

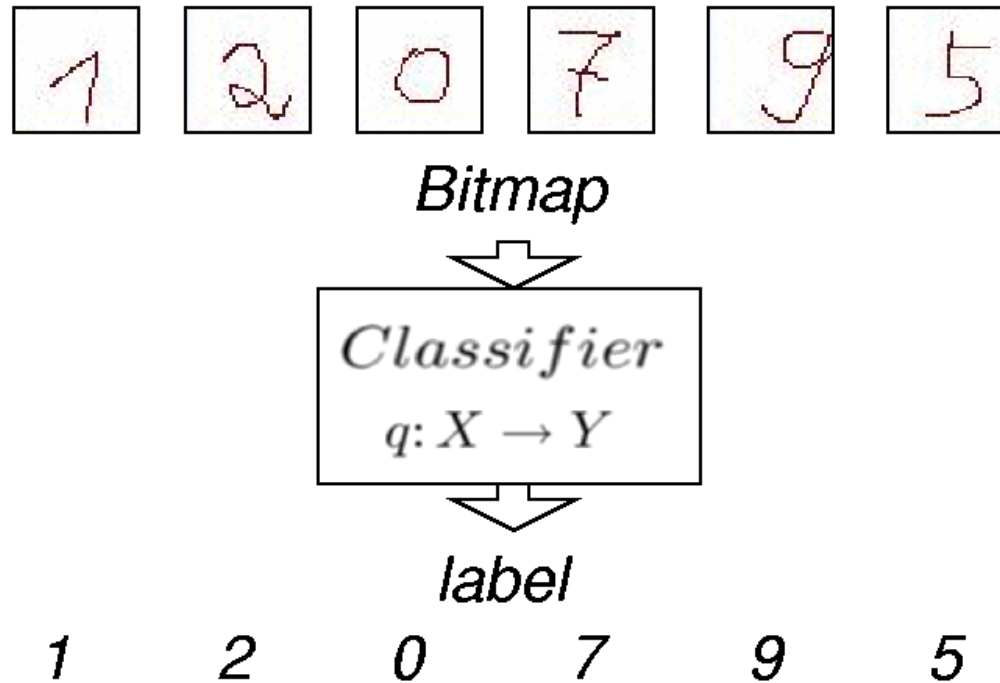


Nine Real Hard Problems We'd Like You to Solve

**Robin Li
August, 2012**

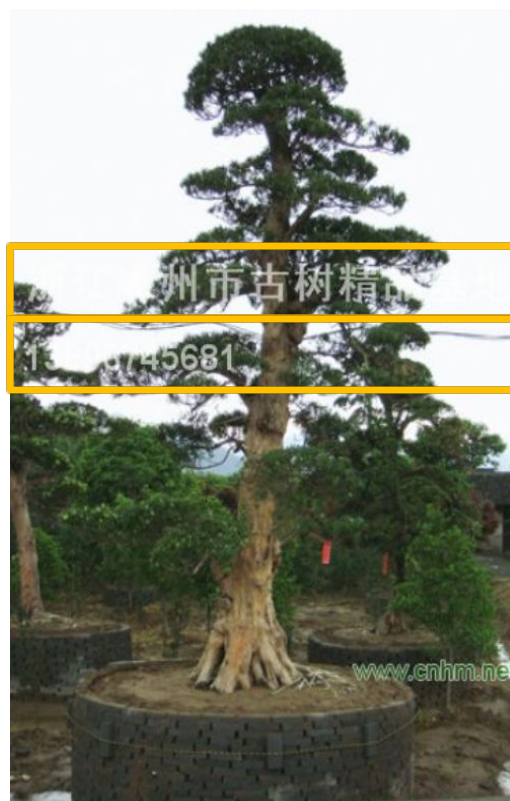
The OCR problem

- And this



The OCR problem

- But now we deal with this



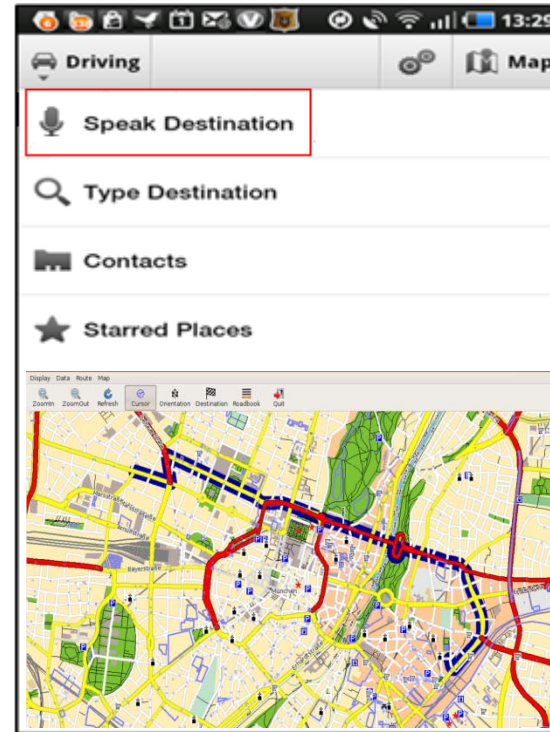
The OCR problem

- And this a lot more



2. The speech recognition problem

- **Traditional Applications: Limited vocabulary**
 - Voice dialing
 - Speech navigation



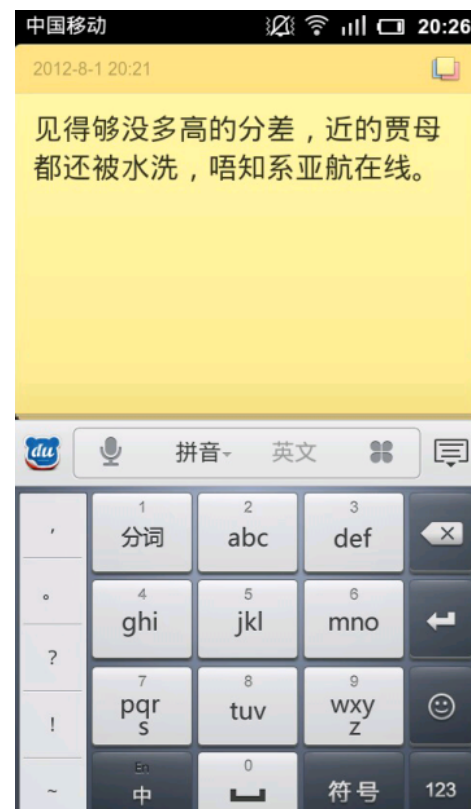
The speech recognition problem

- Descriptive information
 - T like in “Tom”, 弓长 “张”
- Adaptation - dialects/accents/pitch detection
- Background noise
- Dialogue management
- Mixed language input

Speech – Challenges: Examples



Standard Mandarin in Quiet Environment



Dialect and mixed language in noisy environment

3. Content based image retrieval

- Image search page views surpassed web search last year
- Qzone daily upload 320 million pictures
