



互联网图片的挖掘与应用

牛正雨 互联网数据研发部

报告目录

- 图片挖掘概述
- 挖掘技术详解
- 图片挖掘展望

概述

- 为什么做图片挖掘
 - 外因：互联网已进入“读图时代”
信息载体：网站内容、用户交流
 - 内因：产品变革带来的需求
交互方式：输入、展现

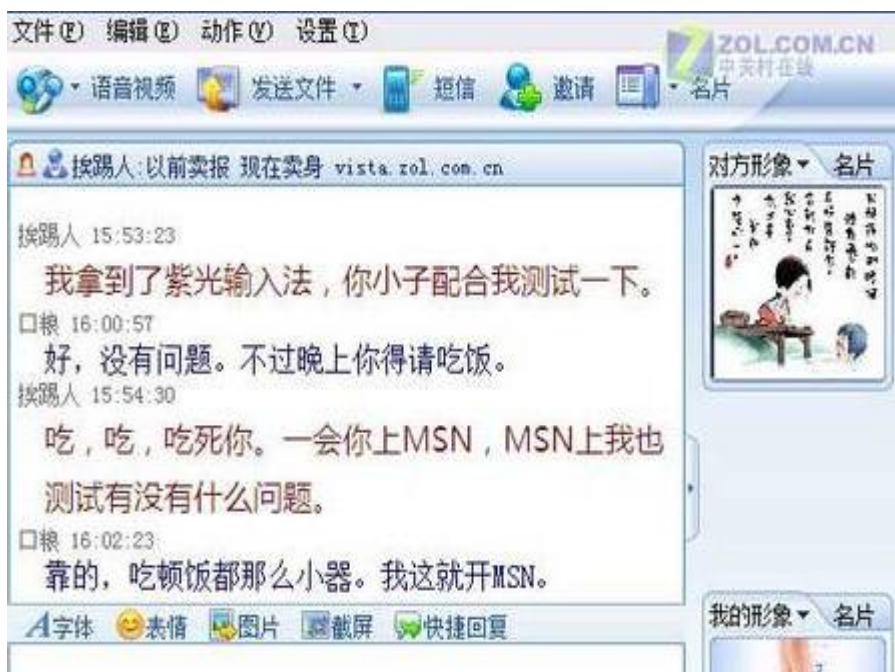
概述-读图时代

- 互联网已进入“读图时代”
— 网络信息载体：从文字到图片



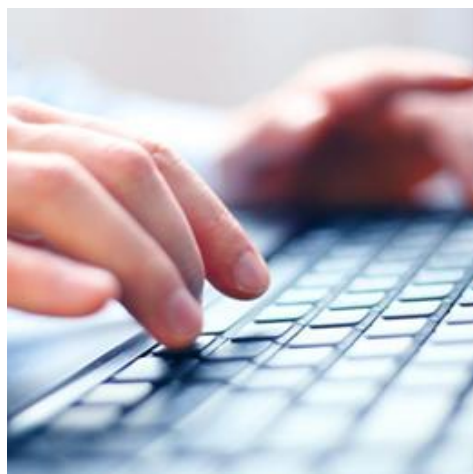
概述-读图时代

- 互联网已进入“读图时代”
— 用户交流媒介：从文字到图片



概述-产品需求

- 产品对图片的需求：
 - 一搜索query：从文字到图片(移动端、信息表达)



概述-产品需求

● 产品对图片的需求： 一搜索结果展现：从文字到图片

无敵汽车网_汽车_改装_高尔夫7

search.modiauto.com.cn/?keyword=高尔夫7

421篇文章 ... Caracere近日便发布了自家针对高尔夫7推出的改装套件。其中包括了外观和动力 两方面, Caracere针对2.0升TDI柴油两驱车型提供了208匹马力、41.8 ...

依然是标杆车型土豆汽车试驾一汽-大众高尔夫7_在线视频观看_土豆...

www.tudou.com/programs/view/t7ebSy48nww/

高尔夫(Golf)是一项运动,同时高尔夫(Golf)还是一种文化。它所代表的是一种生活,是对待自己和家人的态度,是属于年轻人的精神,是一辆车。历经六代的演变, ...

高尔夫7荣获2014最佳国产中级轿车奖-南海拓华

www.tuohuaauto.com.cn/official/201403/64543.html

高尔夫7荣获2014最佳国产中级轿车奖, 2014-03-21 12:33类型: 官方新闻, 2月19日, 目前唯一由汽车消费专业媒体《ams车评?汽车博览》主办的市场调查活动——“ ...

好消息! 一汽大众高尔夫7留学生免税车价格已发布! - 最新动态

www.68011866.com/newsdetail.php?369.html

2014年3月17日 ... 作为大众汽车史上最畅销的车型, 高尔夫品牌早已成为大众旗下的传奇车型, 历经39年历史发展, 高尔夫已经完成七代车型的演变。自2013年底高7 ...

