



# 搜索引擎评估与用户行为分析

彭滔  
2012-01-07



## ✓ Who am I

- 彭滔
- [pengtao@baidu.com](mailto:pengtao@baidu.com)
- 百度网页搜索部
  - 用户行为分析组
- 工作范围
  - 搜索排序、算法评估、系统监控、用户建模、流量分析
  - 与日志有关的一切😊



引子：搜索引擎如何评估

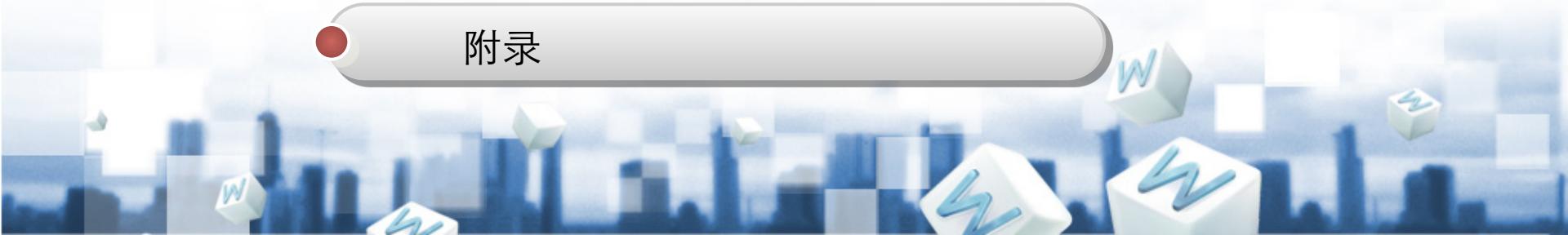
史前时代：一个人战斗

农耕时代：群众的智慧

机器时代：是骡子是马

总结

附录



## 1. 搜索引擎如何评估？

✓ 介绍评估之前

– 搜索在2010年成长互联网第一大应用 (81.9%)

- CNNIC, 《中国互联网络发展状况统计报告》, 2011

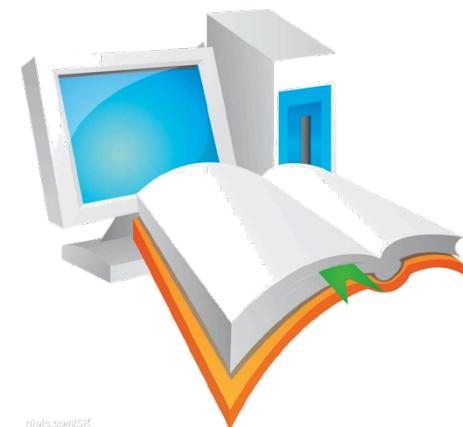
– Google effects on memory

- 面对困难问题：想问题 v.s. 想电脑

- 预期能用电脑：怎么解决问题 v.s. 哪里有电脑

- (Sparrow, 2011)

- *The Internet has become a primary form of external or transactive memory, where information is stored collectively outside ourselves.*



nipia.com/SX

# 1. 搜索引擎如何评估?

- ✓ 搜索引擎的评估
  - 相关性
    - Query  $\leftrightarrow$  url
  - 角度
    - 快、准、全、新
- ✓ 经典方法
  - MAP
  - DCG
  - nDCG
  - ERR
  - ...

Baidu 百度 新闻 网页 贴吧 知道 MP3 图片 视频 地图 更多▼	
给我一片关于春天的作文,要长的,快点	百度一下
<a href="#">谁给我提供一篇写春雨的作文, 谢谢大家帮忙啦, 快点很...百...</a> ☆	1/1 × 2
我无意地抬起头, 遥望见一片新绿。那是树, 那是草, 那是春天的使者。我不由得快步走到那儿。咦, 小草呢? 我奇怪了, 它们为何不见了? 仔细看看后, 笑意不觉...	
<a href="#">zhidao.baidu.com/question/228672053.html 2011-3-3 - 百度快照</a>	
<a href="#">春天的作文 关于春天的作文 描写春天的作文 - 作文网 - 作文网</a> ☆	1/2 × 3
春天的作文栏目是作文网免费为您收集关于春天的作文, 包括大量精选的描写春天的作文, 作文网还为网友提供在线交流平台, 让您了解最新作文热点话题预测。	
<a href="#">zuowen.chazidian.com/chuntian 2011-12-23 - 百度快照</a>	
<a href="#">_春色; 我爱妈妈; 观察日记; 泡蒜; 一张老照片 (范思学) - 作文网</a> ☆	1/3 × 1
作文网专稿未经允许不得转载猜猜看今天的能力课上, 张老师神秘地在讲台上摆了三杯水, 说: 这三杯水里分别装着糖水、白醋、盐水, 我请三个同学来尝这三杯水。同学...	
<a href="#">www.zuowen.com/e/20101018/4cbbb76057bf3.shtml 2011-12-18 - 百度快照</a>	
<a href="#">要一篇关于春的作文, 400字, 不要写的太好的, 快... - 百度知道</a> ☆	1/4 × 2
三、四句描述了一群活泼的儿童在大好的春光里放风筝的生动情景, 他们的欢声笑语, 使春天更加富有朝气。这首诗落笔明朗, 用词洗练。全诗洋溢着欢快的情绪, 给读者以...	
<a href="#">zhidao.baidu.com/question/257393784.html 2011-5-8 - 百度快照</a>	
<a href="#">春天的作文 春天作文, 关于春天的作文 - 中国教育文摘</a> ☆	1/5 × 2
春天的作文由中国教育文摘为您提供, 包括春天作文, 关于春天的作文等, 如果您觉得春天的作文不错的话, 请把春天的作文加到您的收藏夹里面以便下次访问。	
<a href="#">www.eduzhai.net/zuowen/special/chuntiande ... 2011-10-14 - 百度快照</a>	
<a href="#">【三题故事】公园, 春天, 小船【这曾经是一片暑假作文。。。O... - 百度知道</a> ☆	1/6 × 2
6条回复 - 发帖时间: 2011年7月1日	
春天, 小船【这曾经是一片暑假作文。。。OTL】_若君 1楼 RT 其实这是仆...发现原来春天的第一阵风吹起来了, 原本无味的风因为空气中的香味而变得好闻起来, ...	
<a href="#">tishi.baidu.com/n/112704R123 2011-10-28 - 百度快照</a>	

