

最懂你的



— 基于社交挖掘和个性化推荐的新媒体

字节跳动管理层陈述书

2013年1月

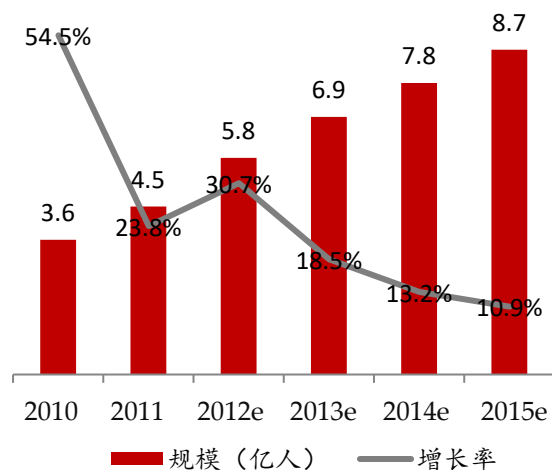
投资亮点

- 1 即将爆发的个性化数字媒体市场
- 2 独创的个性化资讯发现引擎
- 3 领先于世界同类产品的功能和技术
- 4 行业领先的用户粘度和自然增长
- 5 完善的多产品布局，覆盖移动终端和PC端
- 6 具有丰富创业经验和技术功底团队