 - In less than a year, we have to index more pictures than web pages!

CBIR – But there is hope

- Geo information
- Inter-app information
- Semantic understanding of images
 - Object recognition
 - Scene understanding
 - Face recognition

4. Anti-Spam in Social

- Goal
 - Healthy ecosystem
- Problems
 - Content: false/indecent information
 - Means: Images / Martian language



臂『Lj9 ㄥ 座■Ta2 と 駱〇Jo5 ㄥ
【筑姪饹嘍】吧主下台【筑姪饹嘍】
更多款式可以咨询扣扣②④③⑧①③⑧③①~
/删/差评/方/法/ㄥ〇※№♀7691417



ぢ@ㄥ ㄥ {【我们】 删【我们】

【我们】 淘【我们】

【我们】 宝【我们】

【我们】 差【我们】

【我们】 评【我们】

【我们】 方【我们】

【我们】 法【我们】

【发任意邮件到 cnpyjh@qq.com 自动回复】

Anti-Spam in Social

- Challenges
 - Motivation understanding
 - Real-time detection
 - Handling mechanism

5. Box computing – Open Data Platform

- The dark web
- The birth of the Open Data Platform

 [新闻](#) [网页](#) [贴吧](#) [知道](#) [MP3](#) [图片](#) [视频](#) [地图](#) [文库](#) [更多»](#)

菲尔普斯

百度一下

[伦敦奥运会运动员菲尔普斯](#) [百度百科](#)