5.0667

引子：搜索引擎如何评估



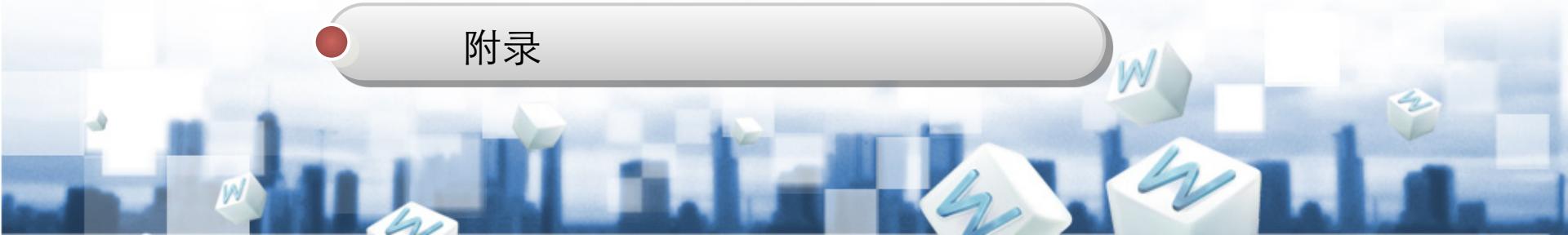
史前时代：一个人战斗

农耕时代：群众的智慧

机器时代：是骡子是马

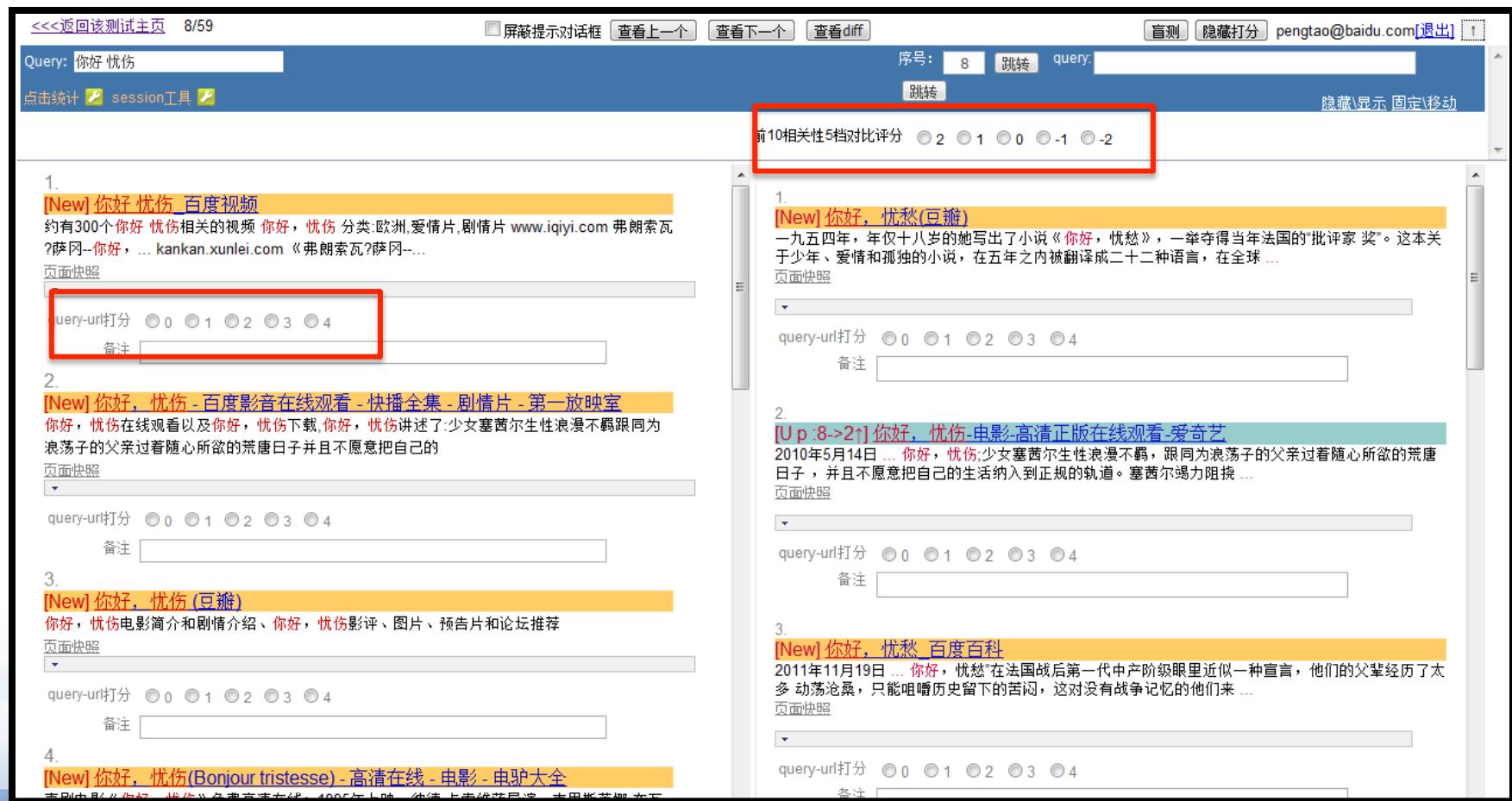
总结

附录



## 2. 史前时代：一个人战斗

## ✓ Side by side 评估



The screenshot shows a side-by-side review interface with two columns of search results for the query "你好 忧伤".

**Left Column (Original Result):**

- Result 1:** [New] 你好 忧伤\_百度视频  
约有300个 你好 忧伤 相关的视频 你好，忧伤 分类:欧洲,爱情片,剧情片 www.iqiyi.com 弗朗索瓦·萨冈--你好, ... kankan.xunlei.com 《弗朗索瓦·萨冈--...  
页面快照
- Result 2:** [New] 你好，忧伤 - 百度影音在线观看 - 快播全集 - 剧情片 - 第一放映室  
你好，忧伤 在线观看以及 你好，忧伤 下载. 你好，忧伤 讲述了:少女塞茜尔生性浪漫不羁跟同为浪荡子的父亲过着随心所欲的荒唐日子并且不愿意把自己的  
页面快照
- Result 3:** [New] 你好，忧伤(豆瓣)  
你好，忧伤 电影简介和剧情介绍、你好，忧伤 影评、图片、预告片和论坛推荐  
页面快照
- Result 4:** [New] 你好，忧伤(Bonjour tristesse) - 高清在线 - 电影 - 由驴大全  
喜剧电影 你好，忧伤 1954年上映 德国 上索维莱导演 吉恩斯勒编 布瓦

**Right Column (Modified Result):**

- Result 1:** [New] 你好，忧愁(豆瓣)  
一九五四年，年仅十八岁的她写出了小说《你好，忧愁》，一举夺得当年法国的“批评家奖”。这本关于少年、爱情和孤独的小说，在五年之内被翻译成二十二种语言，在全球 ...  
页面快照
- Result 2:** [Up 8->21] 你好，忧伤-电影-高清正版在线观看-爱奇艺  
2010年5月14日 ... 你好，忧伤 少女塞茜尔生性浪漫不羁，跟同为浪荡子的父亲过着随心所欲的荒唐日子，并且不愿意把自己的生活纳入到正规的轨道。塞茜尔竭力阻挠 ...  
页面快照
- Result 3:** [New] 你好，忧愁\_百度百科  
2011年11月19日 ... 你好，忧愁 在法国战后第一代中产阶级眼里近似一种宣言，他们的父辈经历了太多动荡沧桑，只能咀嚼历史留下的苦闷，这对没有战争记忆的他们来 ...  
页面快照

**Common Interface Elements:**

- Top navigation bar: <<< 返回该测试主页 8/59, 屏蔽提示对话框, 查看上一个, 查看下一个, 查看diff, 盲测, 隐藏打分, pengtao@baidu.com[退出], 跳转.
- Search bar: Query: 你好 忧伤.
- Rating section: 前10相关性5档对比评分 (radio buttons: 2, 1, 0, -1, -2). The "1" button is highlighted with a red box.
- Bottom rating section: query-url打分 (radio buttons: 0, 1, 2, 3, 4) and 备注 input field. The "1" button is highlighted with a red box.