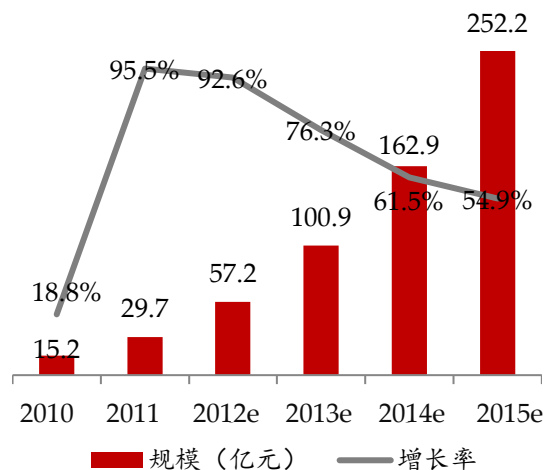
移动互联网开启了崭新的个性化数字媒体市场



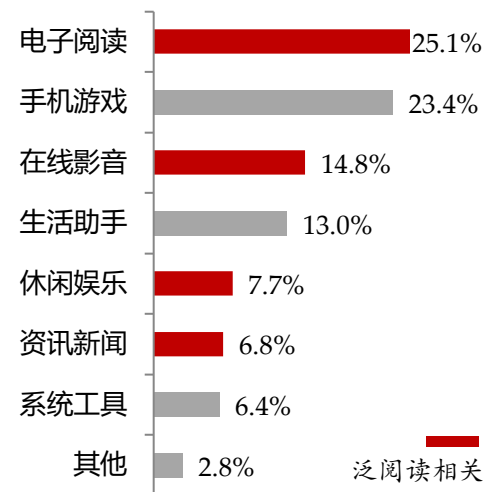
中国手机网民规模
持续稳定增长



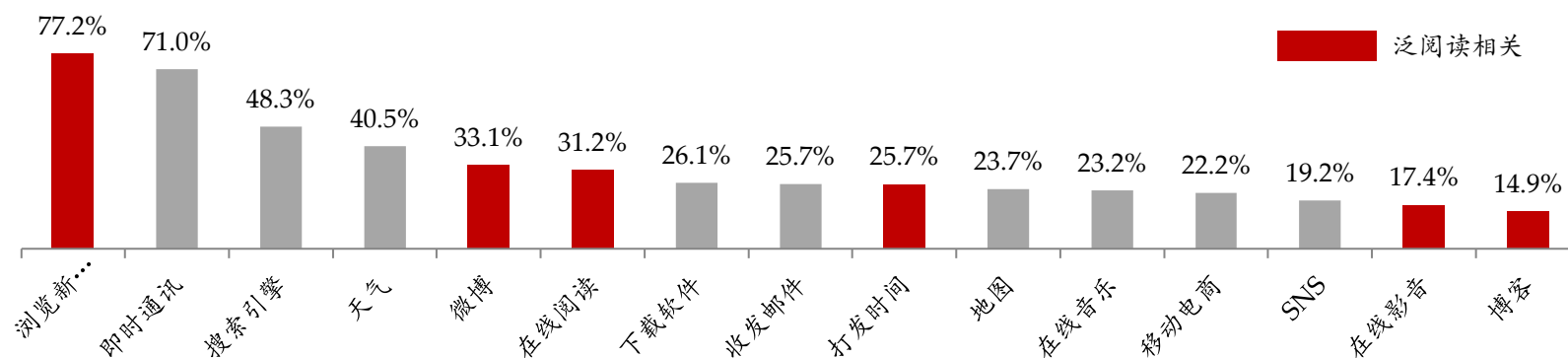
中国移动互联网广告
市场继续高速增长



泛阅读类应用已成为移动广告投放重要媒介⁽¹⁾

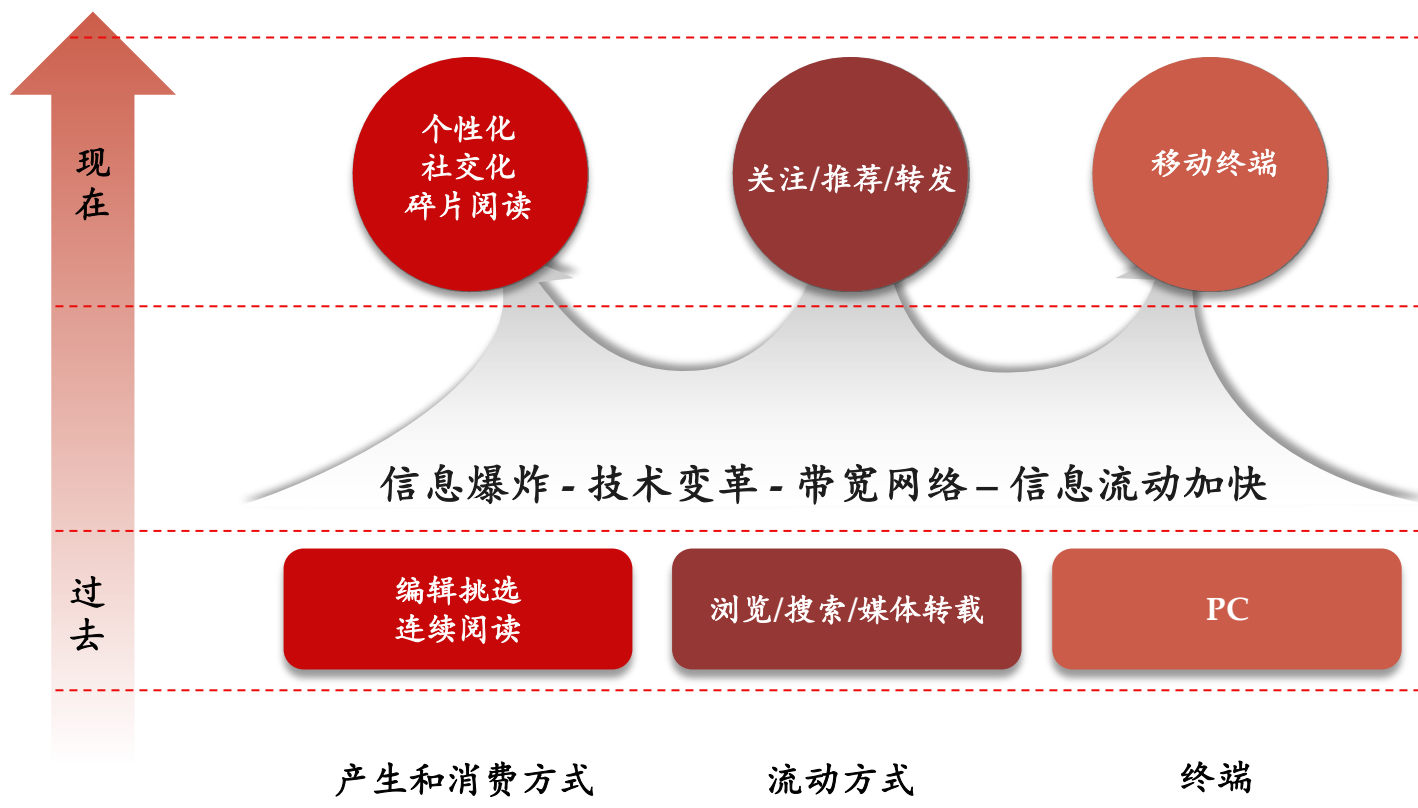


泛阅读是移动互联网最主要的用户行为⁽²⁾



数据来源: (1)多盟 (2)DCCI《2011中国移动互联网蓝皮书》

来源多样化的丰富内容与移动时代的碎片化小屏幕阅读的矛盾



头条 - 独创的个性化资讯发现引擎



最好的满足了更个性、更便捷、更互动的阅读体验



用户的表达、交流与互动信息

头条通过推荐技术、个性化、兴趣图谱等，打造下一代数字媒体

陌生的人
以内容为中心
兴趣信息
能够发现新的信息

熟悉的人
以人为中心
生活工作信息
大量冗余、重复信息

媒体属性

社交属性

豆瓣 douban



蘑菇街
mogujie.com



- 面向兴趣，基于推荐，社交分析热启动并紧密结合sns
 - 内容短，类型丰富
 - 移动优先
-
- 足够大的市场：信息获取消费是足够大的市场，是高频、刚性的需求。领先者能达到千万DAU水平
 - 移动将重新洗牌：传统的门户模式在移动终端都不适用
 - 建立门槛：1) 海量的用户群；2) 个性化的用户数据；3) 技术积累；4) 兴趣图谱

无需订制、点选，启动即可阅读最新最热的个性化资讯，丰富综合的资讯内容，不仅仅是新闻

简单丰富

最懂你

高质量评论

兴趣社区



热门、体育

乔布斯、博客

育儿

用户行为和社交分析，基于多种模型挖掘上万维度兴趣特征，发现属于你的头条，用的越多越懂你

简单丰富

最懂你

高质量评论

兴趣社区

乔布斯
博客

互联网
科技



韩国
娱乐

动物

育儿

具有高质量、丰富及时的评论互动功能，并根据社交关系和影响力及热门程度等属性进行排序

简单丰富

最懂你

高质量评论

兴趣社区



众多聚合评论

好友优先

影响力优先

受欢迎程度



和SNS紧密结合每日十几万分享。同时为用户提供邀请，关注、动态，评论排序等功能建设自己的兴趣社区

简单丰富

最懂你

高质量评论

兴趣社区



提供Web版本，用户可以通过任意设备，随时访问



