

数字时代的学习： 信息搜索



姚 远

yaoyuan@shu.edu.cn

上海大学机自学院

2022/2/17

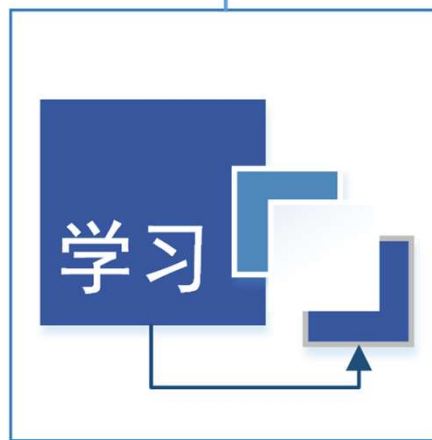
阶段概述



外部环境



思维方式



大纲



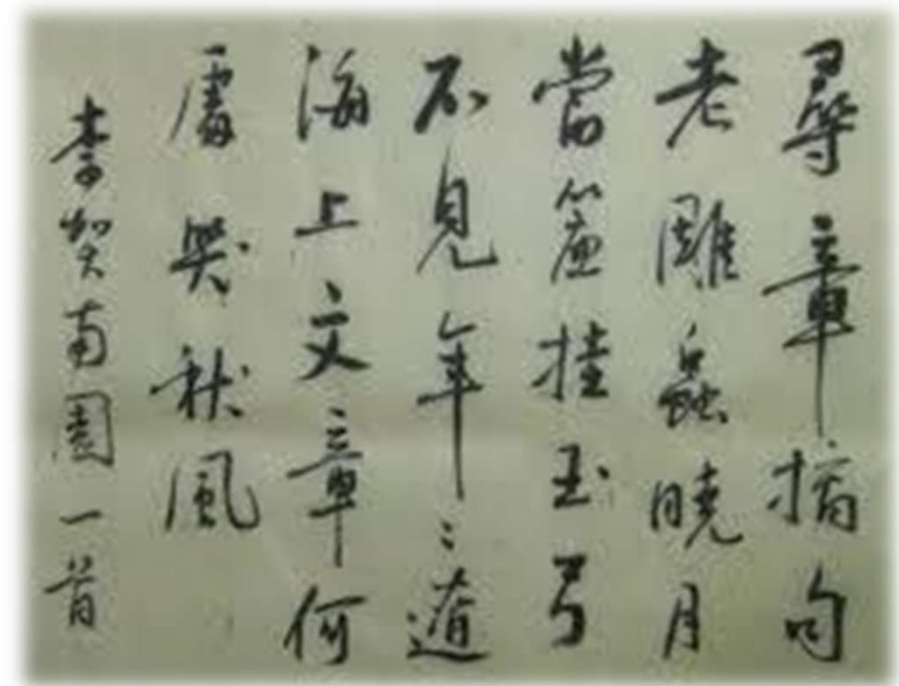