小保养360元起高尔夫7用车成本解析_TOM汽车-Tom汽车频道

auto.tom.com/2014-03-19/0D5K/00095891.html

2014年3月19日 ... 小保养360元起高尔夫7用车成本解析, 收藏(-人收藏1人收藏) 复制链接 打印来源: TOM汽车 作者: xgo.【共条评论共有0条评论】【我要评论】 ...

高尔夫7并不是两厢版速腾两车差异化对比_汽车_环球网

auto.huanqiu.com/testdrive/2014-01/4737962.html

2014年1月10日 ... 2013年12月上市的高尔夫7可以算得上是国内的一款重磅新车了, 不过众所周知, 高尔夫7并没有延续上代高尔夫6的PQ35平台, 而是换装了大众 ...

Searches related to 高尔夫7

高尔夫7报价

高尔夫7代

新高尔夫7

高尔夫7价格

高尔夫7 gti价格

高尔夫7 gti

高尔夫6

高尔夫7上市

大众高尔夫7

高尔夫7上市时间

【图】全新高尔夫7上市_专题_网易汽车



全新高尔夫7于2013年12月10日上市, 届时我们将为您带来全新高尔夫7上市市场的第一手资料, 包括全新高尔夫7报价, 全新高尔夫7论坛, 全新高尔夫7怎么样, 全新高尔夫7参数等信息

auto.163.com/special/... 2013-12-10

高尔夫7上市 这些钱买它们不好吗_新浪汽车_新浪网



【原创】作者: 孔蕊汐 - 发表时间: 2013年12月17日

高尔夫7上市 这些钱买它们不好吗, 高尔夫7上市, 高尔夫7, 新浪汽车讯 2013年12月10日, 被称为新一代大众神车的**高尔夫7**正式上市。 ...

auto.sina.com.cn/news... 2013-12-17

盲测 高尔夫7对比高尔夫6 感受见真招_凤凰网_凤凰网



2013年12月11日 - 高尔夫7内饰用料更考究 乘坐空间变化不大凤凰网车导购 大众旗下的高尔夫车型作为一款经典的两厢家用轿车, 从1974年推出至今已经历过6代车 ...

auto.ifeng.com/daogou... 2013-12-11

【高尔夫7】_高尔夫7报价及图片_高尔夫7怎么样_爱卡汽车

爱卡汽车为您提供**高尔夫7**的基本配置参数和买车参考, 包括**高尔夫7**图片、**高尔夫7**报价、**高尔夫7**参数配置、**高尔夫7**油耗数据, 以及已经购买**高尔夫7**的车友评价和各地价格导购行情...

www.xcar.com.cn/conte... 2014-03-27

【高尔夫-高尔夫报价_论坛_图片_价格】_一汽-大众高尔夫_一汽-大众...



新浪汽车一汽-大众高尔夫栏目为您提供最新**高尔夫**报价、参数配置、**高尔夫**降价优惠团购新闻、**高尔夫**油耗、**高尔夫**图片、**高尔夫**评测试车等信息。想了解**高尔夫**怎么样, 请关注...

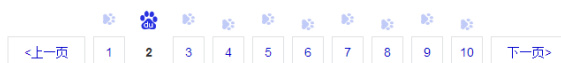
data.auto.sina.com.cn/13/ 2014-03-25

大众高尔夫7 GTI先进口后国产_北京车展见【图】_进口新车快讯...



2013年12月18日 - 盼了一年多的第七代**高尔夫**终于在2013年底上市了, 以此开启了**高尔夫7**众多衍生车型涌入的计划。目前有消息称, 第七代**高尔夫** GTI性能版将在2014年先以进口...

www.pcauto.com.cn/nat... 2013-12-18



相关搜索

高尔夫6

高尔夫

高尔夫gti

高尔夫7什么时候上市

福克斯

高尔夫7gti

高尔夫7

高尔夫7价格

大众高尔夫7

国产高尔夫7



报告目录

- 图片挖掘概述
- 挖掘技术详解
- 图片挖掘展望

详解

- 图片挖掘的目标：
 - 挖掘图片自身特征
 - 挖掘图片与其他数据的关联
- 图片数据的特点
 - 规模、形态、质量
- 图片挖掘技术
 - 图片标注技术
- 产品应用

详解-挖掘目标

- 挖掘图片自身特征：
 - 语义特征：图片的关键词标注等
 - 质量特征：低质图片识别，色情图识别等
- 挖掘图片与其他数据的关联：
 - 图片与图片：识别相同图片，套图识别
 - 图片与网页：识别网页中的代表性图片
 - 图片与实体：为实体义项配图