姓名：迈克尔·菲尔普斯

国籍：美国 **运动项目：**游泳

生日：1985年6月30日

身高：193cm

[个人简介](#) [个性资料](#) [个人履历](#) [主要成绩](#)

菲尔普斯奥运会赛程

结束	08-02 03:43	游泳 男子200米个人混合泳半决赛1	已结束	视频回放	战报
结束	08-02 18:27	游泳 男子100米蝶泳-预赛6	已结束	无视频回放	战报
结束	08-03 03:19	游泳 男子200米个人混合泳决赛	已结束	视频回放	战报
结束	08-03 03:56	游泳 男子100米蝶泳半决赛1	已结束	无视频回放	战报
未开	08-04 02:38	游泳 男子100米蝶泳决赛	无电视转播	无视频直播	无图文直播

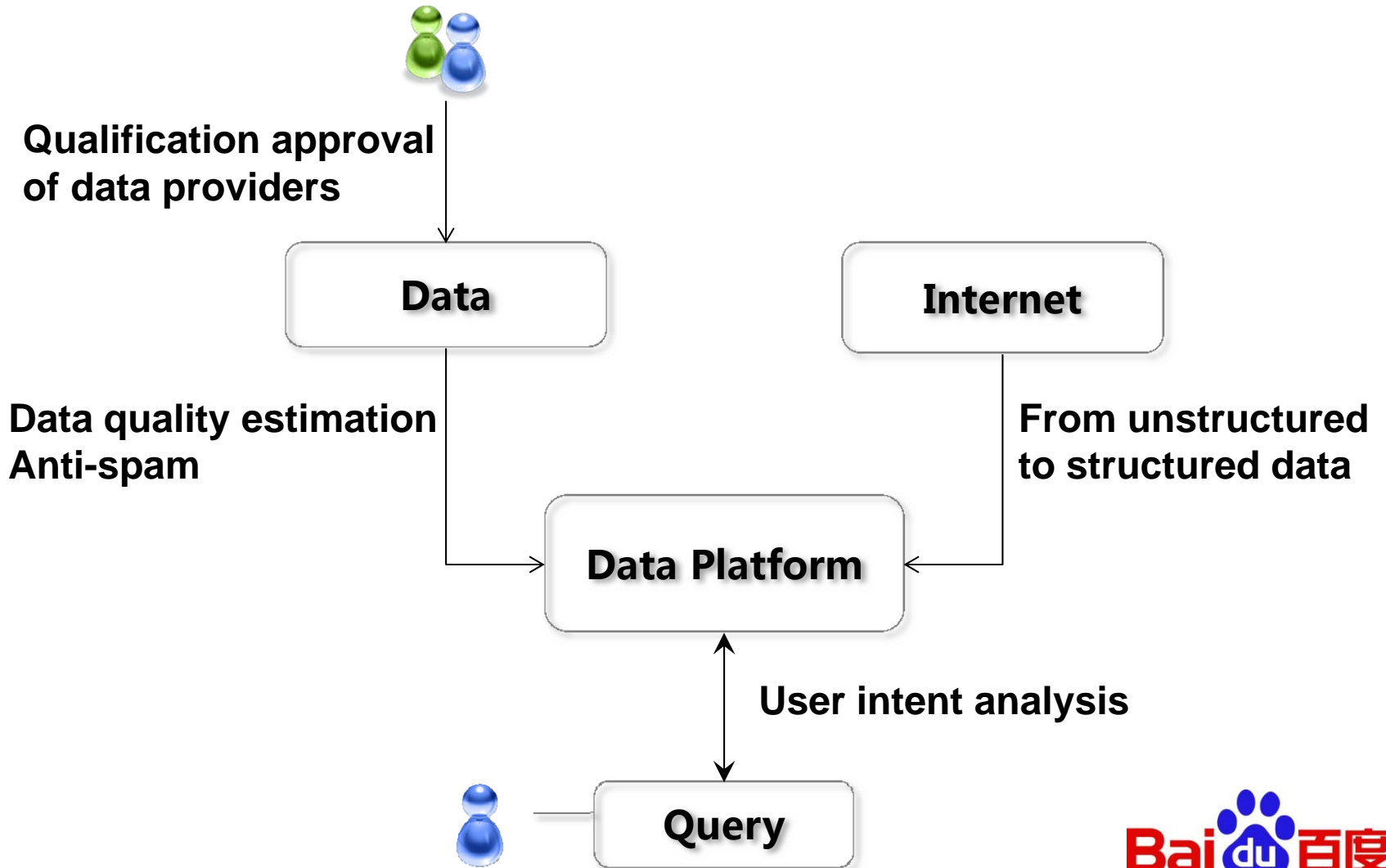
[完整赛程查询](#) [中国队赛程查询](#)