- 1. 搜索能力
- 2. 搜索引擎
- 3. 搜索方法

1. 搜索能力



- 搜索是一种技能

- 学习的重要途径



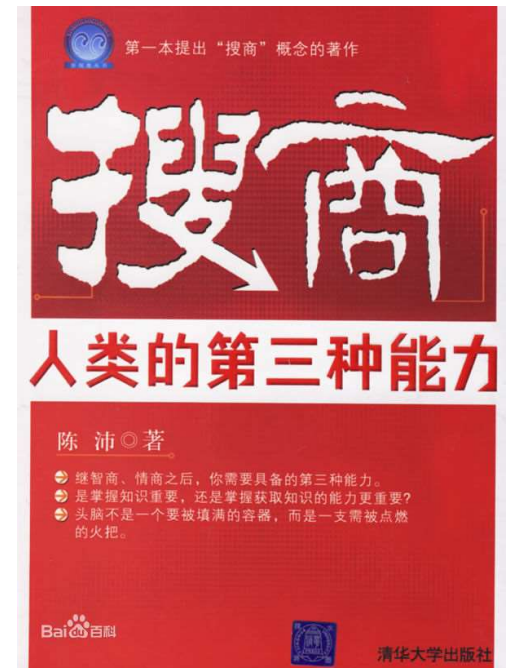
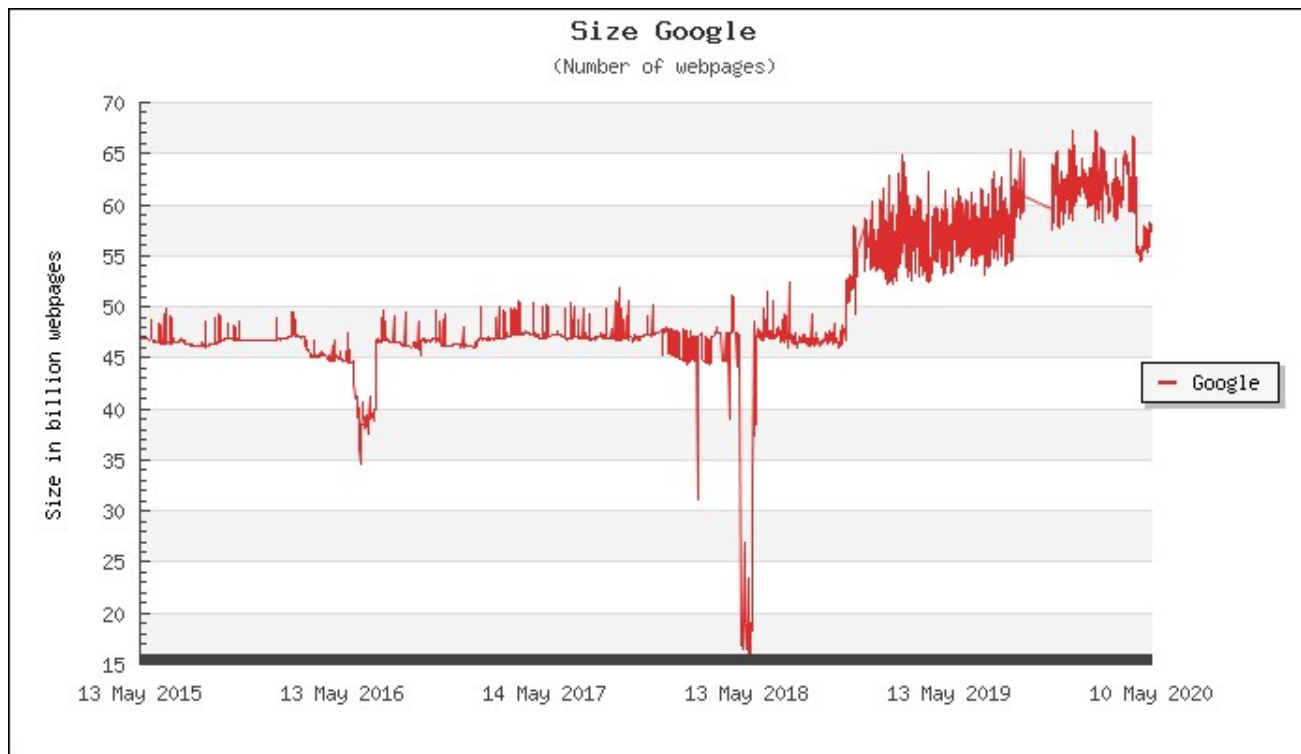
寻章摘句老雕虫，晓月当帘挂玉弓。
不见年年辽海上，文章何处哭秋风