The screenshot displays the Toutiao web application interface. At the top, there's a navigation bar with '头条' (Toutiao) and '给你最想看' (Show you what you want to see). Below this, a sidebar on the left lists various categories like '全部' (All), '社会' (Society), '财经' (Finance), '体育' (Sports), '娱乐' (Entertainment), '科技' (Technology), '军事' (Military), '时尚' (Fashion), '国际' (International), and '汽车' (Cars). The main content area features several news articles. The top article is titled '罗纳德·科斯：计划生育是最奇葩的政策 具有潜在毁灭性影响' (Ronald Coase: The one-child policy is the most bizarre policy, with potential destructive impact). Below it, there's an article about '移动互联网草根狂欢：三四线城市用户价值凸显' (Mobile Internet grassroots狂欢: The value of users in third and fourth-tier cities is highlighted). The interface also includes a '推荐用户' (Recommended users) section on the right, a 'TechWeb网站 诚聘英才' (TechWeb website, urgently hiring talent) banner, and a '24小时热门评论' (24-hour popular comments) section at the bottom right. The overall design is clean and modern, with a focus on user engagement and content discovery.

点击文章链接后，显示来源网页，用户同时可以看到相关评论并参与评论。这样既提供良好的用户体验，又保障内容提供商利益

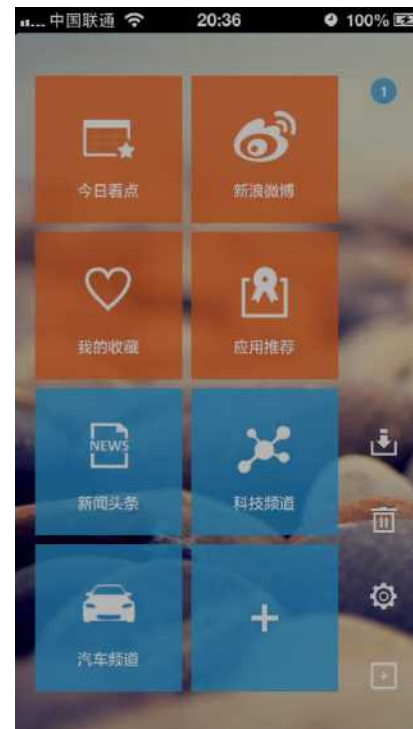
头条提供泛资讯内容，覆盖新闻、博客、影评、段子等；并通过个性化，精确推荐读者感兴趣内容