以上信息由中国奥运会合作伙伴新浪提供，比赛时间为北京时间

2012.sina.com.cn - 



Open Data Platform



Open Data Platform – How far can we go?



Open Application Platform

- System security and stability
- Inter-application resource scheduling on cloud
- Application resource distribution between cloud and clients
- User intent analysis
 - Connect users with applications

7. Box computing – Open Publishing Platform

 [新闻](#) [网页](#) [贴吧](#) [知道](#) [MP3](#) [图片](#) [视频](#) [地图](#) [文库](#) [更多»](#)

今天作业好多，累死了

百度一下

推荐：[把百度添加到桌面](#)

↑ 百度应用 我的应用 << 更多 >> 应用推荐

今天作业好多，累死了

表情 视频 图片 130

腾讯微博

编辑 选择微博 一键发布

收藏 分享 新窗口

应用简介 评论 1

微博在线群发★★★★★ 百度应用

百度搜索框发微博应用是百度官方提供的，一个方便实用的群发微博的应用。可直接使用百度搜索框发布微博内...[显示详细>>](#)

[意见反馈](#) 使用中有问题，欢迎反馈!

 **Baidu** 百度

Open Publishing Platform

- User publishing intent analysis
 - Need to publish?
- Audience and publishing site
 - Categorization of the content
 - Appropriate publishing site
 - Form of publication

8. Box computing – Personalized Homepage



[新闻](#) [网页](#) [贴吧](#) [知道](#) [MP3](#) [图片](#) [视频](#) [地图](#) [百科](#) [文库](#) [更多>>](#)

2012伦敦奥运

我的奥运

[直播](#) | [预告](#)

头条 科比：队友没什么值得我学的 我是队中最强内线

直播 8-10 12:00 奥运午间报道 [直播](#)

结束 8-10 03:30 水球 女子决赛 **美国胜** [重播](#)

奖牌榜

金 银 铜

美国 第1名 39 25 26

中国 第2名 37 24 19

[百度奥运](#) [奥运贴吧](#) [新浪奥运](#)

[查看更多](#)

我的网址

[+添加](#)

[百度贴吧](#)

[百度知道](#)

[百度空间](#)

[百度新闻](#)

[新浪微博](#)

[步步惊心吧](#)

[林志玲吧](#)

[携程网](#)

[百度公司官网](#)

[段林希](#)

[美图拍拍](#)

[菜谱大全](#)

[豆瓣FM](#)

[弹弹堂](#)

[PPTV直播](#)

[黄金矿工](#)

[+](#)

实时热点

刘翔 手术 结束

博尔特 金牌

陈若琳 金牌

王娇 无缘 卫冕

李宗佑迷奸艺人

宁波摩天轮吹垮

国酒争议

光明顶12级大风

喂鸡神曲

男兵团体夺冠

新鲜事



Personalized Homepage

- Home page, no home page?
- Search without a query
- Personalized recommendation
 - User status
 - Temporal-spatial information
 - Scenarios
 - Tasks
- Integration of heterogeneous contents
- Personalization for mobile

Evolution of Box Computing

Transition of Web Content, User Req. and Behavior

Interface
to Web
Contents

Web Information
Platform

Web Application
Platform

Open and
Personalized
Platform

Open Data Platform

Open App Platform

Personalized Homepage
Open Publishing Platform

(Direct Info Plug-in)

(Instant App Plug-in)

("You" plug-in)

Traditional
Search
Engine

Box Computing

Meet user's needs in most simple and
reliable way

9. Monetization for social

- Ads
 - Reach, relevancy, engagement, social context
- Commissions on a platform
- Subscription