--李贺

1. 搜索能力

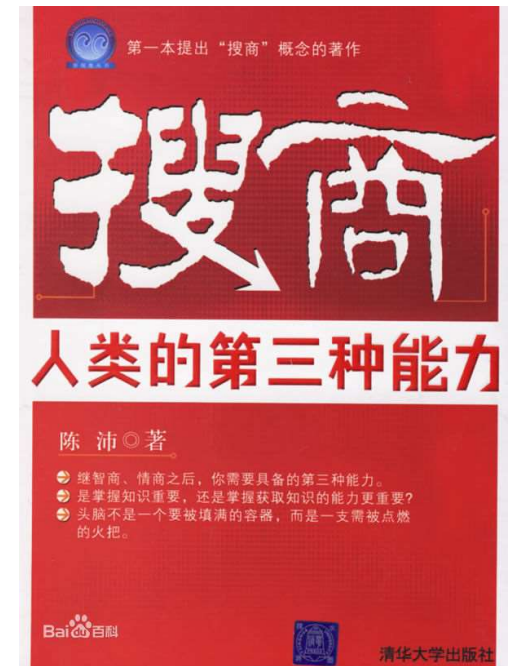
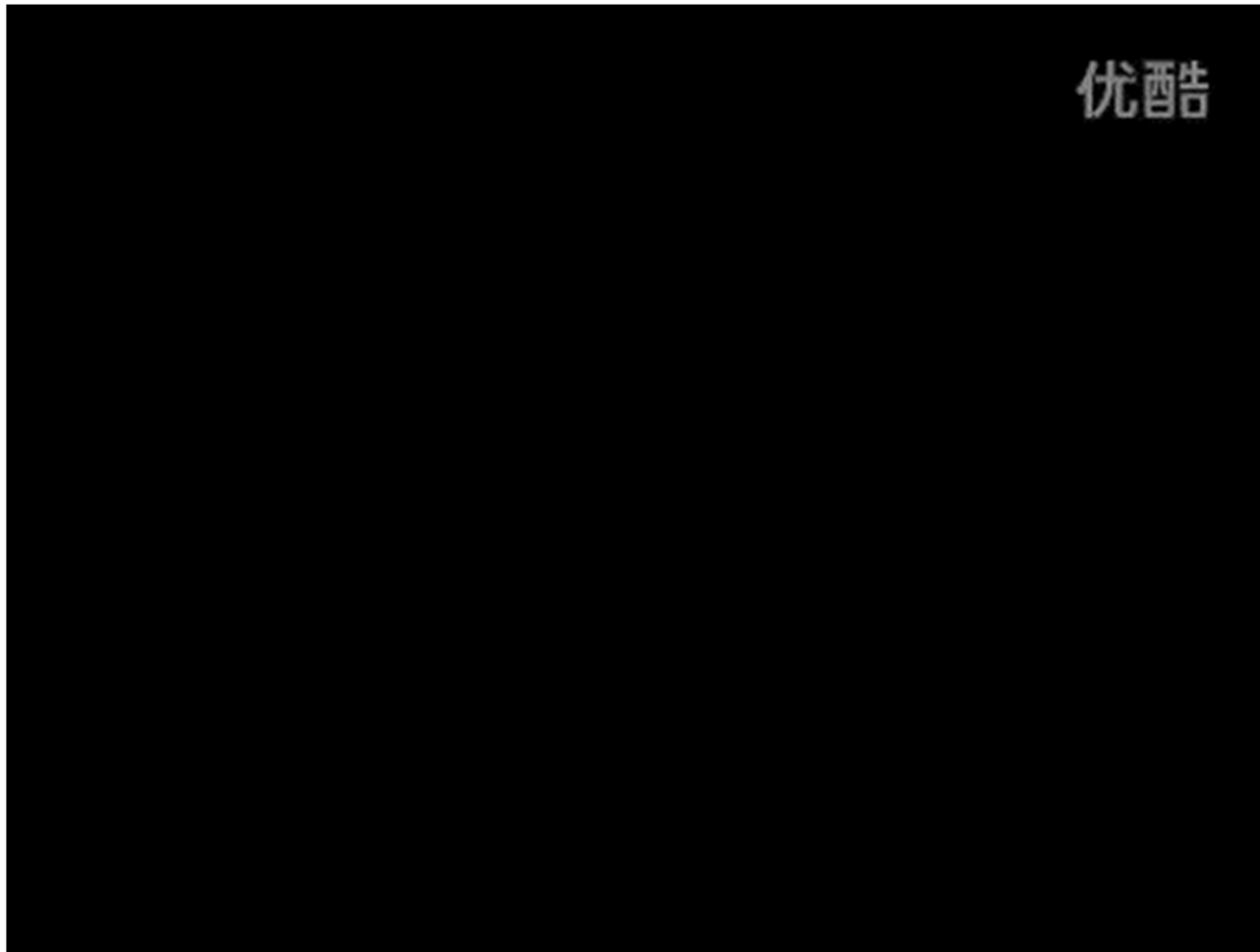
- 重要性与日俱增