详解-挖掘目标

- 语义特征：图片的关键词标注等



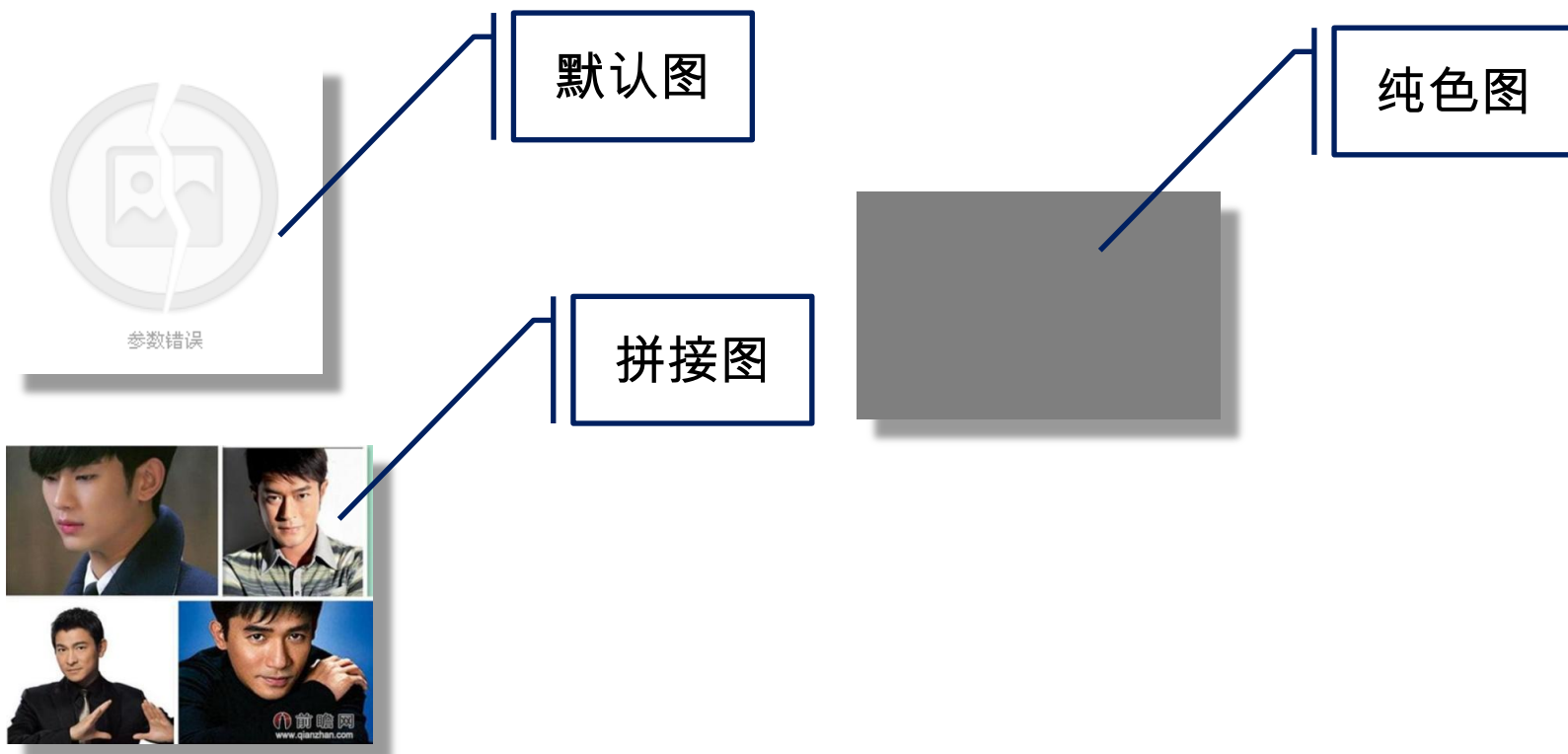
全智贤

荷兰屋/“猫眼”/太阳镜

爱马仕/棕色/斗篷

详解-挖掘目标

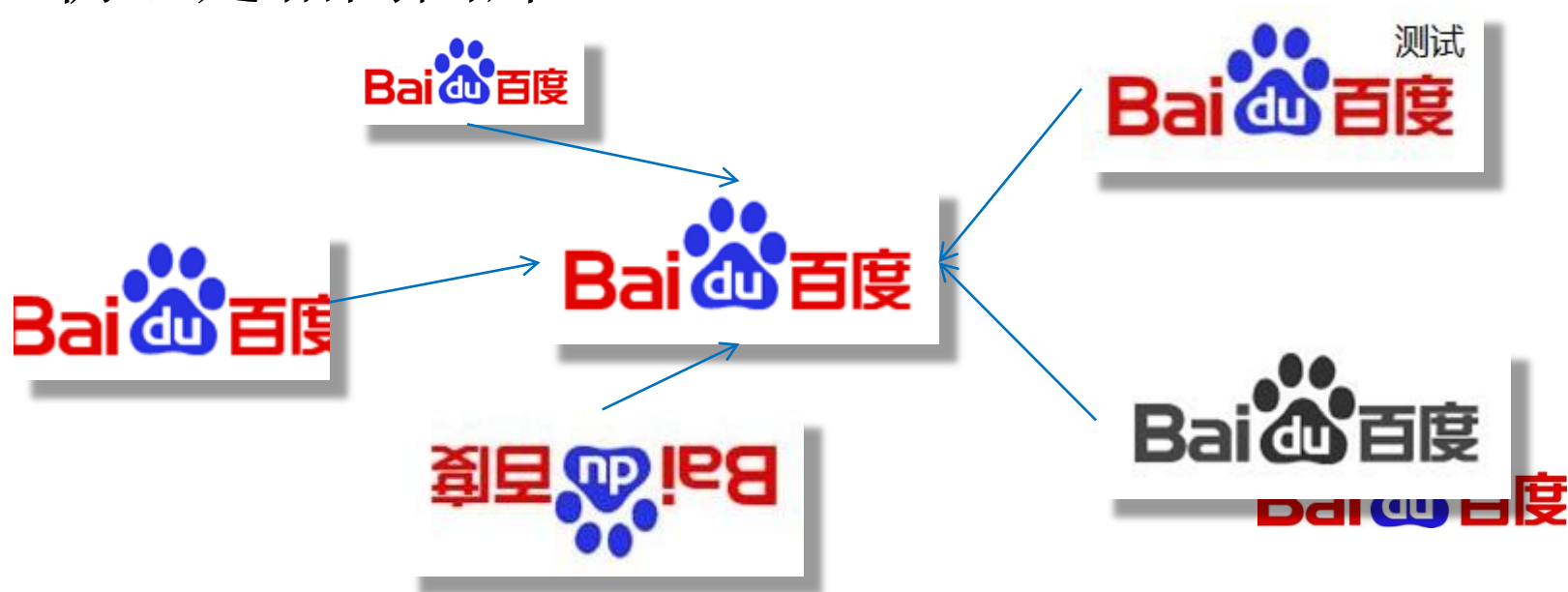
- 质量特征：



详解-挖掘目标

- 图片与图片的关联

— 相同图片识别：即一张图片经过少量编辑行为：如缩放、裁剪、旋转、亮度变化、模糊、颜色改变、加文字、加水印、涂改等，可以认为仍旧是相同图片。

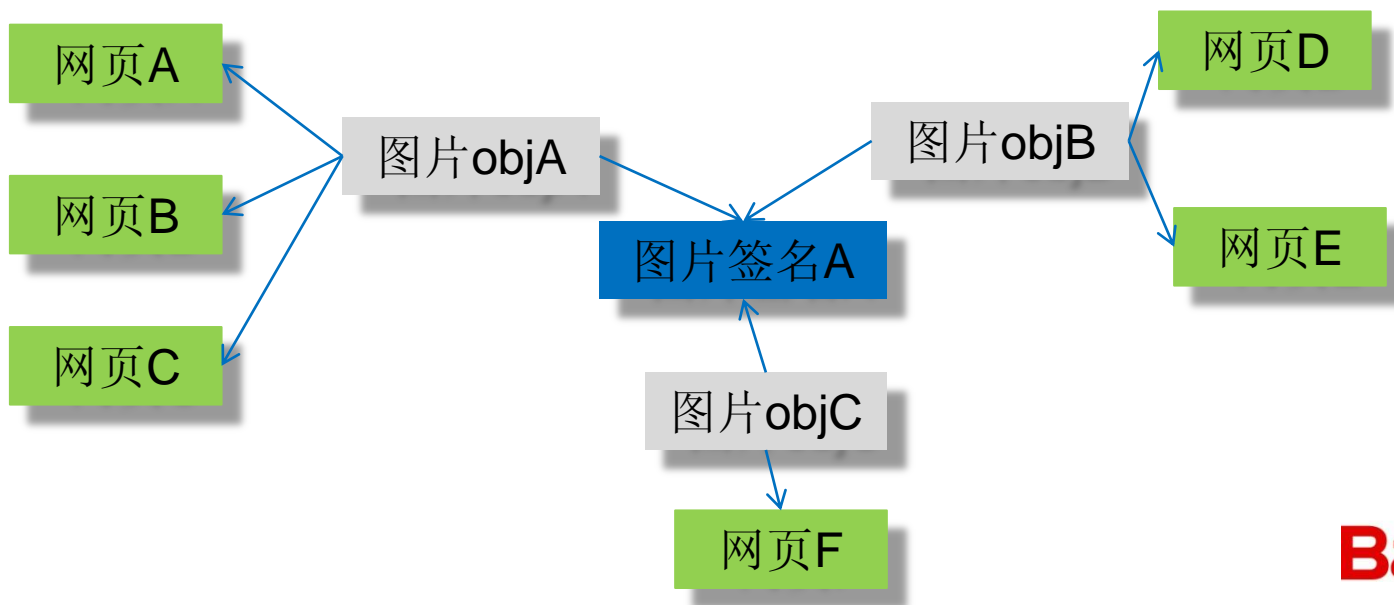


详解-挖掘目标

- 图片与网页的关联

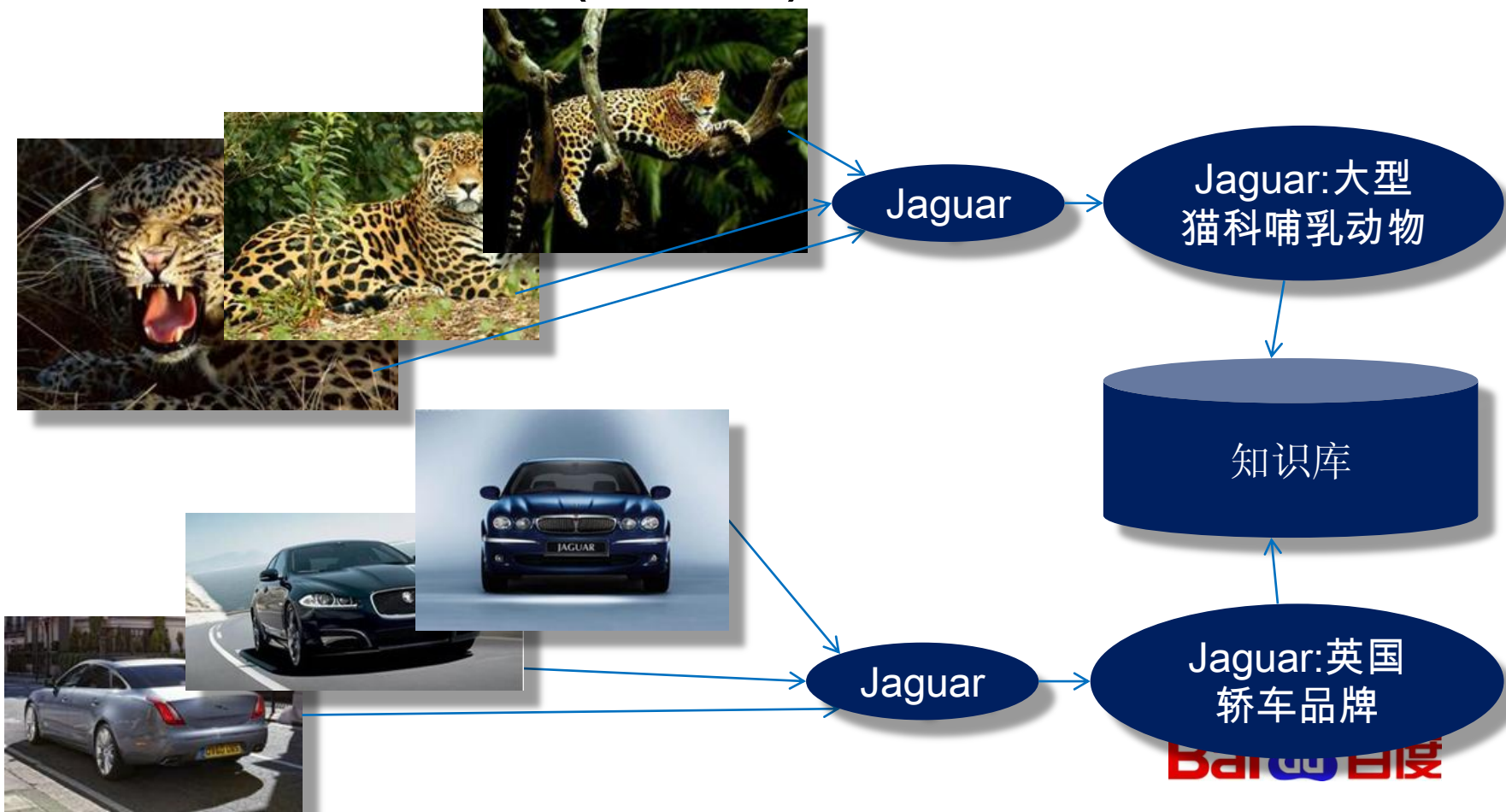
— 层次1：将图片objurl与网页关联，即识别代表性图片

— 层次2：将图片签名与网页关联



详解-挖掘目标

- 图片与实体义项（概念）的关联



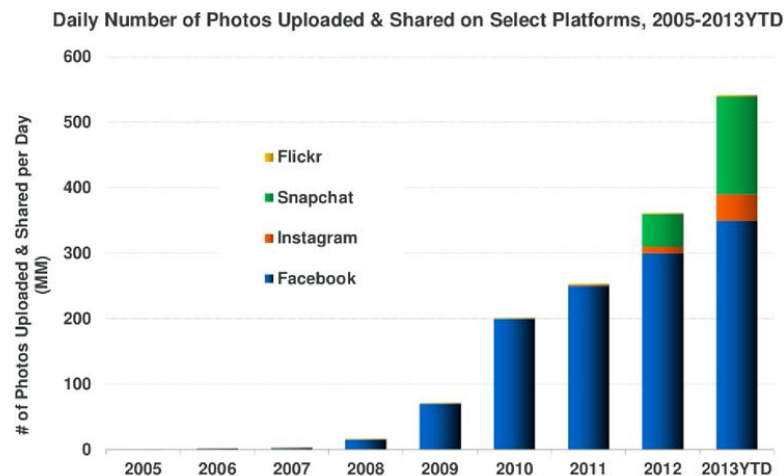
详解

- 图片挖掘的目标
 - 挖掘图片自身特征
 - 挖掘图片与其他数据的关联