## 2. 史前时代：一个人战斗

### ✓ 搜索引擎人工评价：

– 研发人员提出一种新排序策略（方法）

- 实验版本E v.s. 线上基准C

– 10000

- 从log中抽取10000个query，向E和C两个版本同时查询

– 1000

- 10000个query中，有1000个结果排序是不同的

– 100

- 1000个diff中，PM挑选100个进行人工review

- 30 (good) : 50 (same) : 20 (bad)



PM

## 2. 史前时代：一个人战斗

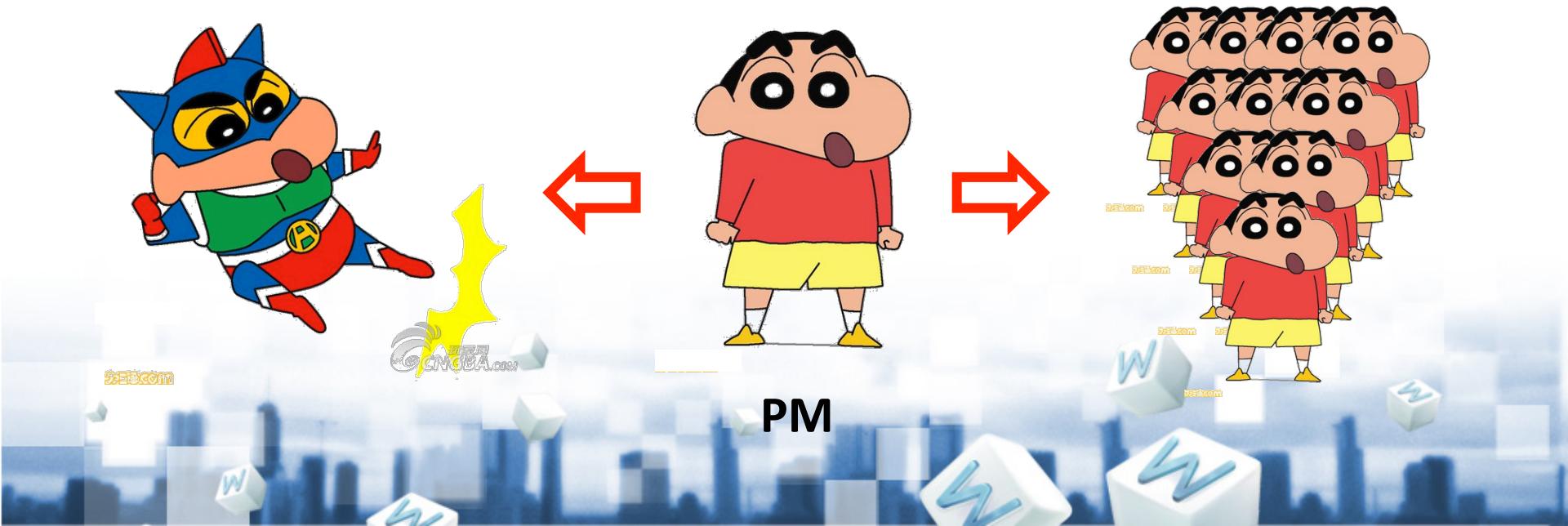
### ✓ 史前时代的惆怅

#### - 数量

- 百度每日的搜索量 v.s. pm评估的query量

#### - 时间

- “下周给出评估结论。”



引子：搜索引擎如何评估

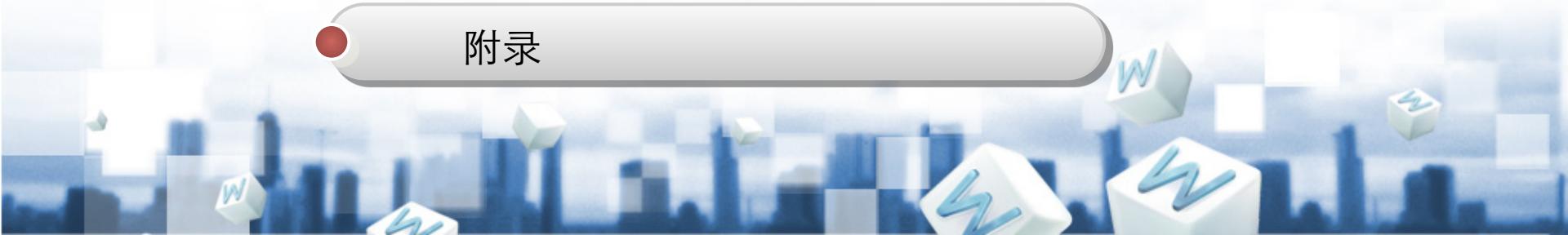
史前时代：一个人战斗

农耕时代：群众的智慧

机器时代：是骡子是马

总结

附录



✓ 众包 (crowdsourcing)

- 与外包公司签订协议，募集兼职人员 (evaluator)。
- 将一个搜索评估任务分解为许多小任务。
- 每个evaluator完成若干任务，最后汇总。

✓ WSE：百度搜索评测平台

- 内部验证
- 线上小规模实验
- 迭代和修改
- 常规运行与监控

## 2. 农耕时代：群众的智慧

## ✓ WSE平台：evaluator的工作环境



The screenshot displays two main windows illustrating the WSE platform's evaluator environment.

**Top Window (Search Results):**

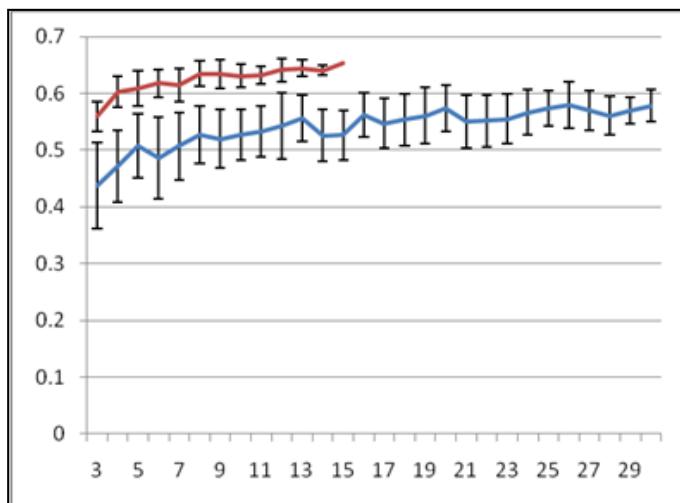
- Header: <<返回该测试主页 1/100, 查看下一个, 隐藏打分, 隐藏结果, jincihang@baidu.com[退出]
- Query: 工号9527为您服务 网页抓取时间: 2011-01-15
- URl: <http://apps.hi.baidu.com/share/detail/14102631> (highlighted with a blue box)
- Buttons: 请打分 (dropdown), badcase (dropdown set to 未选择), 百度搜索, 合成搜索