- 中国网页数量超过2978亿个 (CNNIC 2020.4)
- Google索引的网页个数 (600亿, 2020.5)



1. 搜索能力

- 搜商



1. 搜索能力

- 搜索测试1.1



- 请给出图片所示地点的名字
- 何时命名?

1. 搜索能力

- 搜索测试1.2



- 请找出照片出处?
- 请给出相关人物背景

大纲



- 搜索能力
- 搜索引擎
- 搜索方法

2. 搜索引擎



- 文献搜索引擎

- 专业 (文献、专利、公司信息)
- 收费

- 通用搜索引擎

- 多功能
- 免费

2. 搜索引擎

- 搜索测试2.1

《原料颗粒对磷酸钙骨水泥水化硬化过程的影响》

- 这篇文章的作者信息
- 发表时间
- 被引用数量
- 引用这篇文章的文献列表



2. 搜索引擎

- 搜索测试2.2

《文章抄袭的判定》



- 这篇文章是否涉及抄袭
- 如果有，抄袭了哪篇文章的内容？

2. 搜索引擎



- 知识来源

- 图书
- 会议论文
- 期刊
- 科技报告
- 学位论文
- 科技档案
- 产品资料
- 政府出版物
- 标准文献
- 专利

2. 搜索引擎



- 我们的信息来源渠道

- 网络
- 电视
- 报纸
- 课堂
- 图书馆
- 会议...



▪获得信息的渠道往往是间接的，如何获取有效的知识？

2. 搜索引擎



- 文献搜索

■ 专业文献库

- 万方、维普、知网、IEEE

■ 免费文献库

- <https://arxiv.org/> 预印本

- <http://zh.booksc.org/> BOOKSC

...

■ 文献引擎

- <http://trend.wanfangdata.com.cn/>

- <http://scholar.google.com/>

- <https://www.semanticscholar.org/>

上海大学图书馆
Shanghai University Library

首页 资源 服务 关于本馆 我的图书馆

资源

电子资源

- > 数据库导航
- > 免费资源
- > 试用数据库
- > 电子期刊导航
- > 电子图书
- > 多媒体资源
- > 校外访问
- > 版权公告

纸本资源

- > 新书通告
- > 特色文献
- > 书刊捐赠
- > 书刊荐购
- > 馆藏报刊目录

国内外图书馆

- > 985高校馆

首页 » 电子资源

电子资源

- » 数据库导航
- » 试用数据库
- » 电子图书
- » 校外访问

2. 搜索引擎



- 文献搜索引擎

- 专业 (文献、专利、公司信息)
- 收费

- 网络搜索引擎

- 多功能
- 免费

2. 搜索引擎

- 网络搜索引擎-模式的变迁



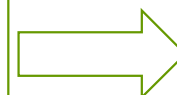
目录式



检索式



多模态



知识搜索

社交搜索

企业搜索

知识图谱



YAHOO!

