樟、李树、
百日红(紫薇)、法桐、木槿、国槐、
黄桷、广玉兰、皂角树、黄连树

果树
桃树、核桃、南丰蜜桔、杏树、
梨树、葡萄、猕猴桃

内乡风景园林果苗基地坐落在内乡县, 风光秀丽的内乡县, 这里交通便利, 地理位置优越, 物产丰富, 人杰地灵, 是著名的苹果、梨、桃产区。种植果树的历史悠久, 深州蜜桃以其色泽艳丽, 香甜可口, 品质一流, 驰名中外, 是旅游、观光的好地方。公司主要繁育各种桃树、苹果、梨、桃、梨、杏、李子、山楂等。[详细]

产品展示

桃树苗 石楠苗 南丰蜜桔树苗 核桃苗

地址: 河南省内乡县内乡镇
联系人: 唐中义
电话: 15036283268
15139082368

本节概要

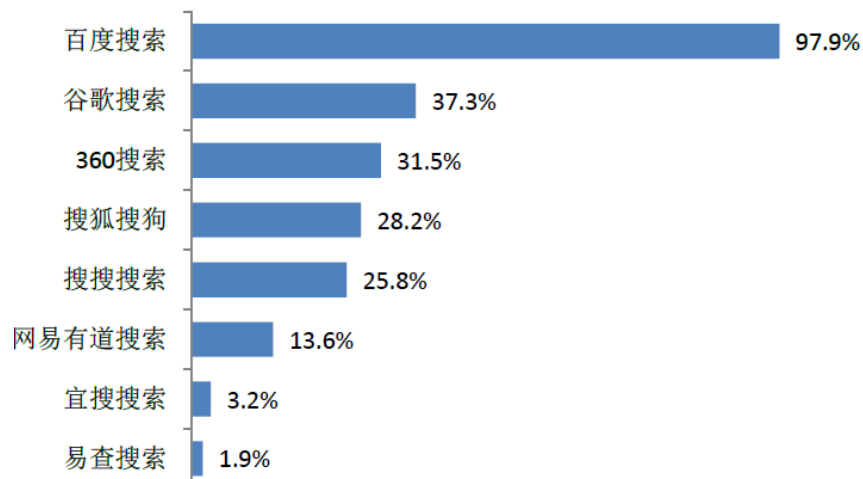
- 搜索引擎性能评价流程
- 用于性能评价的语料库采集
- 用于性能评价的查询样例集合构建
- 用于性能评价的结果相关性标注
- 搜索性能评价指标设计



1. 搜索引擎性能评价流程

• 1.1 搜索引擎性能评价的关注对象

• 作为网络服务供应商的属性



Core Search Entity	Explicit Core Search Share (%)		
	Nov-13	Dec-13	Point Change
Total Explicit Core Search	100.0%	100.0%	N/A
Google Sites	66.7%	67.3%	0.6
Microsoft Sites	18.1%	18.2%	0.1
Yahoo Sites	11.2%	10.8%	-0.4
Ask Network	2.6%	2.5%	-0.1
AOL, Inc.	1.4%	1.3%	-0.1

• 作为网络信息检索工具的属性

- 传统的信息检索系统评价方式如何应用于搜索引擎？如何充分考虑网络数据环境与用户群体的影响？

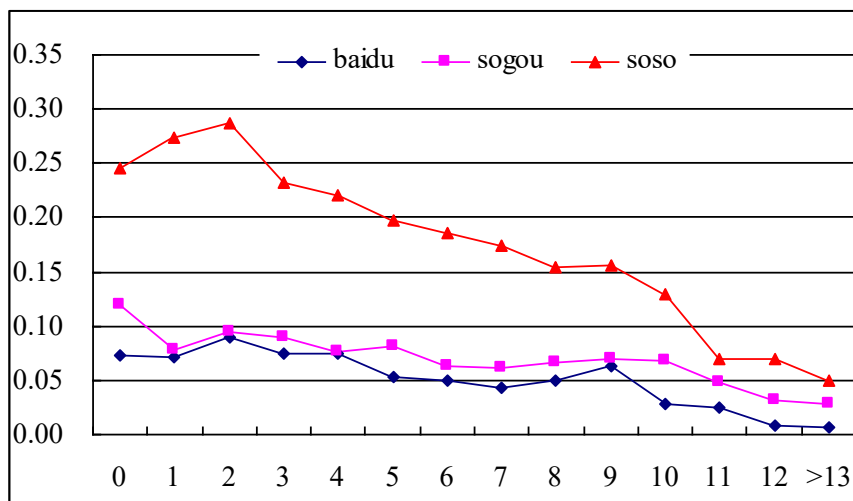


1. 搜索引擎性能评价流程

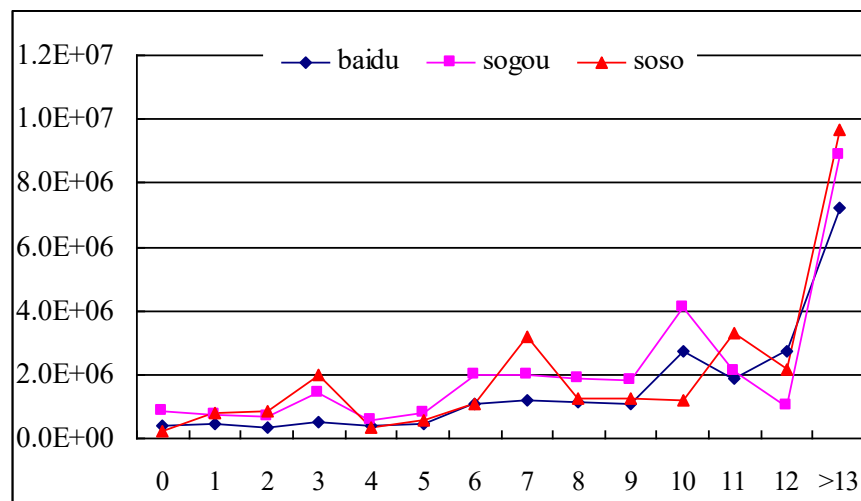
• 1.1 搜索引擎性能评价的关注对象

• 搜索引擎系统运行效率 (Efficiency)

- 用户需求是否得到了很快的响应?
- 为满足用户需求耗费了多大规模的硬件资源?



反馈时间

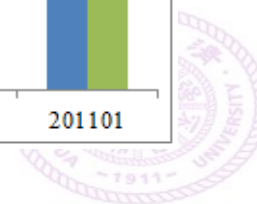
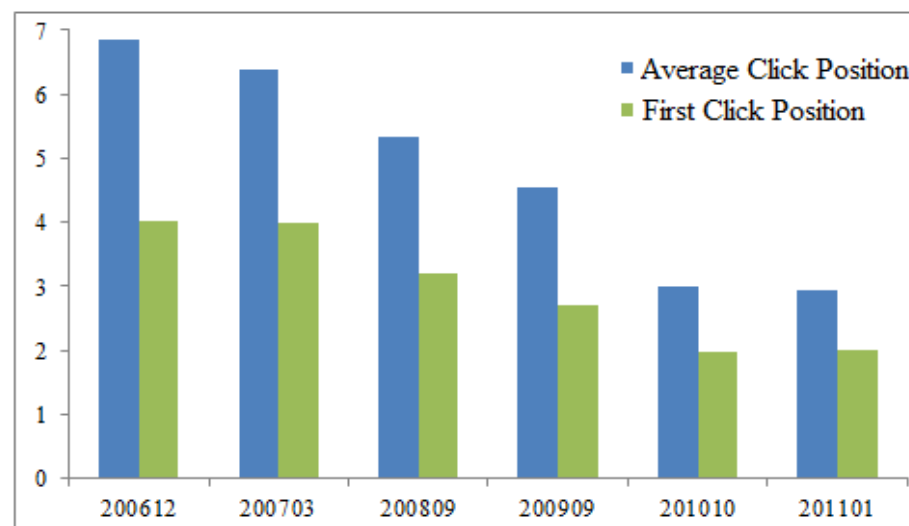
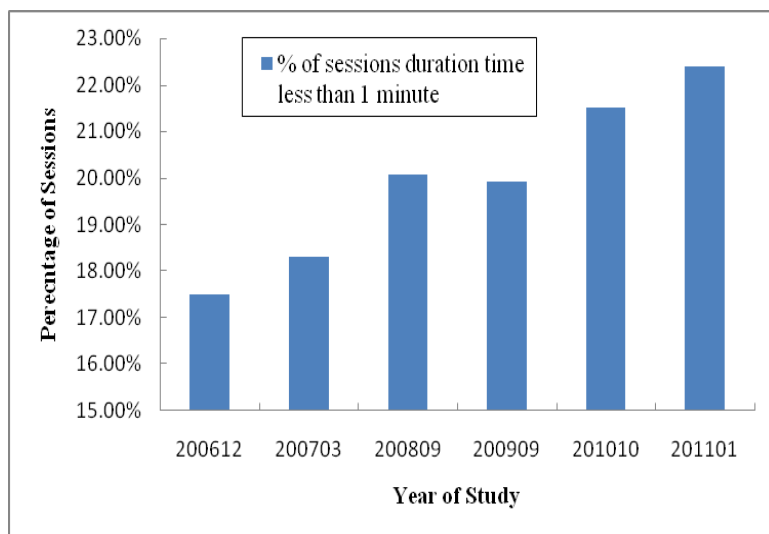


结果条目数



1. 搜索引擎性能评价流程

- 1.1 搜索引擎性能评价的关注对象
 - 搜索引擎系统运行效果 (Effectiveness)
 - 是否满足了用户的信息需求?
 - 用户获取信息所耗费的时间成本如何?