**Bottom Window (Sharing Application):**

- Header: Baidu空间 应用 用户名: 登录 注册
- Left Sidebar: HOHO, 照片PK, 分享, 投票, 测试, 礼物
- Content Area:
  - 分享: 热门分享, 最新分享, 好友的分享, 我的分享, + 我要分享, 如何分享?, 问题反馈
  - 帖子: 枫714茗的分享 > 分享 (highlighted with a blue box)
  - 帖子内容: 超刁骂人歌曲\_工号9527为您服务 (highlighted with a blue box)

### 3.农耕时代：群众的智慧

✓ 众包Lesson1:  
– 培训比人数重要



#### 百度相关性评估培训 : Query-Url 打分

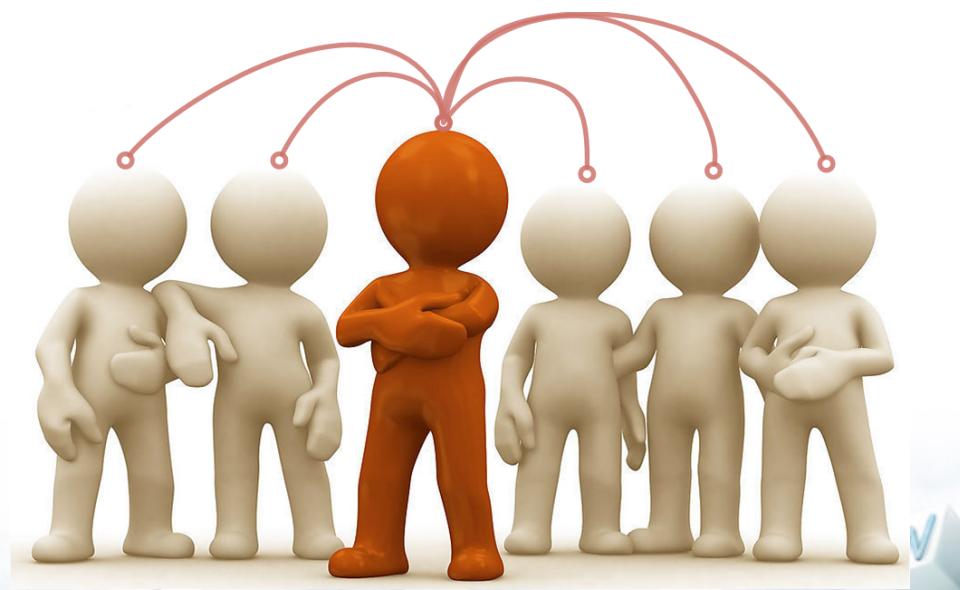
##### 目录

1. 三个步骤.....	2s
2. 五个分值.....	2s
2.1 4—官方原创.....	2s
2.2 3—很棒.....	4s
2.3 2—不错.....	5s
2.4 1—有些帮助.....	6s
2.5 0—很少的帮助、甚至完全没价值.....	7s
3. 打分的注意事项 .....	8s
3.1 跟随用户走到最后一步，判断结果能否真正满足.....	8s
3.2 不确定时，选低的.....	8s
3.3 结果是搜索结果，再判断.....	8s
3.4 注意已有资源的多少.....	8s
3.5 多思考多练习.....	8s
4. 评估维度的运用 .....	8s
5. 搜索评测平台上的注意事项.....	9s
5.1 无法判断或空白，抛弃.....	9s
5.2 判0分时一定要选择原因.....	9s
5.3 两类死链.....	10s
5.4 参考网页抓取时间.....	10s
5.5 页面异常，打开Url查看.....	10s
6. Query 的判断过程举例 .....	10s



✓ 众包Lesson2:

- 人是不可靠的
- 相信共识，不相信人品



✓ 众包Lesson3:

- 机制比技术手段重要
- 不同任务类型的均衡
  - Economics: 理性人假设
- 不同评估者任务量的均衡
  - 我们: 优化正确率
  - evaluator: 优化收入

### 3.农耕时代：群众的智慧

- ✓ WSE的产出
  - 数百评估者
  - 每周10w以上的标注
  
- ✓ 关于crowdsourcing的更多
  - reCaptcha
  - Amazon Mechanical Turk
  - ESP Game
  - Human computation

```
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt(
        "What's fun to see in New York City?
        Ideas so far: " + ideas.join(", "))
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
```