社交兼媒体，信息杂

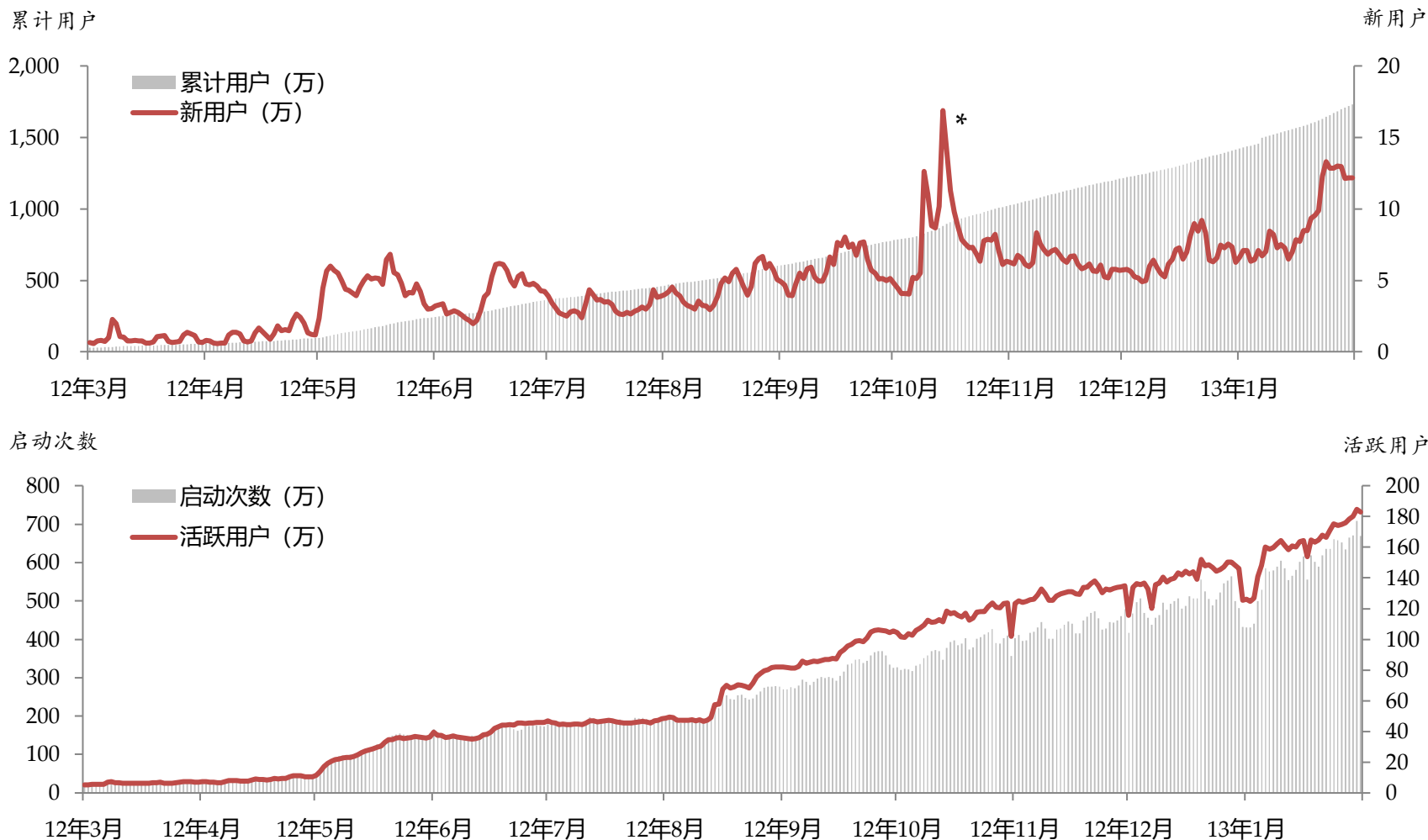


编辑推荐新闻



需手工订阅选择喜好

依靠高用户粘度和良好的用户口碑，在极小规模推广投入下，保持了优异用户增长态势



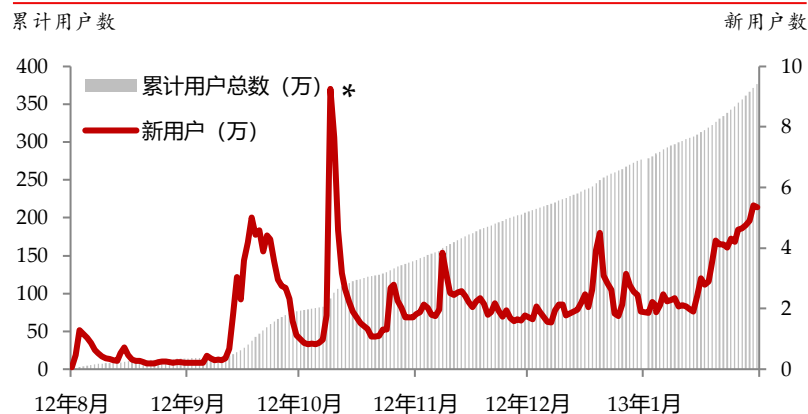
数据来源：友盟统计分析，包含所有应用

* 10月友盟统计偏差导致数据偏高

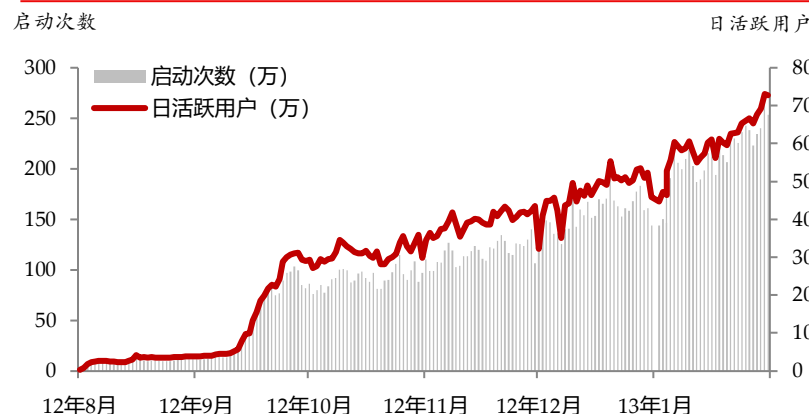
头条上线5月，留存率和用户活跃度已达行业领先水平



累计用户总数



日活跃用户逐月攀升

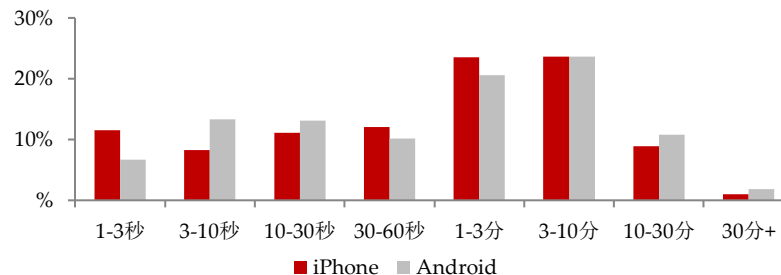
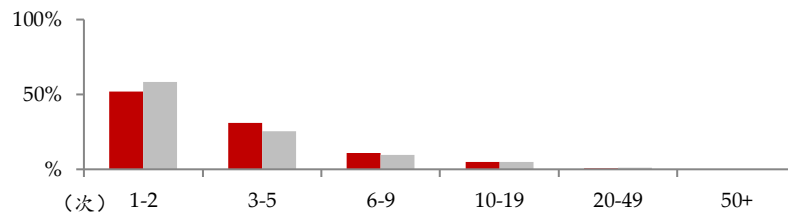


《今日头条》平均次周留存率约60%，8周后留存率约40%

日均启动次数/次使用时长

首次使用时间	1周后	2周后	3周后	4周后	5周后	6周后	7周后
(iPhone, %)							
09/12 ~ 15/12	69.6%	59.5%	56.5%	54.1%	51.1%	47.5%	45.1%
16/12 ~ 22/12	58.6%	51.0%	47.5%	44.3%	40.9%	38.5%	
22/12 ~ 29/12	62.5%	55.3%	50.7%	46.3%	43.1%		
30/12 ~ 05/01	68.6%	59.7%	53.4%	49.4%			
06/01 ~ 12/01	67.6%	57.3%	51.7%				
13/01 ~ 19/01	68.6%	57.0%					
20/01 ~ 26/01	68.8%						