1. 搜索引擎性能评价流程

• 1.2 搜索引擎性能评价的Cranfield体系

- 性能评价的“黑箱”方式

- 给定标准输入情况下，系统输出与标准输出之间的差异

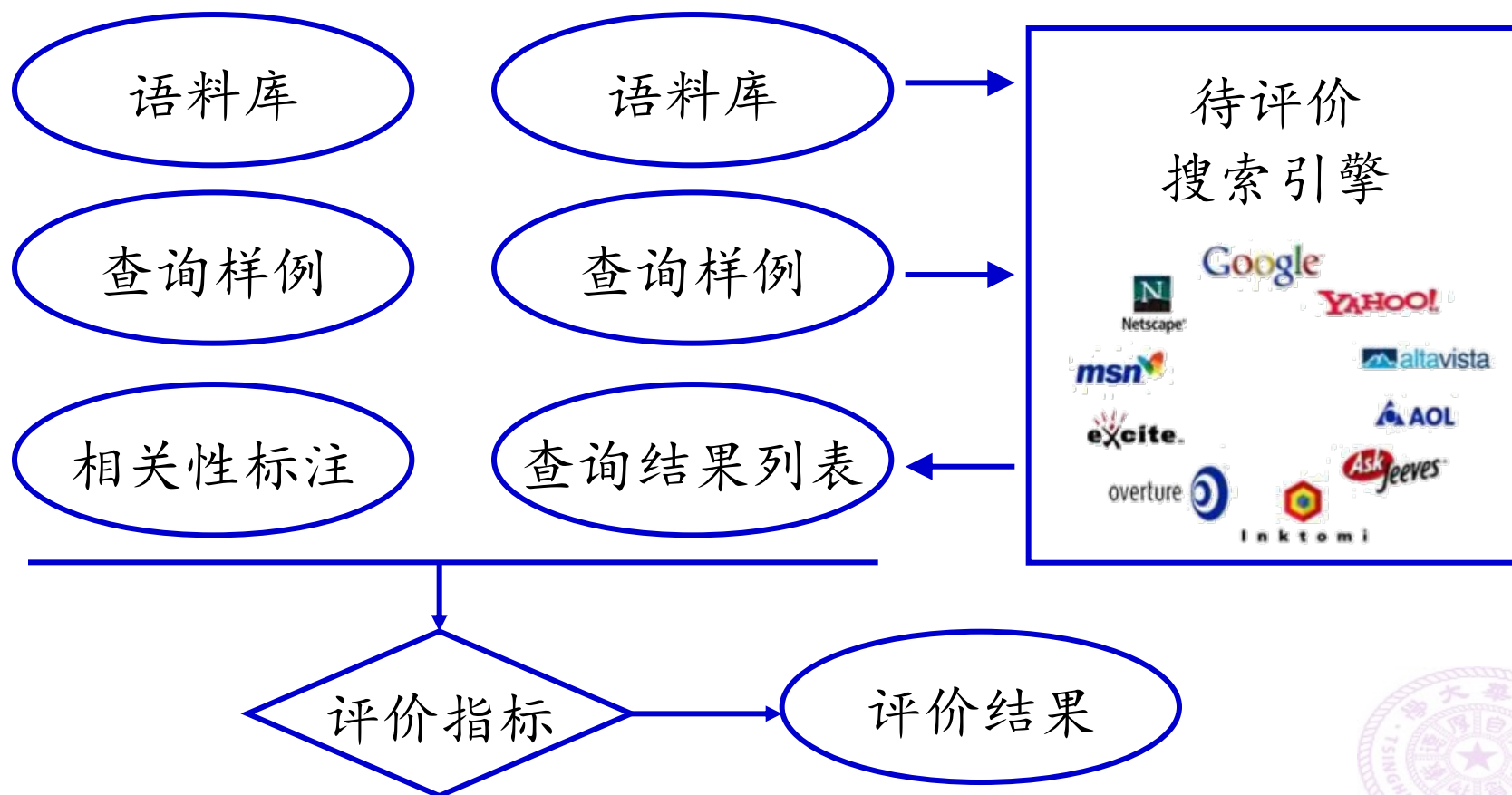
- Cranfield评价体系

- Cleverdon等人于1960年前后在英国Cranfield大学提出



1. 搜索引擎性能评价流程

• 1.2 搜索引擎性能评价的Cranfield体系



1. 搜索引擎性能评价流程

• 1.2 搜索引擎性能评价的Cranfield体系

• Cranfield评价体系的实践应用

- 文本信息检索会议 (Text REtrieval Conference, TREC)
- 1992年开始, 由NIST和DARPA共同主办; UMass, UIUC, CMU, IBM, MSRA/C, ... ; THU, PKU, NUS, TOKYO, ...
- 国立信息技术研究所信息检索评测 (NII Testbeds and Community for Information access Research)
- 1997年开始, 由日本NII主办; 强调跨语言评测任务。
- 其他专题评测: CLEF, FIRE, SEWM, ...



1. 搜索引擎性能评价流程

• 1.2 搜索引擎性能评价的Cranfield体系

- Cranfield体系的优势：

- 复用性：一次标注，多次使用

- 如何用Cranfield评价体系进行检索效果评价

- 如何采集评价语料

- 如何构建查询样例集

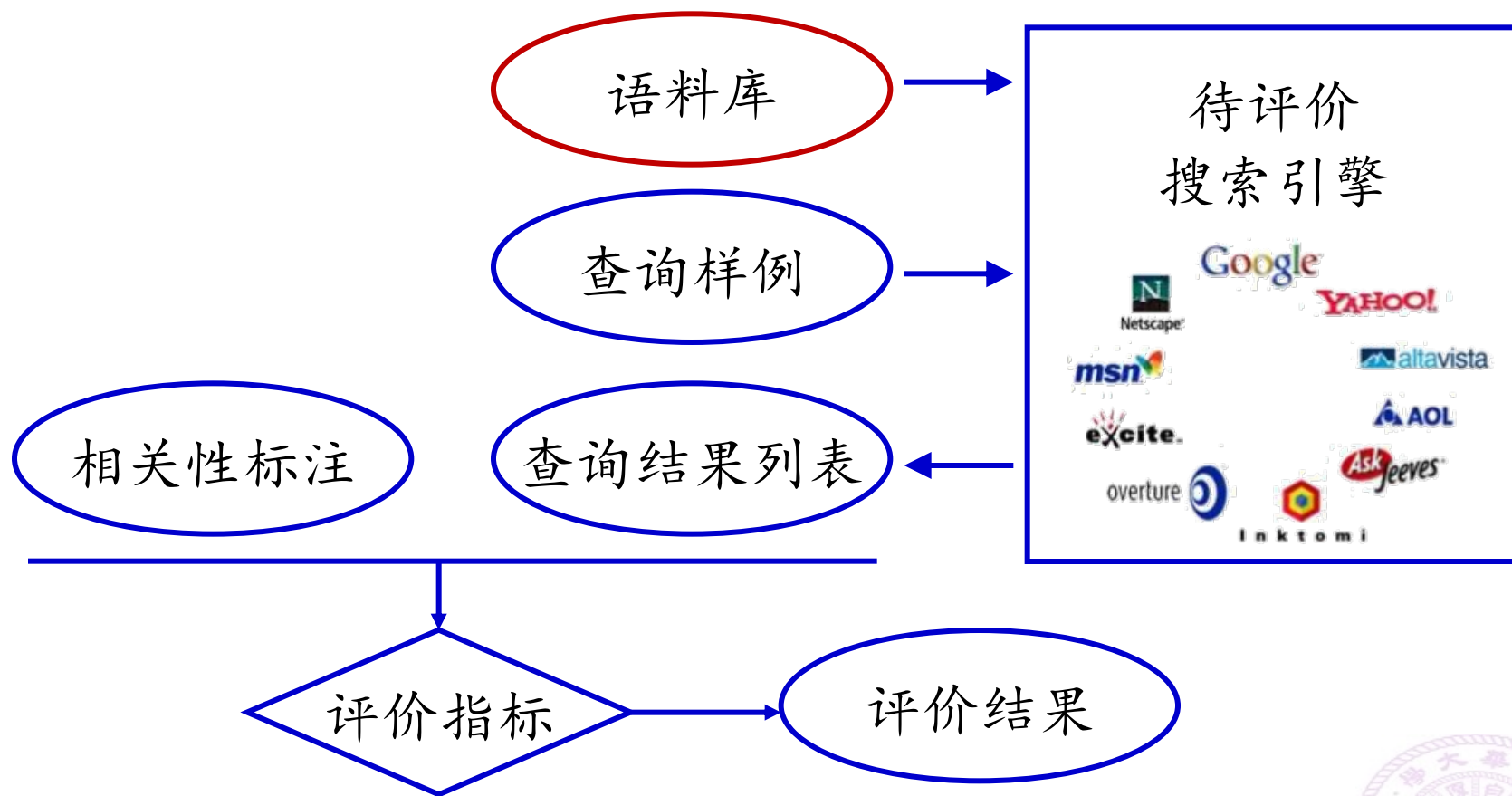
- 如何进行相关性标注

- 如何设计评价指标

面临哪些技术问题？
施行的方法是什么？

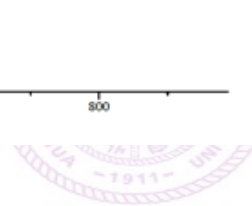
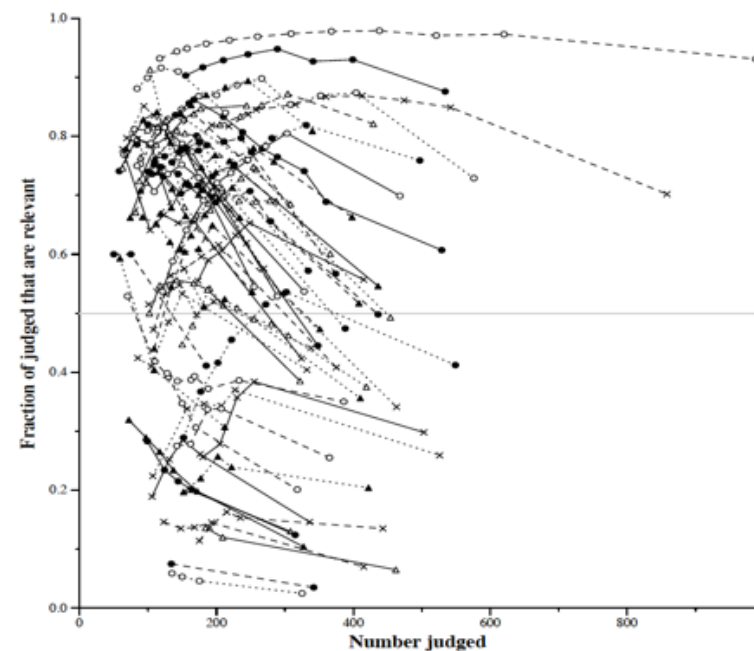