LYCOS

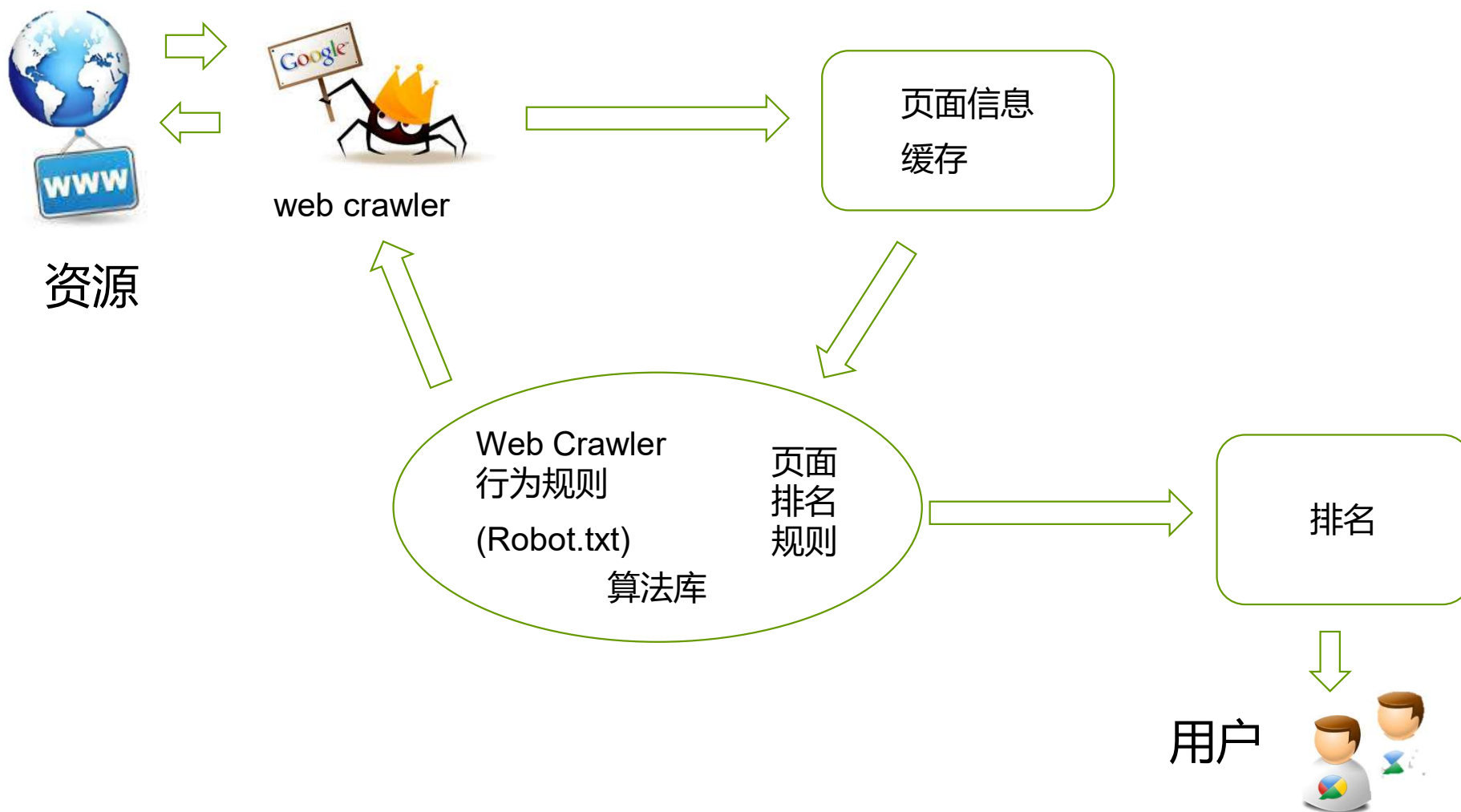


WolframAlpha

2. 搜索引擎



- 网络搜索引擎工作模式



2. 搜索引擎



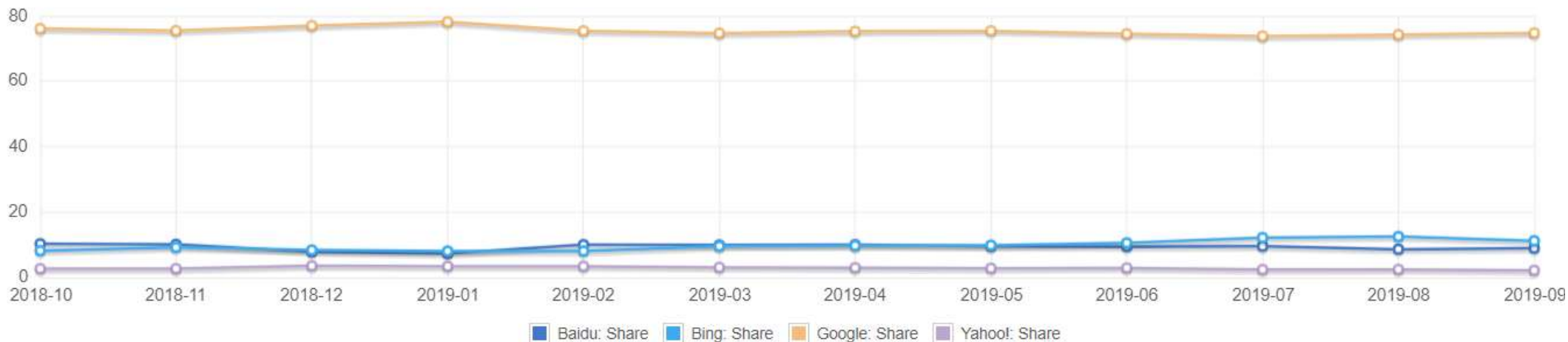
- 搜索引擎的比较

- 收录范围
- 排名算法
- 检索算法
- 交互方式

国内搜索引擎市场数据

搜索引擎	百度	搜狗	神马	360	谷歌	必应
占有率	76.42%	11.35%	4.71%	2.85%	2.4%	2.06%
统计年度	2018年7月1日-2019年7月31日					
发布日期	2019年8月1日					

PC端全球搜索引擎市场数据



2. 搜索引擎

- 搜索测试3



是谁发明了搜索引擎?

有哪些好用的搜索引擎?