引子：搜索引擎如何评估

史前时代：一个人战斗

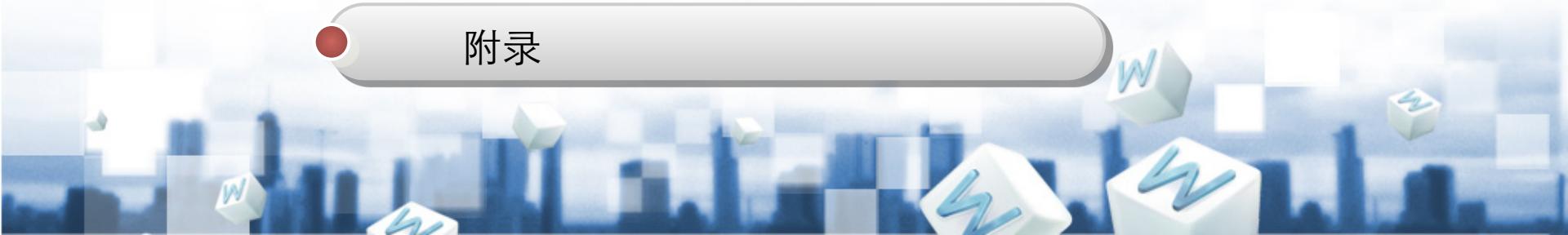
农耕时代：群众的智慧



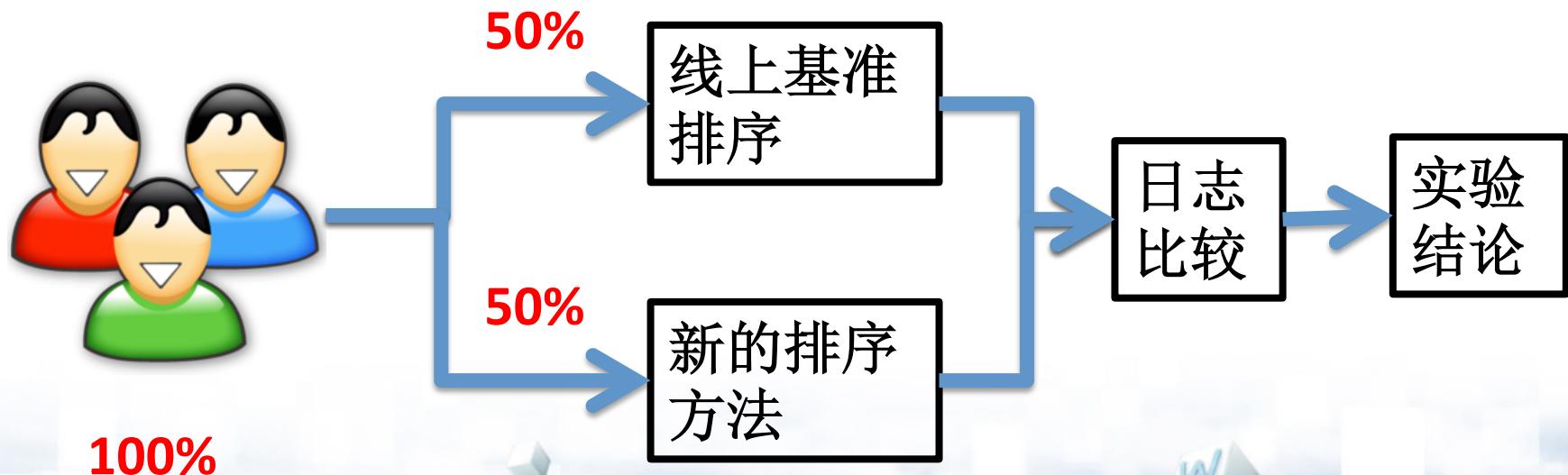
机器时代：是骡子是马

总结

附录



- ✓ 是骡子是马，拉出来溜溜！
  - 利用真实用户反馈，直接验证策略的好坏
  - AB testing, Bucket testing



## 3.机器时代：是骡子是马

- ✓ 为什么要采用AB testing的方法?
  - 因为某些改变，很难人工直接评价

### 每周上网作文 - 走向北大俱乐部 ☆

也给我注入了新的活力。坐在培优班的...快点写作业。”我听后，吓得我直哆嗦，...妈妈听了白了我一眼就走了，我又...我十岁，老师要我以“下雨的时候”...

[demo.whnewnet.com/zxbd\\_old/news.asp?act=c](http://demo.whnewnet.com/zxbd_old/news.asp?act=c) ... 2011-12-12 - 百度快照

### 急求500字作文5篇，快点，500字急求一篇500~600字的作文，题... ☆

2011年12月18日...这时老板出来了，问道：“...春天时，我想放风筝。我会在明媚的下午，独自去长满青草的广场。因为我...急求500字作文5篇，快点，500字急求一篇500~...

[www.53331.com/html/x782tvf.html](http://www.53331.com/html/x782tvf.html) 2011-12-18 - 百度快照

### (╯°□°╰)能不能用英文写一篇关于鲍春来的作文？不用太长... ☆

当前位置：> ----->能不能用英文写一篇关于鲍春来的作文？不用太长-国语4 搜索----->能不能用英文写一篇关于鲍春来的作文？不用太长-国语...

[xrs.net/rs/100036455.htm](http://xrs.net/rs/100036455.htm) 2011-12-22 - 百度快照

### 【写景作文】春天作文|关于春天的作文，700字作文-新东方宝典 ☆

春天作文|关于春天的作文 春天在哪里呀？春天在哪里？春天在那小溪里，春天在那...我爱春天，爱她蕴育万物的精神。上一篇：美丽的徐州 下一篇：我爱海南 ...

[tool.xdf.cn/zw/result\\_ctzwyctdz.html](http://tool.xdf.cn/zw/result_ctzwyctdz.html) 2011-11-27 - 百度快照

### 请大家给我一篇写春的作文要求600字 爱问知识人 ☆

请大家给我一篇写春的作文要求600字 作文要求 远离一场浑然不觉的梦魇，一个与春天即将邂逅的黎明在挣扎中复苏。起身，梳洗，出门。徜徉在和煦的晨曦里，沐浴在香薰...

[iask.sina.com.cn/b/10869873.html](http://iask.sina.com.cn/b/10869873.html) 2011-12-22 - 百度快照

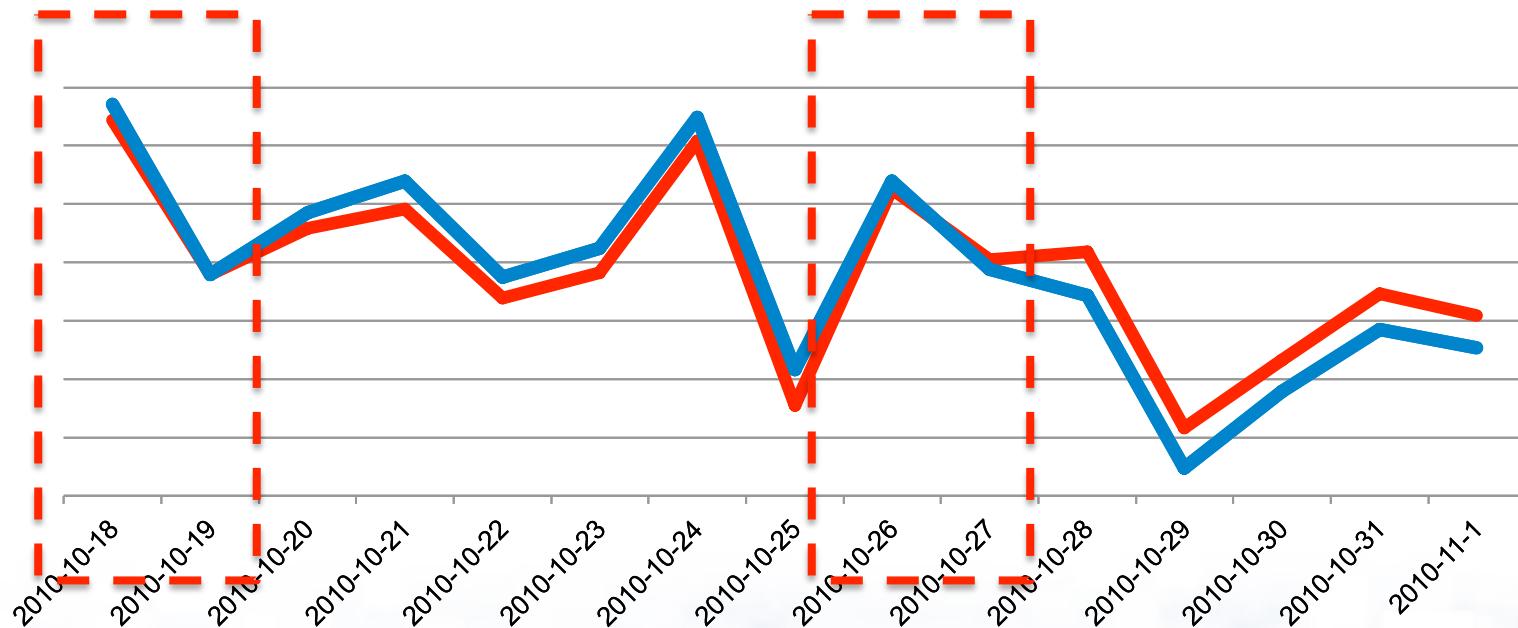
[1](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [下一页](#)