2. 语料库采集



2. 语料库采集

- 对于信息检索系统
 - 提供固定的语料库集合
 - 集合规模适当: VLC2, DL'21
 - 数据质量可靠: ClueWeb
- 对于商业搜索引擎
 - 不提供固定的语料库集合
 - 评价数据抓取子系统性能
 - 索引量战争, 暗网抓取, ...



2. 语料库采集

• 案例：360与百度关于数据抓取的官司

```
User-agent: Baiduspider
Disallow: /w?

User-agent: Googlebot
Disallow: /update
Disallow: /history
Disallow: /usercard
Disallow: /usercenter

User-agent: MSNBot
Allow: /

User-agent: Baiduspider-image
Disallow: /w?

User-agent: YoudaoBot
Allow: /

User-agent: Sogou web spider
Disallow: /update
Disallow: /history
Disallow: /usercard
Disallow: /usercenter

User-agent: Sogou inst spider
Disallow: /update
Disallow: /history
Disallow: /usercard
Disallow: /usercenter
```

<http://baike.baidu.com/robots.txt>

Baidu 百度 新闻 网页 贴吧 知道 音乐 图片 视频 地图 文库 更多»

为什么我的小米3充电充的那么慢?

百度一下

[为什么我的小米3充电充的那么慢? 以前不会这样的 昨天..._百度知道](#)

2个回答 - 提问时间: 2014年03月04日

最佳答案: 需要多久能充满? 放电使用正常么? 换个充电器试试, 如果不是充电器问题就可能是电池或者电源管理电路有问题

zhidao.baidu.com/link?url=CY5FQRepBuHySwm... 2014-03-04

小米3为什么充电很慢	1个回答	2013-10-24
用充电宝充小米3为什么那么慢	3个回答	2014-01-06
为什么小米3用送的充电器充电很慢呢?都充了...	3个回答	2014-01-05

[更多知道相关问题>>](#)

360搜索+ 新闻 网页 问答 视频 图片 音乐 地图 雷电 更多+

为什么我的小米3充电充的那么慢?

搜索一下

[小米3充电非常慢电量剩27%的时候我开始充电用的原装的充电...](#)

小米3充电非常慢电量剩27%的时候我开始充电用的原装的充电器然后就打开WIFI看看... 不知道为什么我也一样,我是双十一才买的米3用的很快,但是冲进来非常慢,但是我每次都...

wenwen.soso.com/z/q501561205.h... 2013-12-13 - 快照 - 59%好评

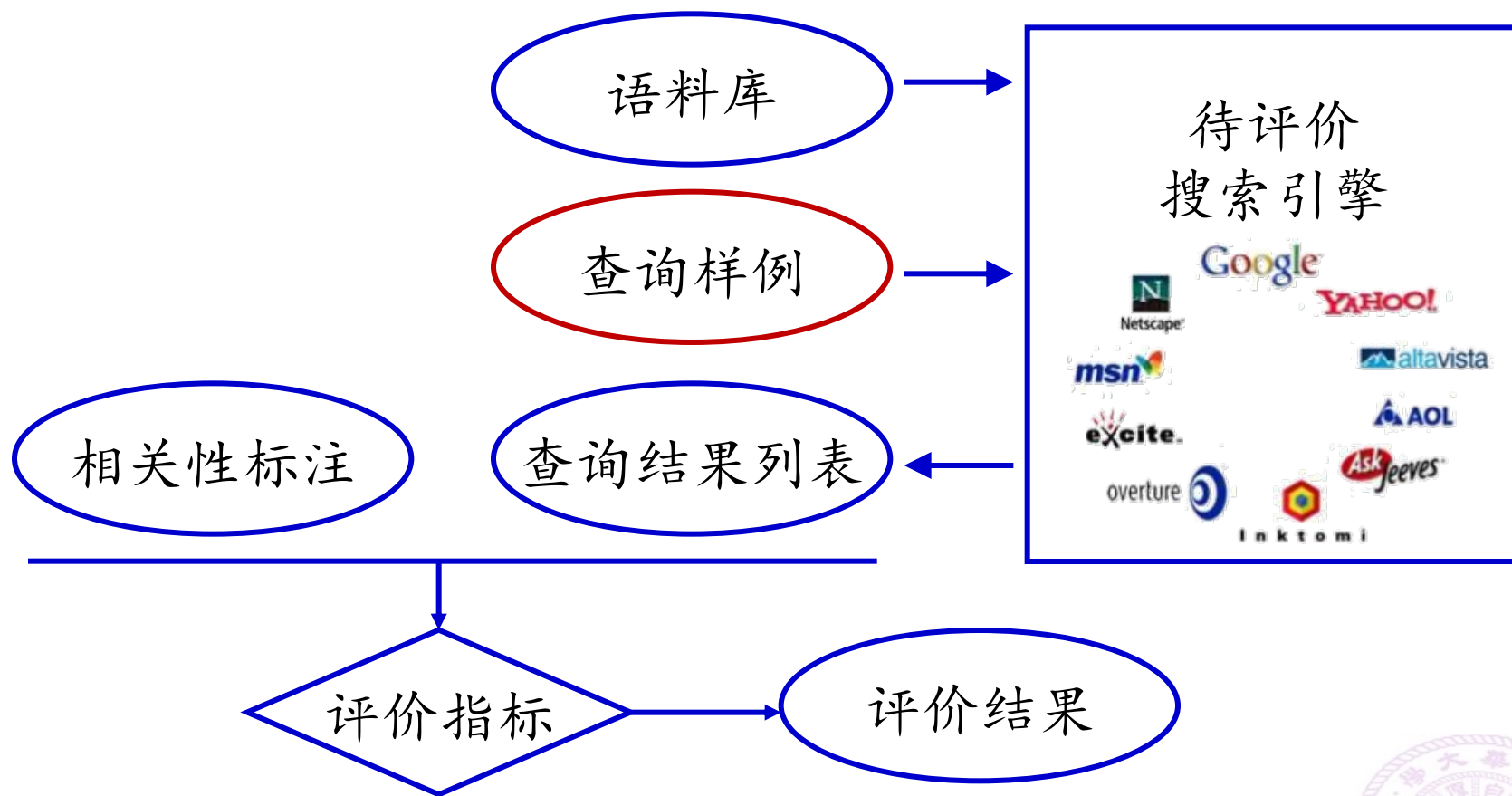
[为什么米3充电这么慢?需要六七个小时才能满,小米3吧,百度贴...](#)

返回小米3吧 为什么米3充电这么慢?需要六七个小时才能满 回复 啊?别的手机一般俩... 充电,不过充电慢,用电也慢,我都两三天一冲,还不错 nverlv23: 怎么我的用电用得那么快!...

tieba.baidu.com/p/2681134860 2013-11-01 - 84%好评



3. 查询样例集合构建



3. 查询样例集合构建

- 用户查询规模庞大
 - ComScore: More than 18.2 billion explicit core searches were conducted per month in U.S.
 - 艾瑞咨询: 每季度中国网页搜索请求量达**775.1**亿次
- 核心问题: 如何采样
 - 真实性: 反映用户实际需求
 - 精确性: 减少相关性标注困难
 - 全面性: 综合评价各方面性能



3. 查询样例集合构建

• 3.1 查询采样的真实性

- 来源：用户查询行为日志
 - TREC 评测：Bing, Yahoo!
 - NTCIR评测：Bing, 搜狗
 - SEWM评测：天网搜索
- 日志收集的隐私保护：AOL案例
- 公开数据资源
 - 查询排行：[百度风云榜](#)，[每日热搜词](#)
 - 搜索日志：SogouQ, WSCD, Yandex



3. 查询样例集合构建

•3.2 查询样例的精确性

```
<top>
<num> Number: 751

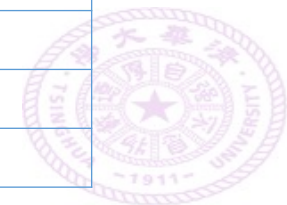
<title> Scrabble Players

<desc> Description:
Give information on Scrabble players, when and where Scrabble is
played, and how popular it has been.

<narr> Narrative:
Give information on the social aspects of the game Scrabble. Scrabble
players may be named or described as a group. Both real and fictional
players are relevant. Mention of a scheduled Scrabble game is
```



Query ID	Query Content
770613	what makes your hands burn
84901	causes for left shoulder and bicep
928755	what year was william shakespeare born
895787	what size sim card does the lg g stylo use
920435	what was the first mammal cloned?
1009016	which elements had complete outer shells