- 图片数据的特点
 - 规模、形态、质量
- 图片挖掘技术
 - 图片标注技术
- 产品应用

详解-数据特点

- 数据规模：已在万亿量级并且在快速增长
— 挑战：兼顾算法的效果与效率

Photos = 500MM+ Uploaded & Shared Per Day,
Growth Accelerating, on Trend to Rise 2x Y/Y...



KPCB

Source: KPCB estimates based on publicly disclosed company data. 14

详解-数据特点

- 数据形态：图片数据与网页信息并存
- 挑战：如何利用网页信息辅助图片分析

刘德华 [编辑]

关于其它主题，请见“[刘德华 \(消歧义\)](#)”。

刘德华，[MH](#)，[JP](#)（[Andy Lau Tak-wah](#)，1961年9月27日－），[香港演员兼歌手](#)，1990年代获封香港乐坛“[四大天王](#)”之一，^[6]也是[吉尼斯世界纪录大全](#)中获奖最多的香港歌手^[6]；电影方面他已三获[香港电影金像奖](#)最佳男主角奖，并获得两座台湾[金马奖](#)影帝，截止2013年参演超过140部电影^[7]。刘德华现在是[映艺集团](#)的老板，作为投资人参与制作了20多部[华语电影](#)。^[8]

1999年，刘德华获得“[香港十大杰出青年](#)”的荣誉，2000年11月则顺利荣登“[世界十大杰出青年](#)”^{[9][10]}，成为获此殊荣的少数几位香港艺人。2006年7月7日，[香港演艺学院](#)因他“是香港最受尊重和喜爱的演艺名人之一，对香港电影及音乐贡献良多。其严谨专业的工作态度，足以成为年轻人的典范”，并且为了“表彰他在表演艺术方面的成就”而授予刘德华荣誉院士称号，^[11]他也因此成为少数几位获此荣誉的香港艺人之一^[12]。

刘德华笃信[佛教](#)，[法号](#)“[慧果](#)”，热心公益，时常参与[慈善](#)活动。2008年，刘德华获[香港特别行政区政府](#)委任为[太平绅士](#)^[13]。2010年4月23日，刘德华获任[中国残疾人福利](#)



摄于2005年2月24日，第29届香港国际电影节

刘德华

男艺人

罗马拼音 Lau Tak-wah

英文名 Andy Lau

昵称 华仔、华哥、华Dee、刘天王、刘德华