找到相关结果约61,000,000个

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [下一页>](#)

找到相关结果约710,000个

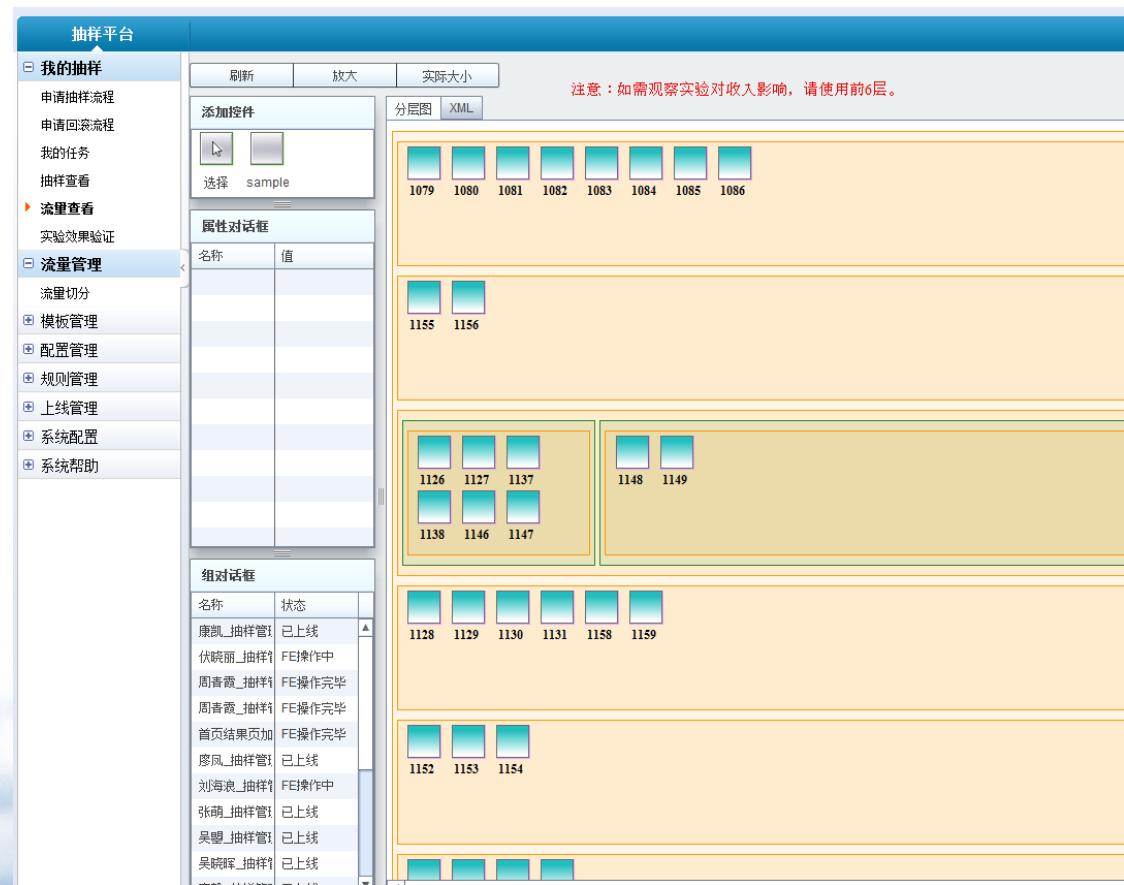
- ✓ 为什么要采用AB testing的方法?
  - 因为单个策略的效果很微弱



### 3.机器时代：是骡子是马

#### ✓ AB testing的流程

- 准备实验：模块+日志
- 流量分配
- 实验上线
- 日志分析



The screenshot shows the Baidu Sampling Platform interface. On the left, a sidebar menu includes '我的抽样' (My Sampling), '申请抽样流程' (Apply Sampling Process), '申请回滚流程' (Apply Rollback Process), '我的任务' (My Tasks), '抽样查看' (Sampling View), '流量查看' (Traffic View), and '实验效果验证' (Experiment Effect Verification). The '流量管理' (Traffic Management) section is currently selected. The main area displays a grid of colored squares representing different sampling groups or users. A legend at the top right indicates colors for different categories. Below the grid, a table lists experiment names and their current status.

名称	状态
康凯_抽样管	已上线
仇晓丽_抽样管	FE操作中
周青霞_抽样管	FE操作完毕
周青霞_抽样管	FE操作完毕
首页结果页加	FE操作完毕
廖凤_抽样管	已上线
刘海浪_抽样管	FE操作中
张萌_抽样管	已上线
吴鹏_抽样管	已上线
吴晓晖_抽样管	已上线

✓ AB testing的日志分析模式

– 1T 原始数据

- cubeproducer, disql, hadoop

– 1G olap数据

- infobright, mondrian

– 1M 终端数据

- ABreport

### 3.机器时代：是骡子是马

#### ✓ AB testing指标的解读

##### – 两种思路

- 一个指标：
  - Overall Evaluation Criteria
    - » (*Crook, 2009*)
  - Queryrank:
    - » 用户行为聚类
    - » 每个类别上训练决策树
- 多个指标：
  - 利用多个指标交叉验证，综合解读
  - 人工决策树



### 3.机器时代：是骡子是马

- ✓ 海量数据解读的经验
  - 在解读前给出预期
    - 数据可以海量，结论只有一个
  - 检查一切
    - 实验设计
      - 空转：AA test
      - 反转
    - 环境检查
      - 日志解析的差别
      - 代码效率的差别

```
try:  
    record = parser(line)  
except:  
    continue
```

### 3.机器时代：是骡子是马

- ✓ 海量数据解读的经验
  - 代码效率的差别



Baidu 百度 新闻 网页 贴吧 知道 MP3 图片 视频 地图 更多▼

宝马 百度一下

**BMW中国官方网站**  
全新BMW 1系运动型两厢轿车炫酷上市,UN1QUE.唯1.推出都市、运动两款车型,个性各选.12.31前在线订购全新BMW1系运动型两厢轿车苹果限量大礼  
[www.bmw.com.cn](http://www.bmw.com.cn) 2012-01 - 品牌推广

运动设计套装 都市设计套装 点击观看视频

● 全新BMW M5雷霆上市,铸就先发之势  
生于赛道深藏不露,新车与跑车兼得,以强势的姿态证明实力

● 全新BMW 5系旅行轿车,平衡自从容  
收放自如,境由心生,突破极限,超越期望,您旅途中的新乐趣

官方微博 BMW Z4 资讯二手车 车型比较 BMW M

全新宝马 1系炫酷上市 UN1QUE.个性各选 [www.bmw.com.cn](http://www.bmw.com.cn)

全新BMW 1系运动型两厢车上市,预订有好礼,首创BMW设计套装,运动和城市间自由选择。

**【宝马X6】宝马X6\_宝马X6报价\_宝马X6论坛\_油耗-易车网**  
宝马X6频道为您提供宝马X6最新报价,宝马X6参数配置,宝马X6图片,宝马X6油耗,二手宝马X6,宝马X6经销商信息,同时还提供宝马X6论坛和关于宝马X6怎么样等信息,查宝马...  
[car.bitauto.com/baomax6/](http://car.bitauto.com/baomax6/) 2011-12-23 - 百度快照