大纲



- 搜索能力
- 搜索引擎
- 搜索方法

3. 搜索方法

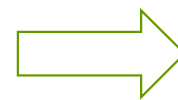


- 对搜索目标应具有明确的认知

**明确
任务**



**选择搜
索引擎**



**确定搜
索方案**



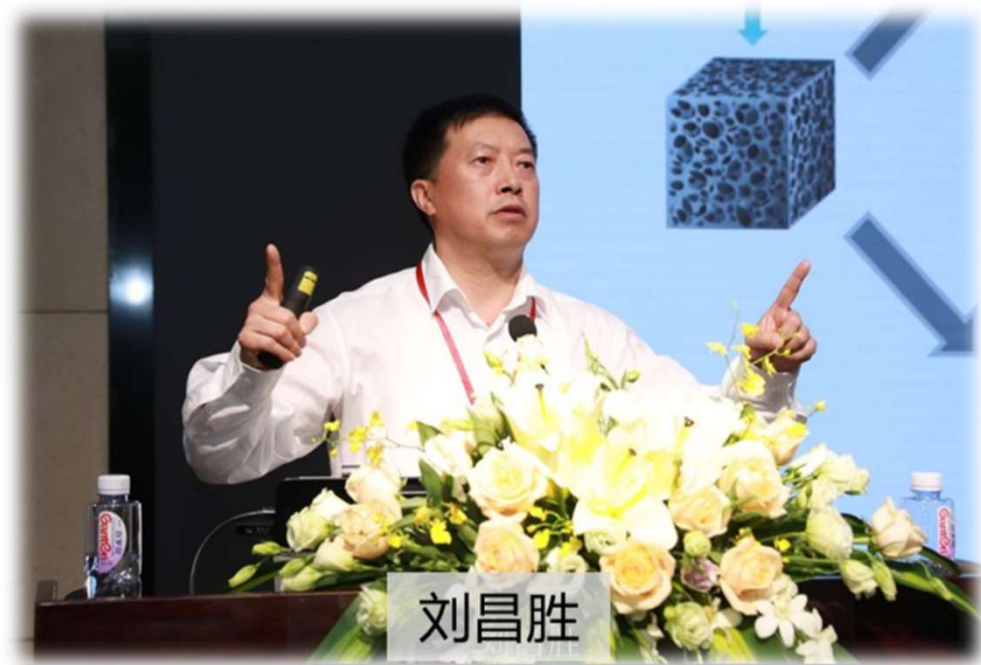
过滤结果

搜索流程

3. 搜索方法

- 搜索测试4

请给列出10个上海大学校长**最近一个月**内参加的**活动**。



3. 搜索方法



1) 明确任务

▪任务类型

- 地址
- 图像
- 文本
- 视频
- 音频
- 数据
- 计算

- 论文
- 报告
- 专利
- 书籍
- 人
- 企业
- 定理

3. 搜索方法



2) 选择搜索引擎

- GOOGLE(scholar, trend...)
- 百度 (快速)
- 360 (好搜)
- Sogou (微信/知乎)
- Bing
- Magi
- Duckduckgo

3. 搜索方法



- 搜索测试5

1) 请给出**上海大学**、**复旦大学**、**上海交大**和**华东理工**四个学校的关注度，以及随时间变化的趋势

2) 请给公众对**新冠病毒**、**非洲猪瘟**这二个名词的关注情况，以及随时间变化的趋势

3. 搜索方法

- 技巧

- 关键字搜索
- 逻辑关系搜索
- 命令使用与拓展



3. 搜索方法



- 关键字搜索

- 归纳关键词，逐步精炼
- “双引号”限定
- “*” eg. “largest * in the world”

- 逻辑关系搜索

- 命令使用与拓展

3. 搜索方法



- 关键字搜索
- 逻辑关系搜索
 - AND
 - OR
 - -
 - +
- 命令使用与拓展

3. 搜索方法



- 关键字搜索
- 逻辑关系搜索
- 命令使用与拓展
 - 指定目标位置: Allinanchor, allintext, allintitle, allinurl, link, insubject ...
 - 指定搜索类型 define, filetype, info, time, weather