Baidu 贴吧 新闻 网页 贴吧 知道 MP3 图片 视频 百科 地图

奥运会 进入贴吧

奥运会吧 精品 图册 投票

体育相关话题目录排名: 本吧昨日第 1 名, 巅峰之上, 一览众山小! | 周排名: 1 | 月排名: 1

LONDON 2012 今日看点

08-07 17:10	田径	男子110米栏预赛	刘翔	110米栏	田径
08-07 21:00	体操	男子双杠决赛	张成龙	体操	
08-07 21:47	体操	女子平衡木决赛	眭禄	邓琳琳	体操
08-07 22:00	花样游泳	双人自由自选决赛	花样游泳		
08-07 22:30	乒乓球	女团金牌赛	乒乓球	李晓霞	郭跃

3141 【冰力十足 一冰最棒 千万贴吧友送鼓励】 直播 LONDON 2012

127 奥运第11日, 中美冲金点对决
中国今日有一定把握的项目是男子单人三米板和乒乓球女子团体, 其它如男子双杠、女子平衡木、男子单杠要奋力一搏了。

998 奥运全民主题曲发布, “歌神”张学友与朗朗联手演绎!
7月12日, 历时三个月, 由近两亿个“中国节拍”汇集而成的「可口可乐」2012伦敦奥运中国加油主题曲——《中国节拍》震撼世界! 在北京首发。现场, 朗朗用钢琴演奏, 歌神张学友倾情演唱了主题曲, 有来自56个民族和不同领域的普通人演绎了自己的节拍。为了给奥运健儿加油, 可口可乐公司发起了“加入中国节拍, 助威2012伦敦奥运”活动, 征集到近2亿个“中国节拍”。如果你想了解或加入“中国节拍”请猛戳这里: <http://www.icoke.cn> 您也可以观看下方视频了解首发仪式盛况:

423 陈一冰, 是冠军! 是中国人就陈一冰, 是冠军! 是中国人就用这

718 为一个瘦小的外国女人盖楼, 看到有无知的童鞋揶揄她的长相, 我只要有一点良知和人性, 都不会口出恶言

327 这是13亿人的比赛! 刘翔孙杨何姿领衔, 运动员集体致谢 顶 可口可乐 卡卡丹儿

80 阿尔及利亚名将800米消极比赛被剥夺参赛资格 跑步怎么个消极法???

1 2 3 4 5 6 7

列表页

奥运会吧 精品 投票

1 2 3 4 5 6 7 8 9 10 下一页 尾页 有614篇回帖 返回奥运会吧>

奥运全民主题曲发布, “歌神”张学友与朗朗联手演绎!

只看楼主 收藏 回复

7月12日, 历时三个月, 由近两亿个节拍汇集而成的「可口可乐」2012伦敦奥运中国加油主题曲——《中国节拍》震撼世界! 在北京首发。现场, 朗朗用钢琴演奏, 歌神张学友倾情演唱了主题曲, 有来自56个民族和不同领域的普通人演绎了自己的节拍。为了给奥运健儿加油, 可口可乐公司发起了“加入中国节拍, 助威2012伦敦奥运”活动, 征集到近2亿个“中国节拍”。如果你想了解或加入“中国节拍”请猛戳这里: <http://www.icoke.cn> 您也可以观看下方视频了解首发仪式盛况:

采集到花瓣



592

真是全民一起为奥运健儿加油啊~中国必胜!

可口可乐

详情页

我的本吧头衔

喜欢 我喜欢

喜欢本吧 马上点高头衔

我喜欢的吧 更多>>

仓央嘉措 暗黑3
李毅 百度相册
百度推广 百度网盘

奥运会奖牌榜

国家	金	银	铜	总
1 中国	31	19	14	64
2 美国	29	15	19	63
3 英国	18	11	11	40
4 韩国	11	5	6	22
5 法国	8	9	9	26

奥运会赛程表

08-06 08-07 08-08

02:00 举重男子105kgA组
02:00 田径女子撑杆跳高决赛
02:00 跳水男子单人3米跳...
02:56 摔跤男子古典式120...
05:15 篮球男子A组预赛

相关图片推荐

有关张学友的图片 中国节拍一起来的可口可...



And the 10th Problem is...

Robin Li
August, 2012