**【宝马】宝马报价\_宝马价格\_汽车之家**  
汽车之家报价库提供宝马报价,宝马价格等宝马相关信息,了解最新宝马信息请登录汽车之家报价库。宝马共13个在产车系,86款在产车型  
[car.autohome.com.cn/price/brand-15.html](http://car.autohome.com.cn/price/brand-15.html) 2012-1-2 - 百度快照

**宝马\_百度百科**  
BMW,全称为巴伐利亚机械制造厂股份公司(德文: Bayerische Motoren Werke AG),是德

jumpout\_arthur 百度首页

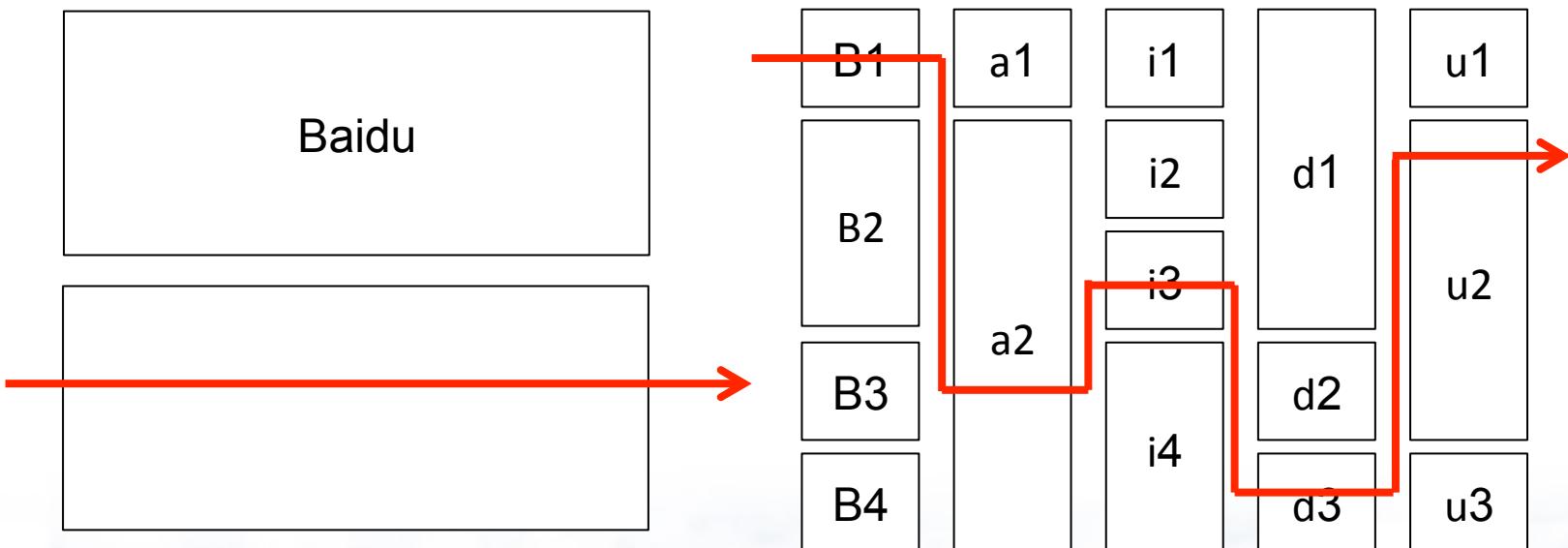
全新BMW 1系  
预约试驾  
姓名  
手机  
邮箱  
有无驾照  
预约试驾  
成功提交后,我们会与您联系。  
分享到:

推广链接  
渣打银行贷款利率优惠,限时3...  
最高50万,最快1天放款,最长5年,轻松月供!渣打银行无担保个人贷款。  
[scbank.omsys.com.cn](http://scbank.omsys.com.cn)

▶来百度推广您的产品  
咨询热线: 400-800-8888  
[e.baidu.com](http://e.baidu.com)

✓ 流量竞争

- 50% v.s. 50% : 首页改版
- 其他实验怎么办?