3. 查询样例集合构建

• 3.3 查询采样的全面性

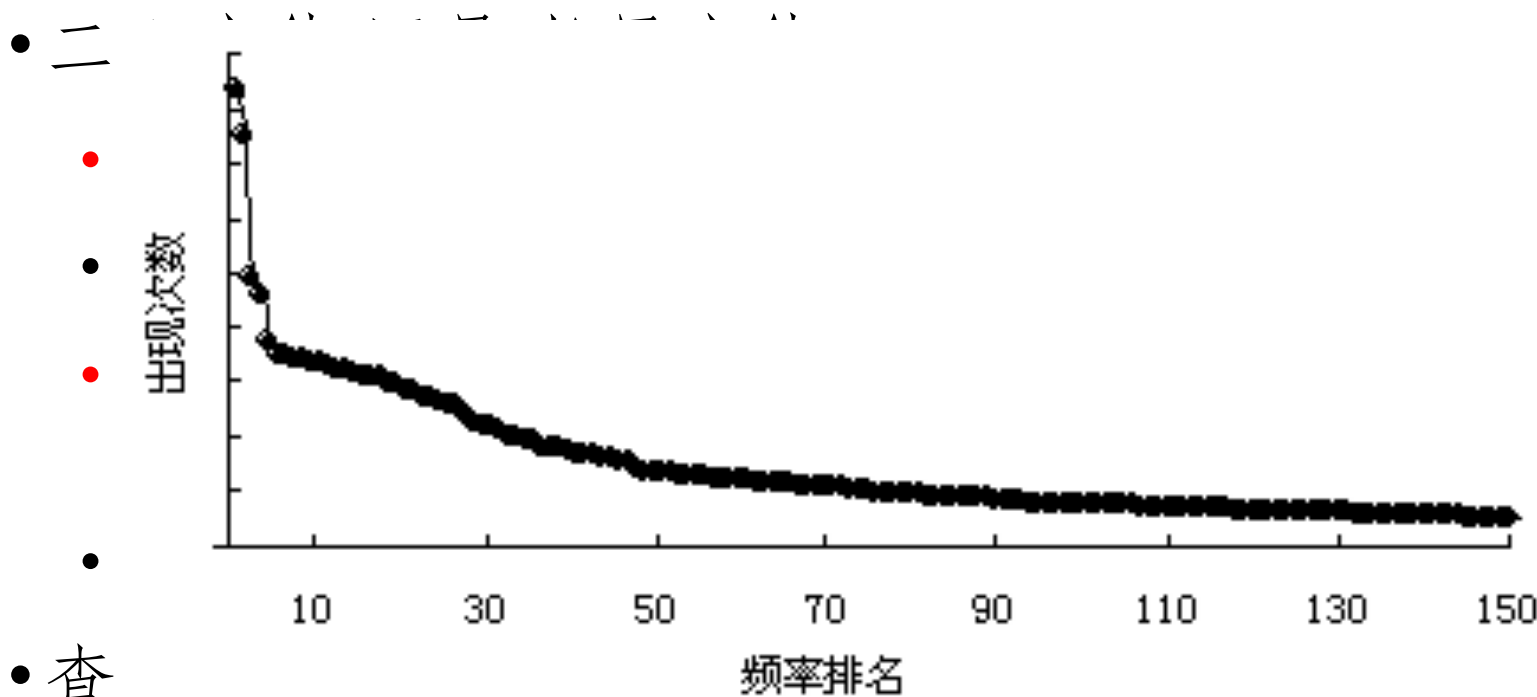
- 用少量的查询样例代表大多数需求类别
- 通常考虑的采样依据有哪些？
 - 查询的内容类别
 - 用户查询内容分类体系：KDDCup 2005
 - 网络内容分类体系：Yahoo! 网页目录
 - 查询的热门程度：查询频度分布
 - 查询的需求类型：用户需求分布



3.3 查询样例集合构建：采样全面性

• 3.3.1 搜索引擎用户查询频度分布规律

- 实际分布数据（搜狗搜索，2008年6月）
 - 总数：1500万；最热门的1万个查询覆盖56%的用户需求



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

- 查询：沟通繁杂的数据环境与丰富的用户意图
- 查询信息需求(**information need**)
 - 用户查询背后的不同类型的信息获取需要
 - 决定了用户使用搜索引擎的满意程度
 - 直接反映在用户与搜索引擎的交互行为上
- 案例：学堂在线主页/**NBA**在线观看/什么是大数据
- 案例：魔兽争霸



3.3.2 搜索引擎用户信息需求分布

- 案例：魔兽争霸
 - 用户1：到达某些特定站点

点击次序	被点击结果的排序	URL
1	9	war3xp.xiyou.net/
结束查询		

点击次序	被点击结果的排序	URL
1	7	www.aomeisoft.com/war3/wc3/
结束查询		

3.3.2 搜索引擎用户信息需求分布

- 案例：魔兽争霸
- 用户2：游戏下载

点击次序	被点击结果的排序	URL
1	4	www.it.com.cn/f/hotweb/053/17/88017.htm
2	3	war3.ogame.net/
3	2	www.gamedge.net/
4	1	war3.cga.com.cn/
5	5	games.tom.com/zhuanti/war3/
6	7	www.aomeisoft.com/war3/wc3/
7	10	war3xp.xiyou.net/
8	魔兽争霸3下载	ewnf.com/1/war.htm

3.3.2 搜索引擎用户信息需求分布

- 案例：魔兽争霸
 - 用户3：获取资讯

点击次序	被点击结果的排序	URL
1	1	war3.cga.com.cn/
2	4	www.it.com.cn/f/hotweb/053/17/88017.htm
3	3	war3.ogame.net/
4	6	www.pcgames.com.cn/fight/warcraft/



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

- 查询的信息需求分类体系 (Broder, 2003)
 - 导航类 (Navigational, Known item search)
 - 查找某个已知存在的页面/资源
 - 信息类 (Informational, Key resource finding)
 - 查找与某个主题相关的关键信息资源
 - 事务类 (Transactional, Task-oriented search)
 - 查找与完成某个特定任务相关的资源



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

• 导航类查询案例

• 主页查找(Homepage Finding)

- 百度 www.baidu.com

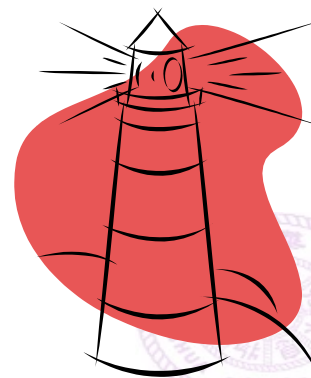
- 北京大学 www.pku.edu.cn

• 其他资源查找(Named-page Finding)

- 清华大学2014年研究生招生简章

- 护照申请表

- 2012年高考数学试卷



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

• 信息类查询案例

• 获取相关信息，没有确定的查询目标

- 香港股市：hk.finance.yahoo.com；cn.biz.yahoo.com/stock/hk.html

• 往往需要不止一个结果

- 麦迪：www.t-mac.cn；
tieba.baidu.com/f?kw=麦迪

• 在查找的过程中逐渐深化认识

- 关节肿胀 => 痛风 => 嘌呤



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

• 事务类查询案例

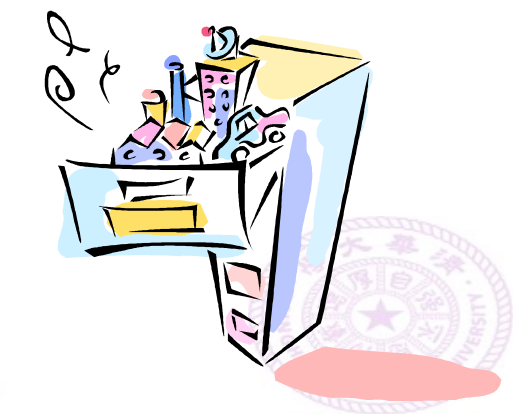
- 与完成特定任务相关，没有确定

- Ultraedit 下载：www.onlinedown.net/soft/22314.htm
www.skycn.com/soft/22314.htm

- 往往一个结果就可以满足需求

- 狂飙：
https://www.iqiyi.com/v_xkt6z

- 垂直搜索引擎服务的对象



3.3 查询样例集合构建：采样全面性

• 3.3.2 搜索引擎用户信息需求分布

- 信息需求分布估计

- Broder: 20%/50%/30% (导航/信息/事务)
- Rose et. al.: 14.7%/60.9%/24.4% (导航/信息/事务)
- 对搜狗搜索的抽样标注：导航类约占30.6%，
信息/事务类约占69.4%