别名 福荣（入学所取的学名）、紫慧（作词笔名）、慧果（佛教法名）

国籍 中国（香港）

详解-数据特点

- 数据质量：图片质量良莠不齐
一挑战：如何自动预测/计算图片质量



详解

- 图片挖掘的目标
 - 挖掘图片自身特征
 - 挖掘图片与其他数据的关联
- 图片数据的特点
 - 规模、形态、质量
- 图片挖掘技术
 - 图片标注技术
- 产品应用

详解-挖掘技术

- 图片与实体义项的关联：

- 新任务

- 存在很多应用需求，包括实体配图/图片检索等

- 学术界工作集中于：

- 为图片生成标题
 - 为图片生成关键词
 - 为图片判定语义类别

详解-挖掘技术

- 难点：

- 一问题拆解

- 无法一步解决
 - 拆分为两个子问题：图片->实体名,实体名->义项

- 一子问题形式化

- 图片、实体数量庞大,无法采用分类方式解决(训练数据,类别数目)
 - 转化为关联度打分问题

详解-挖掘技术

- 方案

- 图片映射到实体名

- 将图片聚簇，得到相同图片集合
 - 基于多页面文本信息校验，获取候选tag
 - 基于多特征，计算图片的关联实体

- 实体名映射到实体义项(entity linking)