 - 指定搜索区域 site, [cache]
 - 相关性搜索 related
 - 计算性搜索 单位转换, 统计数据, 数据计算

3. 搜索方法



- 数据集搜索

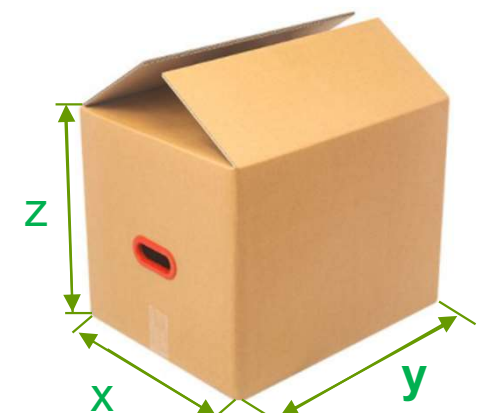
- [Google Dataset Search](#)

- 自然语言命令

- time in CN now
- weather Shanghai
- population in China

- 计算

- 1 美元 = ? 人民币 $\cos(x+y)$
- Minimize $2(x+y)*z+x*y$ Subject to $x*y*z = 50$



3. 搜索方法



- 搜索测试6

如何成为一名“宽客”？以此为目标，借助搜索引擎制定一份学习计划。

3. 搜索方法

- 知识搜索

<http://www.wolframalpha.com/>

- 一种“计算知识引擎”
- 自动搜集与人工干预
- 架设在超级电脑R Smarr上（全球排名66）

- 涵盖领域

- Dates & Times (日期和时间)
- Places & Geography (地方与地理)
- Socioeconomic Data (社会数据)
- Health & Medicine (健康与药物)
- Food & Nutrition (食物与营养)
- Words & Linguistics (文字与语言学)
- Astronomy (天文学)
- Earth Sciences (地球科学)
- Life Sciences (生命科学)
- Units & Measures (单位与测量)
- Money & Finance (货币与财政)
- Culture & Media (文化与媒体)
- Mathematics (数学)
- Physics (物理学)
- Chemistry (化学)
- Materials (材料)
- Engineering (工程学)
-



Stephen Wolfram

总结



- 不同形式的数据具有不同的特征
- 准确表述搜索问题
- 掌握搜索的流程和技巧
- 综合利用各种搜索引擎