- 中英文用户行为习惯差异

- 采样方法差异

- 信息需求 v.s. 热门程度

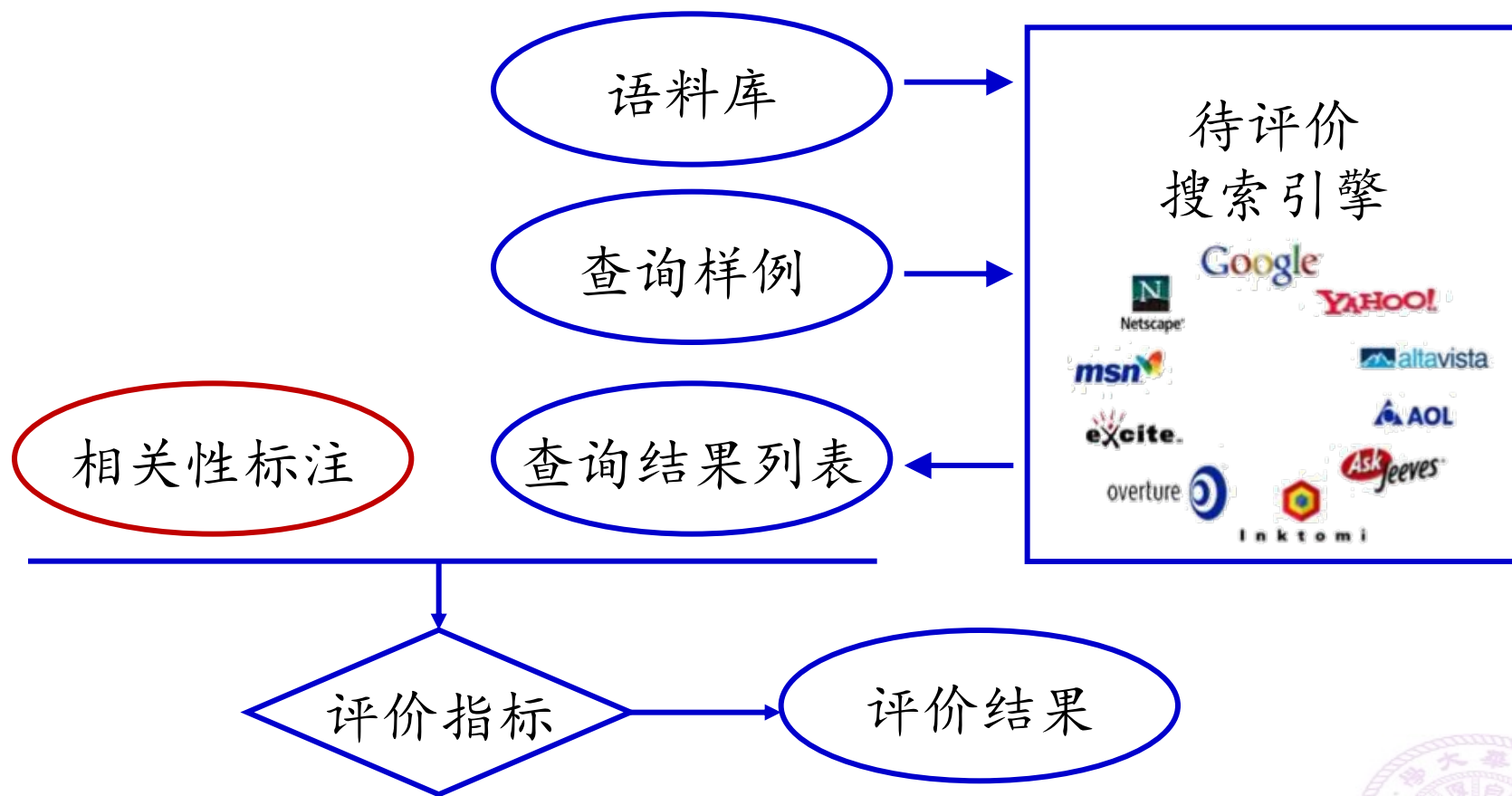


3.3 查询样例集合构建：采样全面性

- 查询热门程度
 - 充分重视热门查询的作用
 - 必须有适当的冷门查询代表
- 查询信息需求
 - 包含导航类、信息类、事务类三种不同类型的查询信息需求



4. 结果相关性标注

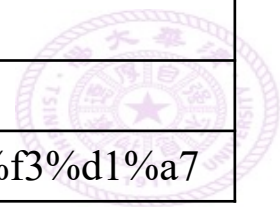


4. 结果相关性标注

•4.1 结果相关性标注的标准

- 需要注意的问题： 相对相关性 v.s. 绝对相关性

序号	结果标题	结果URL
1	欢迎光临清华大学	www.tsinghua.edu.cn
2	清华大学_百度百科	baike.baidu.com/view/1563.htm
3	清华大学_新浪院校库	kaoshi.edu.sina.com.cn/college/c/10003.shtml
4	清华大学图书馆	www.lib.tsinghua.edu.cn
5	清华大学地图	map.baidu.com/#word=%c7%e5%bb%aa%b4%f3%d1%a7
6	國立清華大學	www.nthu.edu.tw
7	清华阳光	www.thsolar.com
8	清华美术学院	ad.tsinghua.edu.cn
9	清华大学公共管理学院	www.sppm.tsinghua.edu.cn
10	清华大学的相关新闻	news.baidu.com/ns?word=%c7%e5%bb%aa%b4%f3%d1%a7



4. 结果相关性标注

• 4.1 结果相关性标注的标准

- 对相关性结果的共性要求
 - 结果所提供的信息应当及时新、真实可靠
 - 结果的标题和摘要应当方便阅读并有效引导用户阅读
- 以信息需求类别为指导

查询信息需求类别	相关结果评价标准
导航类	结果即为用户的检索目标
信息类	结果能够为用户信息需求所涉及的主题提供权威性的信息和引导
事务类	结果能够协助用户顺利完成所需任务



4. 结果相关性标注

• 4.1 结果相关性标注的标准

- 不同标注人员的判定标准存在差异
 - TREC 2008 案例：在可能相关的200个文档中，仅有85个意见一致

	标注人员B：相关	标注人员B：无关
标注人员A：相关	85	51
标注人员A：无关	64	567

• 解决思路

- 查询样例集合构建时添加精确需求描述
- 专业标注人员 v.s. Crowd sourcing



4. 结果相关性标注

• 4.2 结果相关性标注的人力成本

- TREC对人力成本的估计

- 一个规模为**800万**的文档集合；针对**1**个查询主题的相关性评判；需要耗费**1**名标注人员**9**个月的工作时间

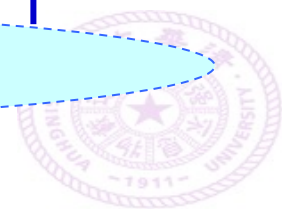
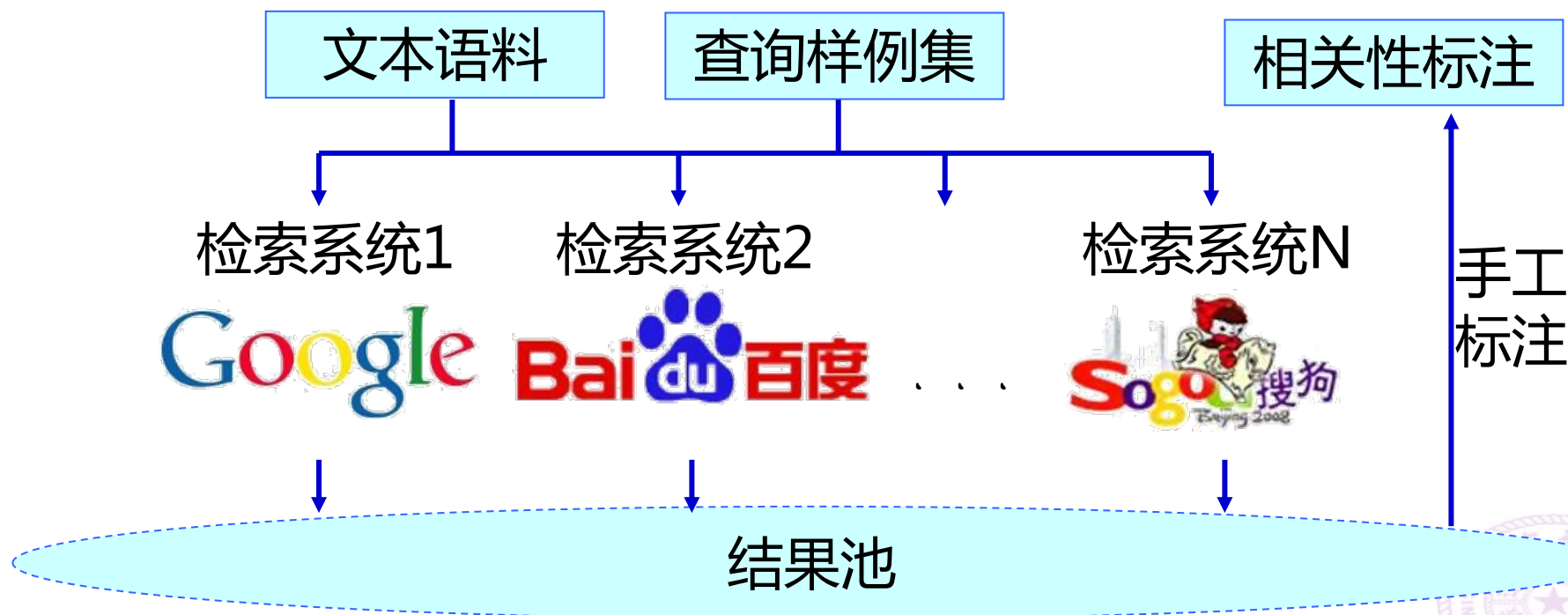
- 如何提高相关性标注集合构建的效率？

- **Pooling**方法：保证评价结果可靠性的基础上减少工作量
 - 假设：没有被任何一个系统检索出的答案对于评价检索系统的性能没有意义



4. 结果相关性标注

- 4.2 结果相关性标注的人力成本
 - 结果池过滤方法(Pooling)



4. 结果相关性标注

• 4.2 结果相关性标注的人力成本

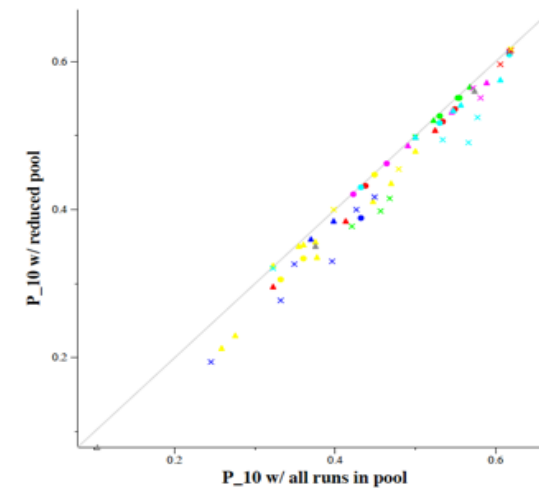
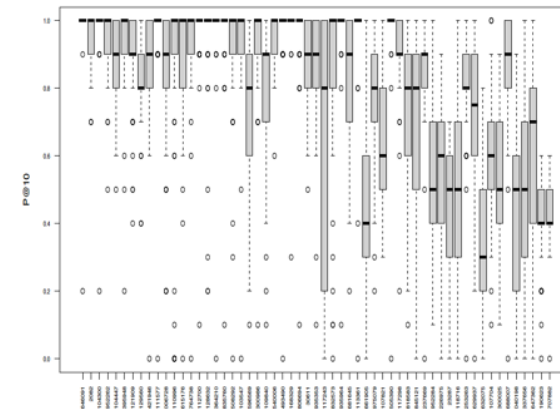
- 结果池过滤方法(Pooling)

- 如何确定结果池深度?