- 基于实体上下文信息

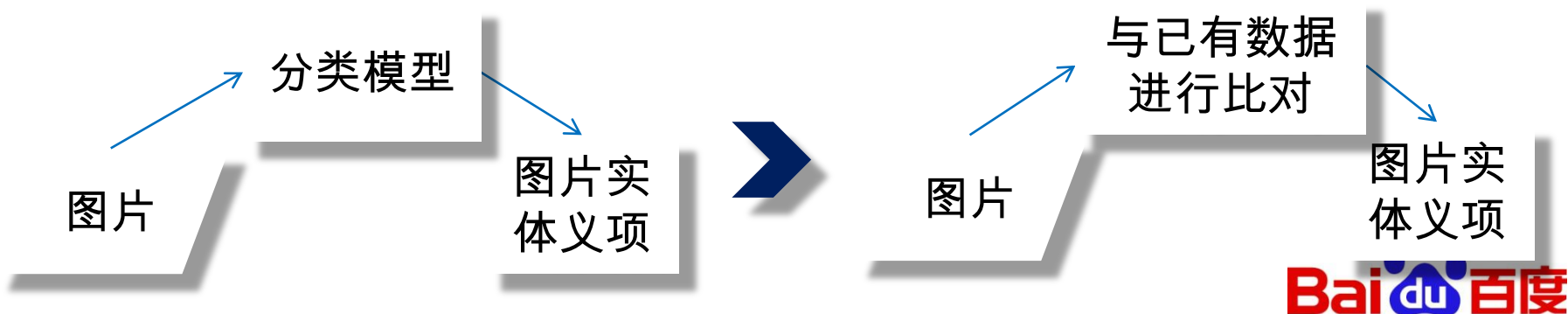
详解-挖掘技术

- 特色

- 从单特征到多特征



- 从模型驱动到数据驱动



详解

- 图片挖掘的目标
 - 挖掘图片自身特征
 - 挖掘图片与其他数据的关联
- 图片数据的特点
 - 规模、形态、质量
- 图片挖掘技术
 - 图片标注技术
- 产品应用

详解-应用

● 搜索：搜索结果出图

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

如何翻墙

百度一下

[怎样使用最方便的翻墙术（FireFox篇）_百度经验](#)

1 在MyEtnel中设置SS H账号打开MyEtnel... 2 安装AutoProxy插件打开 Firebox, 访问Aut... 3 为AutoProxy设置右键单击Firebox的状态... 4 让AutoProxy使用MyEtnel代理"确定"...

[jingyan.baidu.com](#)

[怎么翻墙上外国网站_百度知道](#)

3个回答 - 提问时间: 2012年05月24日

最佳答案: 用VPN。在我的百度空间有一款,你看看,个人一直在用,速度不错,有6个IP可选,建议先不买,免费用先用,好的再买。包年的话会比包月划算好多,电脑,手机...

[zhidao.baidu.com](#) 2012-05-25

请问怎么翻墙啊?	3个回答	2012-08-13
怎样翻墙	3个回答	2013-09-04
怎么翻墙上网	10个回答	2012-07-29

[更多知道相关问题>>](#)

[谷歌浏览器怎么翻墙 图文教你谷歌浏览器翻墙 浏览下载软件教程...](#)

谷歌浏览器翻墙一词成为了广大网友议论的焦点,目前强烈推荐的是使用 goagent 进行翻墙作业,接下来介绍Windows平台下使用Chrome浏览器翻墙的具体步骤,感兴趣的朋友可以参考下

[www.jb51.net](#) 2014-03-04 - 百度快照

详解-应用

● 广告：面向图片的广告

www.ddjkw.com/a/201403/5461.html

创意礼物 肠胃健康网 宠物健康网



公共营养师—绿色白领职业
三个月圆您高薪白领梦想!

人保授权认证机构
4000名学员共同见证

尚科营养师 400-000-0262

女性都希望自己有着红润白皙的皮肤，但随着年龄的增长，皮肤会逐渐失去弹性，下面就跟着小编了解下把。

造成脸色暗黄的原因有很多，其中最主要的一个原因是气血亏虚引起的。因此通过食疗法对身体进行调节能够起到比较有效的作用。

脸色暗黄怎么调理

美味可口的樱桃就是助你打造粉嫩肌肤的必备之物。樱桃有着艳丽的色彩，味道甜美，而最重要的是它的铁含量位居众水果之首。

我们都知道铁是合成人体血红蛋白的重要原料，只有让体内的血红蛋白维持在一个较高的水平，才能够有效地促进血红蛋白的合成。因此每天食用适量的樱桃能够帮助调理出好肤色哦。

食用鸡鸭肝帮助“藏血”

肝脏是属于藏血的器官，当其运行正常的时候，肝气舒缓，我们的脸颊就会呈现出自然健康的浅粉色，有着如此美好的色彩，即使我们不打腮红，也能够展现出白里透红的好肤色。

而在中医的角度来讲，一向有着“以形补形”的说法，因此平常我们可以通过使用动物肝脏来补肝。每天食用五十克左右的鸡肝或者鸭肝，就能够起到很好的作用。另外平常也多煮一些鸭血汤来喝，能够滋阴补血哦。

豆类和主食可以多吃

详解-应用

● 配图：实体的配图

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

成龙 百度一下 推荐:用手机随时随地上百度

成龙_百度百科

 成龙，1954年4月7日生于香港特别行政区香港岛中西区。祖籍安徽省，国家一级演员，大中华区影坛和国际功夫影星。1960年进入中国戏剧学校学习戏曲，以武师身份进入电影圈。1970年自该校毕业，因为自幼在影片中跑过...[查看百度百科>>](#)