引子：搜索引擎如何评估

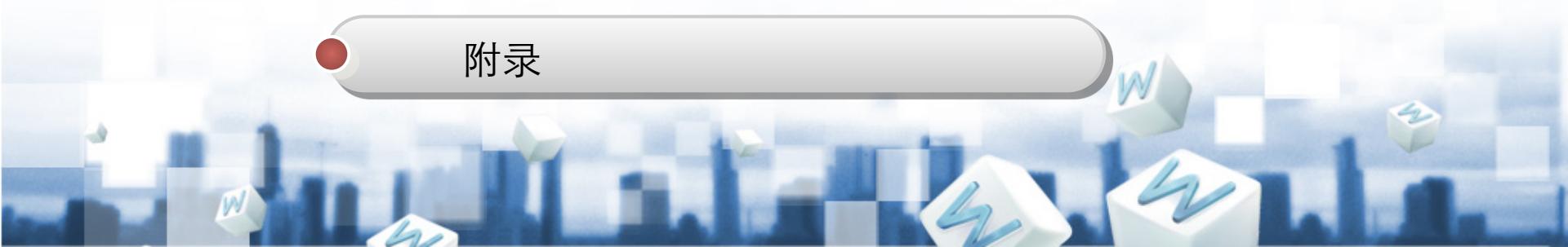
史前时代：一个人战斗

农耕时代：群众的智慧

机器时代：是骡子是马

总结

附录



- ✓ 搜索引擎评估
  - 相关性：快，准，全、新
  - 经典方法：DCG
  
- ✓ 百度实践
  - 原始时代：PM review
  - 农耕时代：crowdsourcing
  - 机器时代：AB testing

✓ 挑战：

- 人工评价 v.s. AB testing 指标存在冲突
- 交互与社会化元素如何评价
- 短时用户喜好 v.s. 长期用户成长

谢谢！



引子：搜索引擎如何评估

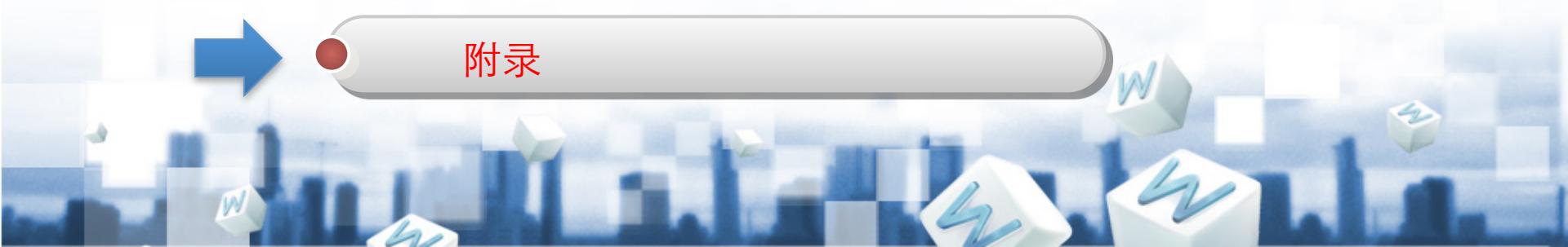
史前时代：一个人战斗

农耕时代：群众的智慧

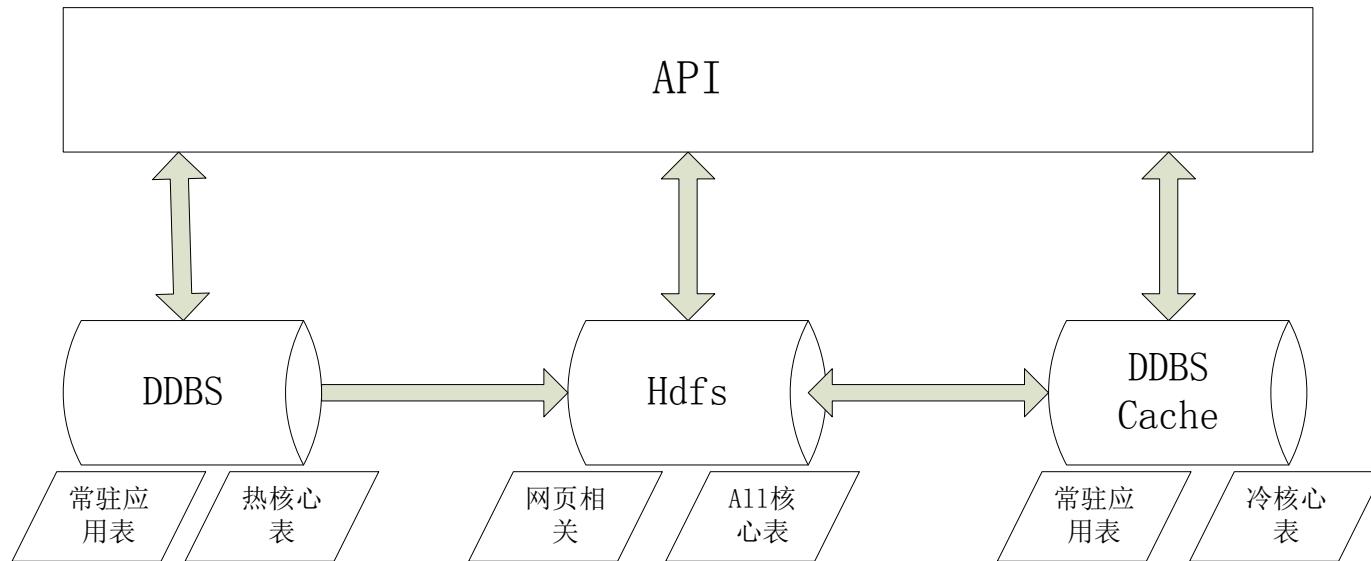
机器时代：是骡子是马

总结

附录



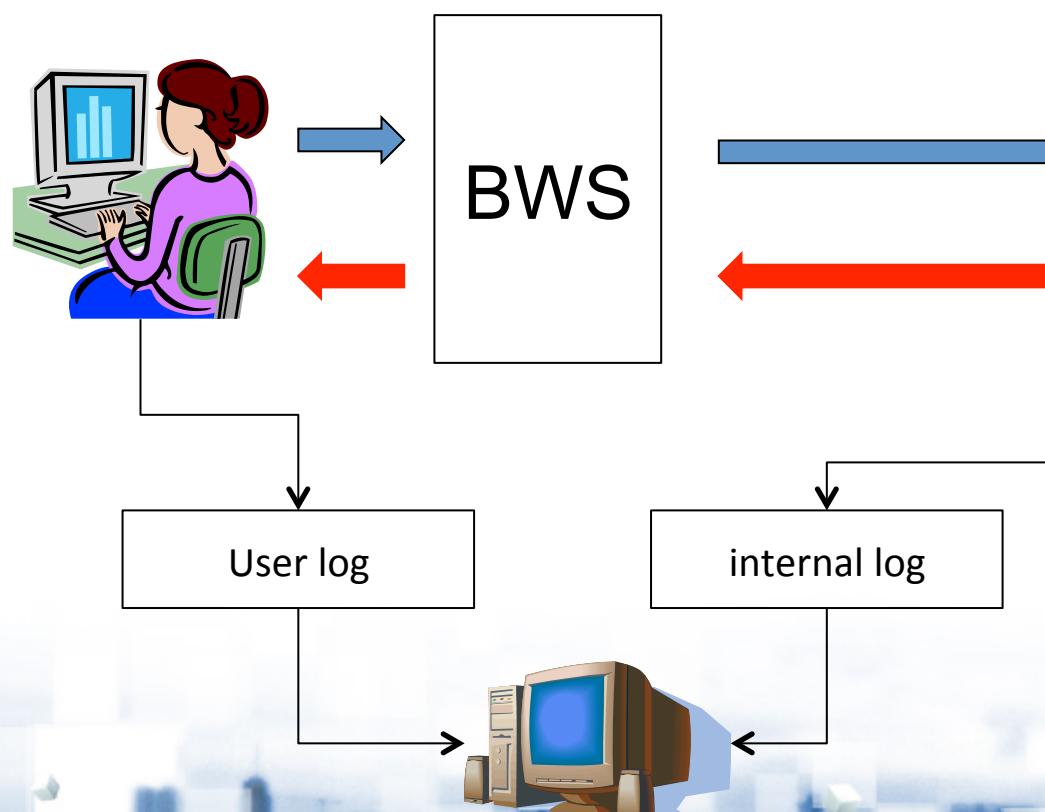
DYDC



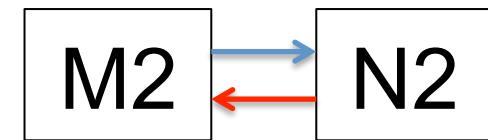
数据调度

## 附录2：抽样框架架构示意图

用户/Cookie  
切分流量  
 $Sid=1001$



变量映射，  
根据sid，采用x或者x'



BWS 下游模块

# 百度技术沙龙



关注我们：[t.baidu-tech.com](http://t.baidu-tech.com)

资料下载和详细介绍：[infoq.com/cn/zones/baidu-salon](http://infoq.com/cn/zones/baidu-salon)

“畅想•交流•争鸣•聚会”是百度技术沙龙的宗旨。百度技术沙龙是由百度与InfoQ中文站定期组织的线下技术交流活动。目的是让中高端技术人员有一个相对自由的思想交流和交友沟通的平台。主要分讲师分享和OpenSpace两个关键环节，每期只关注一个焦点话题。

讲师分享和现场Q&A让大家了解百度和其他知名网站技术支持的先进实践经验，OpenSpace环节是百度技术沙龙主题的升华和展开，提供一个自由交流的平台。针对当期主题，参与者人人都可以发起话题，展开讨论。

## InfoQ 策划·组织·实施

关注我们：[weibo.com/infoqchina](http://weibo.com/infoqchina)