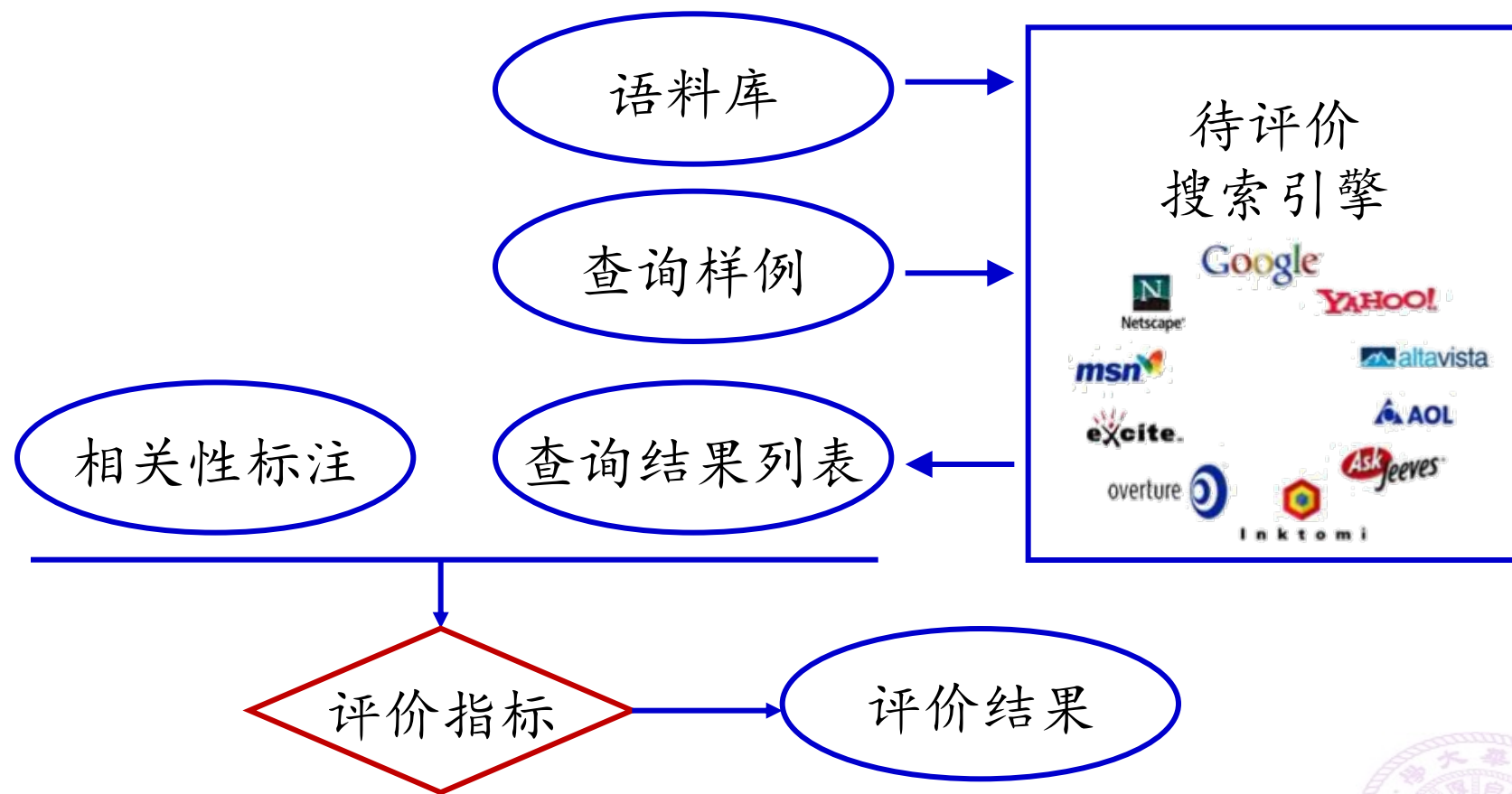
- 过深: 造成较大的标注成本
- 过浅: 无法实现性能区分

- 如何确保结果池复用性?

- LOU: leave-out-uniques test
- 将参与构建结果池的某系统贡献的独特结果完全过滤掉, 评价其性能损失。



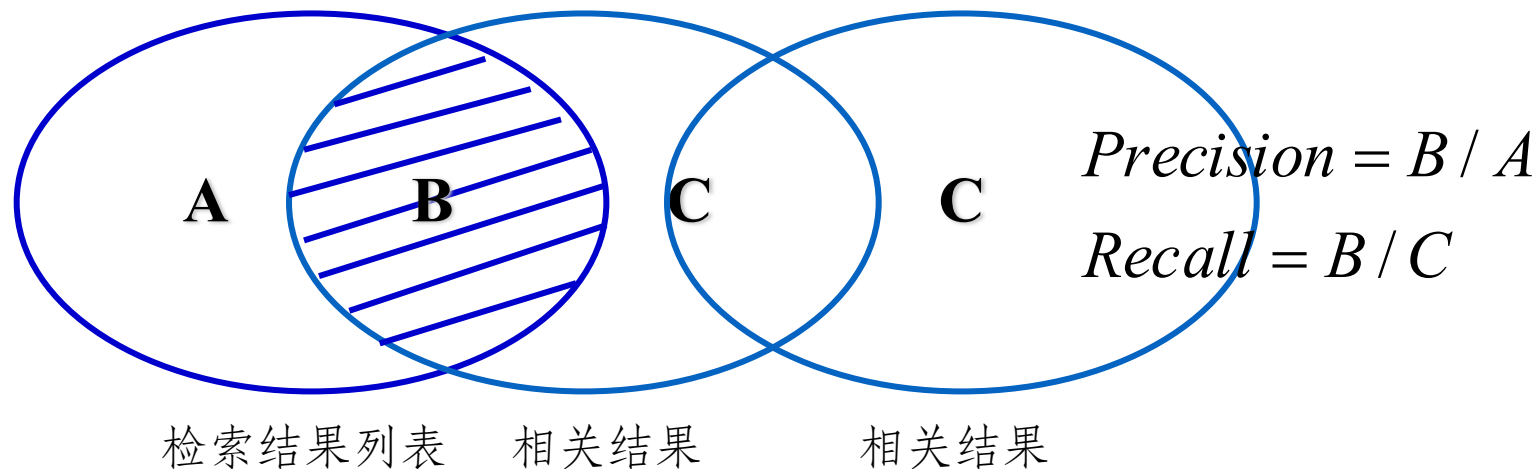
5. 评价指标设计



5.1 信息检索系统评价指标

- 准确率/召回率

- Kent等人1955年提出了关于Precision和Recall的概念
- 准确率(Precision): 找到的是否准确
- 召回率(Recall): 找到的是否全面