[早年经历](#) · [演艺经历](#) · [个人生活](#) · [主要作品](#) · [社会活动](#)

[baike.baidu.com/2014-04-02](#)

成龙的最新相关信息

[李兆会往事:年少接班后刚愎五叔与成龙是多年好友](#)

 李兆会往事:年少接班后刚愎五叔与成龙是多年好友”此外,自海鑫钢铁遭遇高额逾期贷款被报道以来,事件的主角李兆会一直未现身,行踪成谜。李海仓突...
凤凰网 1天前

[文章成龙似震男星出轨被曝光后的雷人说辞 金鹰网](#) 1天前

[明星出轨结局最惨烈:成龙张学友搞定老婆 韩寒最奇... 人民网](#) 3小时前

[成龙迎60大寿 赴非洲帮助当地发展农业\(图\) 深圳新闻网](#) 6天前

[成龙成都“淘金” 凤凰网娱乐](#) 6天前

成龙的新浪微博

 **成龙 V** 粉丝: 2200万 微博: 751
上海艺术馆... [查看全文>>](#)
3天前 · 新浪微博

[转发\(890\) | 评论\(919\)](#)

成龙电影全集国语版高清_成龙最新电影_成龙主演的电影列表【2345...】

成龙 林正英 周星驰 刘德华 章子怡 舒淇 李连杰 甄子丹 李小龙 周润发 葛优 更多>> 排序方式: 人气| 好评| 最近更新 全部 共有124部 私人订制 5...
[dianying.2345.com 2014-03-21](#) · [百度快照](#)

香港武打演员

[展开](#)

 [李连杰](#)  [洪金宝](#)  [周星驰](#)  [甄子丹](#)

相关香港演员

[展开](#)

 [刘德华](#)  [向华强](#)  [周润发](#)  [廖启智](#)

相关人物

 [林凤娇](#)  [范冰冰](#)  [赵文卓](#)  [林志颖](#)

文化人物风云榜

排名	搜索指数
1 张杰	26460
2 彭丽媛	22390
3 刘恺威	18110
4 周星驰	15150
5 成龙	12180

报告目录

- 图片挖掘概述
- 挖掘技术详解
- 图片挖掘展望

展望

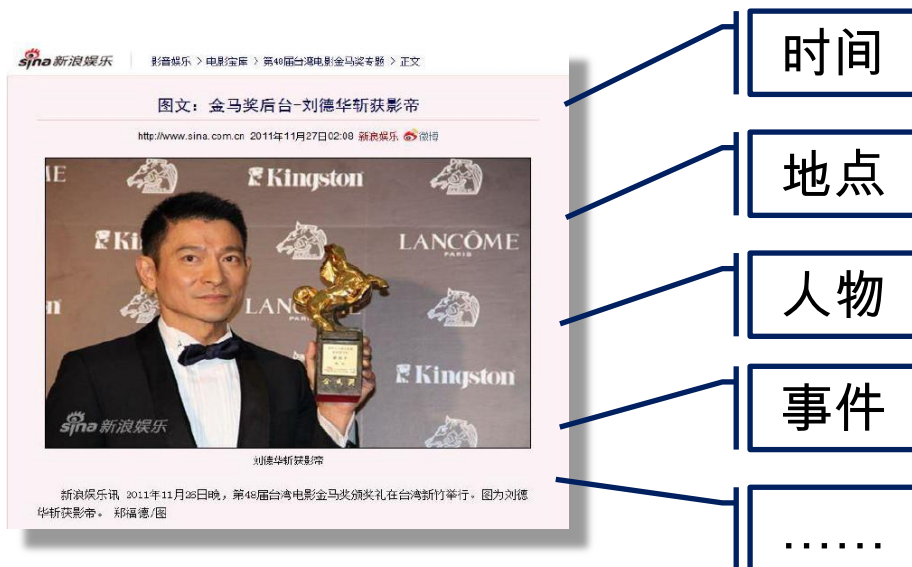
- 支持数据建模
 - 建立图片知识库
- 支持用户的任务与决策
 - 支持购物、决策等

展望

- 支持数据建模

- 一分析图片重要属性，形成结构化数据

- 一建立面向图片的知识库,并与实体知识库关联



展望

- 支持用户的任务与决策

- 分析看图背后的需求：购物、决策等

- 满足用户的需求



分析
用户
状态、
需求

整合
信息
满足
需求



天气



出行
方式



攻略

总结

- 回顾

- 图片挖掘概述：图片已成为主流信息载体
- 挖掘技术详解：大部分技术问题已解决
- 图片挖掘展望：深挖图片知识/支持用户任务
- 面临的技术挑战：图片知识挖掘



谢谢！