5.1 信息检索系统评价指标

- 准确率/召回率

- 垃圾电子邮件识别

- 准确率：有多少识别出的信件真的是垃圾信件
 - 召回率：是否所有的垃圾邮件都被识别出了

- 疾病的医学诊断

- 准确率：诊断患病的病人有多少真的患病
 - 召回率：有多大比例的患病的病人被诊断出

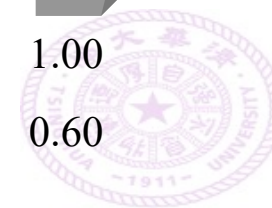
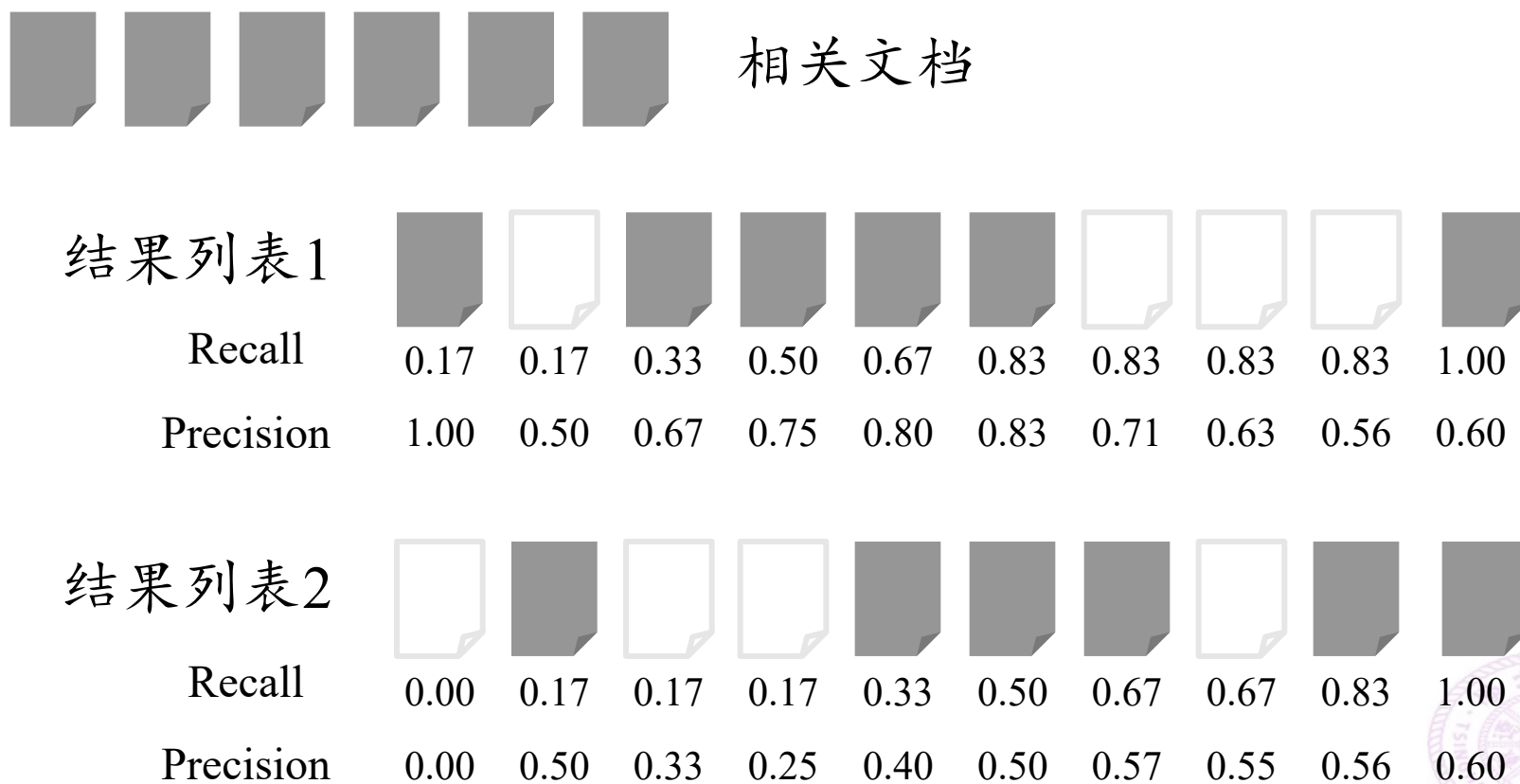
- 信息检索强调“序列”的关系

- 在结果序列中计算准确率与召回率



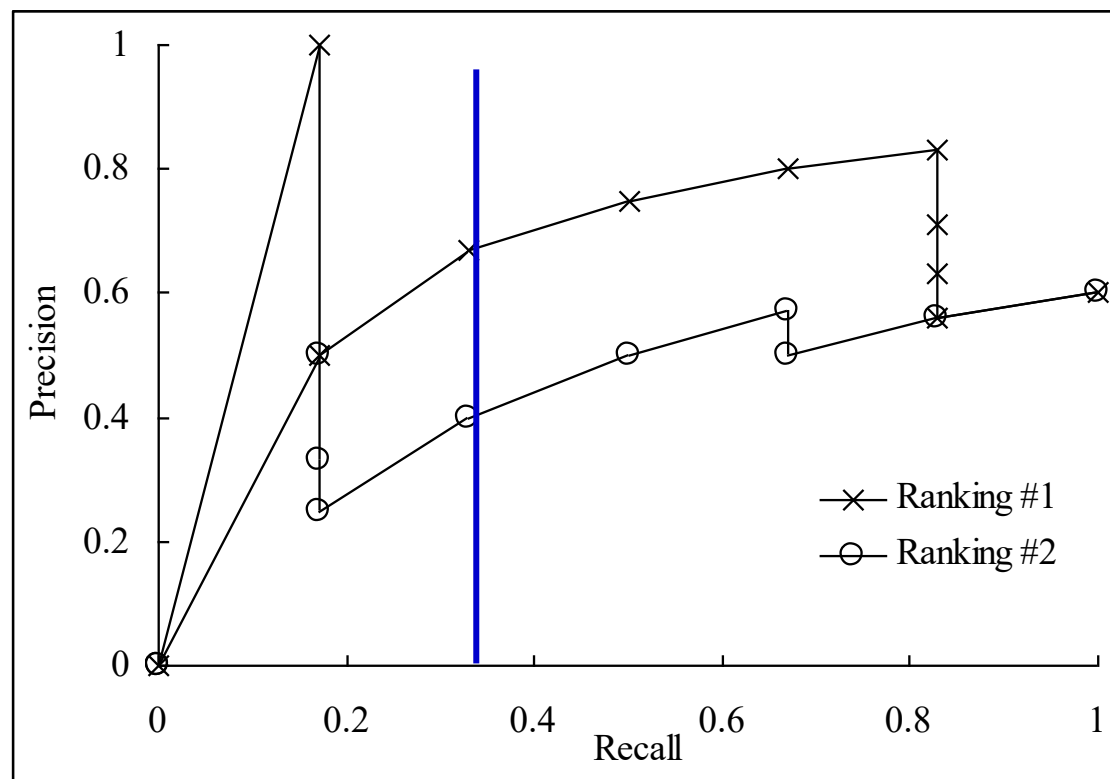
5.1 信息检索系统评价指标

- 检索结果序列中准确率/召回率的计算



5.1 信息检索系统评价指标

- 准确率/召回率曲线
 - 比较不同系统之间的性能差异



5.1 信息检索系统评价指标

- 综合考虑准确率与召回率的指标











- 平均准确率 (Average Precision, AP)

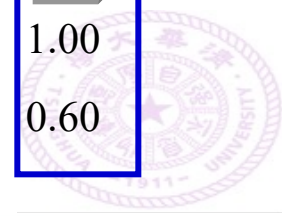
$$AP = \frac{1}{N} \sum_{i=1}^N Precision(i)$$

- N 为已知相关结果的总数

- $Precision(i)$ 为系统找到第 i 个答案时的 Precision

结果列表1











										
Recall	0.17	0.17	0.33	0.50	0.67	0.83	0.83	0.83	0.83	1.00
Precision	1.00	0.50	0.67	0.75	0.80	0.83	0.71	0.63	0.56	0.60



5.1 信息检索系统评价指标










• 平均准确率(AP)的计算

结果列表1

										
Recall	0.17	0.17	0.33	0.50	0.67	0.83	0.83	0.83	0.83	1.00
Precision	1.00	0.50	0.67	0.75	0.80	0.83	0.71	0.63	0.56	0.60

$$AP = (1.00 + 0.67 + 0.75 + 0.80 + 0.83 + 0.60) / 6 = 0.78$$

结果列表3

									
Recall	0.17	0.17	0.33	0.50	0.67	0.83	0.83	0.83	0.83
Precision	1.00	0.50	0.67	0.75	0.80	0.83	0.71	0.63	0.56

$$AP = (1.00 + 0.67 + 0.75 + 0.80 + 0.83) / 6 = 0.68$$



5.2 搜索引擎性能评价指标

- 信息检索评价方法如何应用于搜索引擎？
 - 搜索引擎用户行为的特殊性
 - 约**85%**的用户只翻看搜索引擎返回的前**10**个结果。
 - 大部分用户甚至不会滚动网页查看其他结果
 - 搜索引擎用户信息需求的差异性
 - 导航类信息需求的用户仅关注特定检索目标
 - 信息类信息需求的用户关注全面而权威的信息
 - 事务类信息需求的用户关注自己的任务是否可以顺利完成。

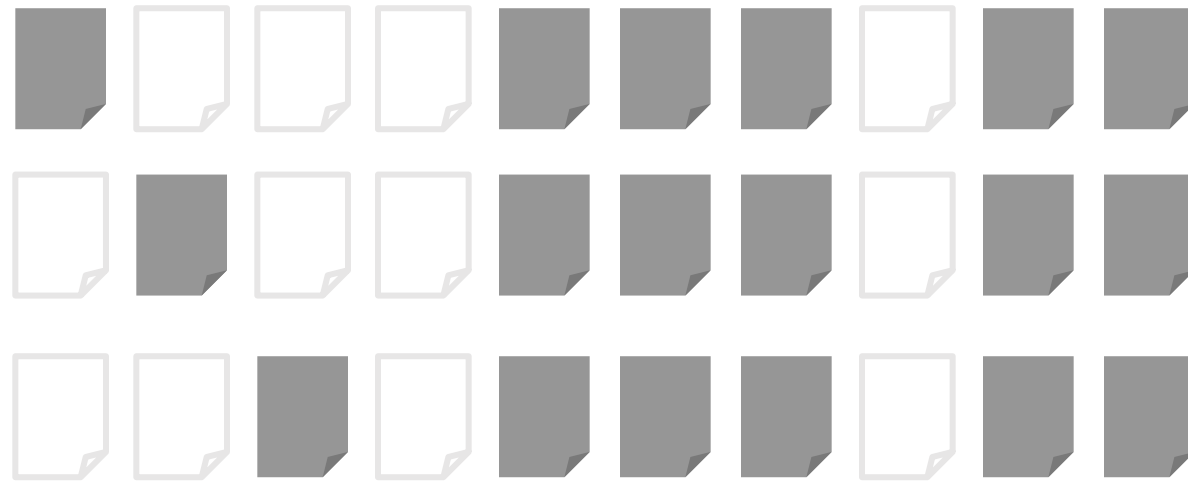


5.2 搜索引擎性能评价指标

• 5.2.1 首位相关结果倒数(Reciprocal Rank)

$$RR = \frac{1}{Rank(1)}$$

出现第一个相关性标注的
排序倒数



对排
名靠
前的
结果
有利

- 适用于导航类型的查询信息需求（用户在找到搜索目标前需要浏览多少结果？）



5.2 搜索引擎性能评价指标

- 5.2.2 前N位准确率(Precision@N)

- N=10, 20: 第一页结果中的准确程度

- N=5: 用户不滚动屏幕时结果的准确程度

结果列表1



结果列表2



结果列表3



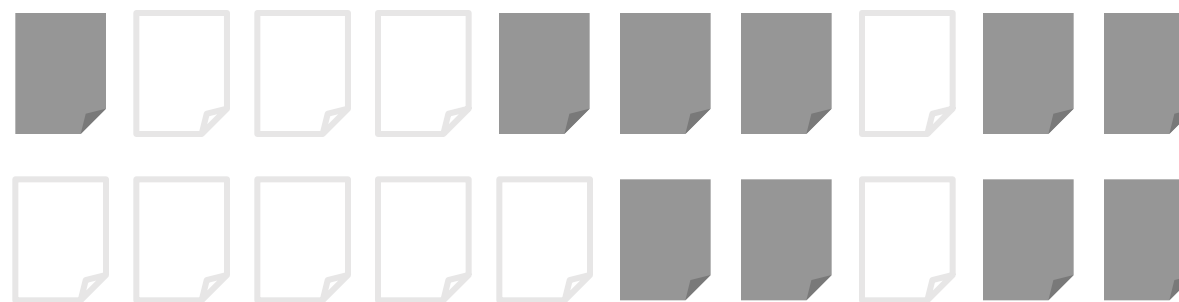
- 适用于信息类型的查询信息需求（结果列表中多大比例信息能够满足用户需求？）



5.2 搜索引擎性能评价指标

• 5.2.3 前N位成功率(Success@N)

- $N=10, 20$: 第一页中是否有满足用户需求的结果
- $N=5$: 不滚动屏幕时是否有满足用户需求的结果
- $N=\infty$: 搜索引擎是否有满足用户需求的结果



- 适用于事务类型的查询信息需求（用户是否能够利用给出的结果完成自己所关注的事务？）



5.2 搜索引擎性能评价指标

• 5.2.4 其他常用评价指标

- NDCG@N:

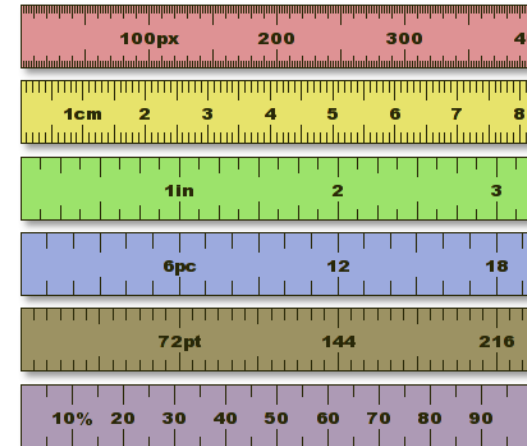
- Normalized Discounted Cumulative Gain
- 对结果进行多级相关性标注时适用

- ERR: Expected Reciprocal Rank

- Supported by Yahoo! and Google
- 可以获取用户行为数据时适用

- B-pref: Binary preference

- 相关性标注不完整时适用



6. 总结与展望

- Cranfield评价体系可能的弊端

- 以“查询—结果”的相关性标注为核心
 - 信息检索系统用户：逐一认真阅读结果文档
 - 搜索引擎用户：尽快获得信息，不一定访问结果网页
- 与用户的搜索引擎使用行为不完全匹配
 - 通过摘要直接获得信息
 - 通过垂直搜索服务获得信息
 - 通过知识图谱/框计算结果直接获得信息



6. 总结与展望

- 不同于Cranfield体系的其他评价方案
 - 在线性能评价 v.s. 离线性能评价
 - 用户偏好性测试
 - Interleaving 测试

Presented Ranking	
1. Napa Valley – The authority for lodging... www.napavalley.com	Click
2. Napa Country, California – Wikipedia en.wikipedia.org/wiki/Napa_Valley	
3. Napa: The Story of an American Eden... books.google.co.uk/books?isbn=...	
4. Napa Valley Wineries – Plan your wine... www.napavalley.com/wineries	
5. Napa Valley Hotels – Bed and Breakfast... www.napalinks.com	Click
6. Napa Balley College www.napavalley.edu/homex.asp	
7. NapaValley.org www.napavallev.org	

完成一题，再接再厉

搜索内容为 **程序员**

查询意图为

1. 程序员_网站 2. 程序员_介绍 3. 程序员_职业 4. 程序员_课程 5. 程序员_招聘

四川在线人才频道

;(2008/04/15) 来源: 四川在线人才频道 对于**程序员**的学历, **程序员**一般要求本科毕业, 有些公司对于自学的**程序员**, 如果有工作经验的话不受限制, 通常测试的是**程序员**掌握的语言技能
http://job.scol.com.cn/info/html/20080721/1216631788218.htm

程序员单身之谜 细数程序员“六宗罪” [软件新闻行...
行业新闻**程序员**单身之谜细数**程序员**“六宗罪”[软件新闻行...CSDN2007-11-16 16:00文/WriterLink(门): 有人曾说过**程序员**是IT行业发展的基石, 这话可算是把**程序员**的角色
http://it.yq.sx.cninfo.net/edu/0711/16/508856.htm

2004软考改革 TOM科技

[报价论坛头条 专题]观点 专栏 登载 访谈 调查 人物 图片 行业观察 海外论坛 壁纸 创投天下 财经 **程序员**高薪之路现在随便登陆一个人才网站, 都可以看到招聘软件人才的信息铺天盖地, 但是不同企业
http://tech.tom.com/zhuanti/ccidedu_tom/rkzx/cxygczl.htm

程序员软件考试专题

留学[简历模板]文化历史[地图推荐]快来网吧学习俱乐部往日头条在线实验职场名人微访谈大学明星超级女声题库下载游情黄健翔网由入门学习园地视觉盛宴 软件资格和水平考试各地查询地址 **程序员**设计 (初级**程序员**级
http://www.enet.com.cn/edu/hot/kaoshi/

招聘PHP程序员/软件工程师-北京无限新空信息技术...
注册企业注册您现在的位置: 首页>职位搜索>北京无限新空信息技术有限公司>PHP**程序员**/软件工程师公司简介所有职位PHP**程序员**/软件工程师发布时间: 2008/08/11 浏览: 3190次
http://www.cnithr.com/personal/user_show_job.php?id=476290

招聘对日开发java程序员-世纪新动信息技术(北京...
注册企业注册您现在的位置: 首页>职位搜索>世纪新动信息技术(北京)有限公司>对日开发java**程序员**公司简介所有职位对日开发java**程序员**发布时间: 2007/06/29 浏览: 15705
http://www.cnithr.com/personal/user_show_job.php?id=365322

SCJP--Sun认证Java程序员考试专题[中国...

首页|资讯中心| IT培训|远程教育|技术专题| 搜索 解决方案|网络产品|学习下载|在线测试|在线实验|IT社区 SCJP--Sun认证Java**程序员**考试专题
http://www.chinaitlab.com/www/special/scjp.asp

www.pudn.com - 最大的源码下载中文网...

程序员5名,企业信息系统设计与开发[威正软件]高级**程序员**2名, **程序员**设计[北京计算机图书]文档管理员20名,编写计算机书籍[惠州华臣.net]文档管理员1名,美工[em beddedsoftware]**程序员**
http://www.programsalon.com/default.asp

左边好+4 左边好+3 左边好+2 左边好+1 难分辨 右边好+1 右边好+2 右边好+3 右边好+4



实验1: 搜索引擎性能评价

- 实验步骤

1. 每人一组

2. 构建查询样例集合：利用网络资源(<http://top.baidu.com>等)和个人使用经验构建查询样例集合，查询样例集合需覆盖不同查询热门程度（冷门/热门）和各种类型的用户查询需求（导航类/信息类/事务类），样例集合的规模为**10**个查询，各类比例为**2:5:3**，并根据个人经验，撰写每个查询样例的信息需求内容。



实验1: 搜索引擎性能评价

- 实验步骤

3. 构建**Pooling**: 学生根据其构建的查询样例集合, 抓取常用的两个中文搜索引擎(百度、必应搜索、搜狗等搜索引擎中选取两个)对这部分查询词的查询结果, 每个搜索引擎抓取查询结果的前十位结果, 并利用这些结果构建**Pooling**。
4. 构建相关性标注集合: 根据步骤2中撰写好的信息需求, 对**Pooling**里的结果进行标注, 标注为“相关”和“非相关”两类即可。



实验1: 搜索引擎性能评价

- 实验步骤

5. 根据标注结果, 依据MAP, P@10, MRR等评价指标对各个搜索引擎的查询性能进行评价, 并对搜索引擎满足不同信息需求的情况加以比较。

评价指标可参考: https://trec.nist.gov/trec_eval/
<https://research.nii.ac.jp/ntcir/tools/ntcireval-en.html>

6. 扩展内容: 可以尝试对搜索引擎处理非中文查询、有错别字查询等情况的不同策略进行分析、比较。
7. 实验截止日期: 3月20日22:00, 提交实验报告、查询样例集合、Pooling集合和相关性标注集合





Welcome to visit our homepage

<http://www.thuir.cn/>

祝身体健康、学习顺利！