首次使用时间	1周后	2周后	3周后	4周后	5周后	6周后	7周后
(Android, %)							
09/12 ~ 15/12	60.2%	46.4%	42.3%	39.4%	35.5%	32.2%	29.2%
16/12 ~ 22/12	52.2%	40.9%	36.8%	32.4%	28.9%	26.0%	
22/12 ~ 29/12	40.4%	30.6%	25.4%	22.4%	19.1%		
30/12 ~ 05/01	53.7%	40.3%	34.2%	29.7%			
06/01 ~ 12/01	50.8%	37.5%	31.2%				
13/01 ~ 19/01	47.0%	32.9%					
20/01 ~ 26/01	45.5%						



数据来源：友盟统计分析

* 10月友盟统计偏差导致数据偏高

1 独创的数据处理和推荐技术底框架

2 自然语言、多媒体信息处理

3 高维度用户兴趣建模

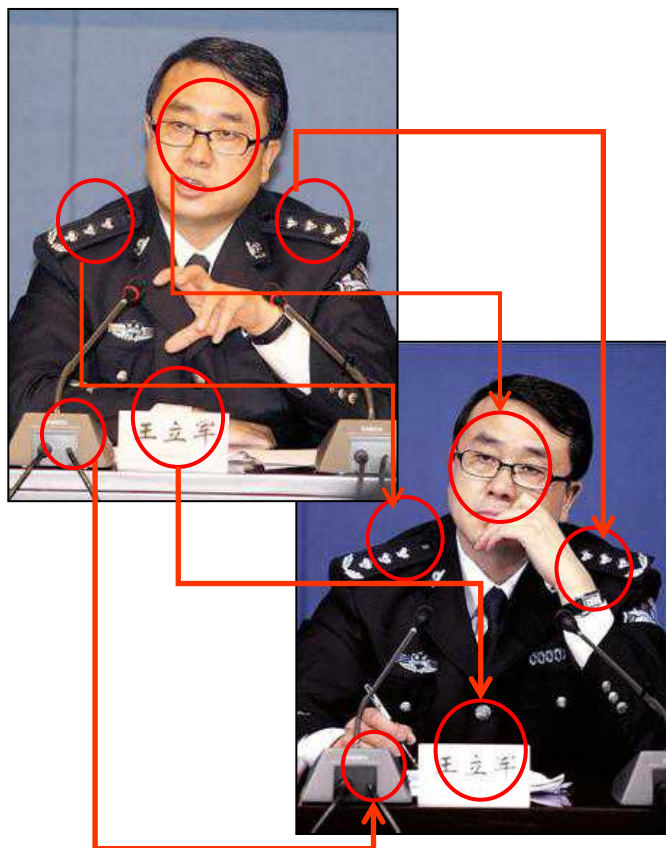
4 高性能的实时大规模数据运算

5

The diagram illustrates a complex data processing and user interaction system. It is organized into several main sections:

- Data Sources:**
 - Web Data & Social Data:** Represented by a large grey arrow pointing into the **Spider** component.
 - Fundamental Service & Other Tools:** A vertical list of services including Store, Log, Cache, Monitor, Calculation, Other Tools, User Operation Sys, Product Analysis Sys, BD Sys, Data Analysis Sys, and OP Sys.
- Data Processing Pipeline:**
 - Spider:** Processes raw data using API Based, Web Based, and Visual Based Crawlers.
 - Extractor:** Extracts Video, Image, and Text data, and performs ReGarbage, ReDup, and Main Info Extraction.
 - Miner:** Performs High Dimension Matrix Operation, Cluster, Reduce Dimension, Global Historical Statistics, and Classify.
 - Recommender:** Uses Indexer, Sorter, and Merger to process data.
 - Ad Recommender:** A dashed box representing the final recommendation stage.
- User Platform Server:**
 - API:** Acts as the central interface for the User Platform Server.
 - Log Processor:** Processes user actions and impressions.
 - Analyzer:** Analyzes user data and interests.
 - Privacy Control:** A dashed box ensuring data privacy.
 - User Data & Interest DB:** A database storing user data and interests.
- Output & Interaction:**
 - Web & Apps:** The system interacts with web and app environments.
 - Plugin:** A dashed box representing a plugin interface.
 - Series:** The final output of the system, represented by a series of data points.

自然语言和多媒体信息聚合能够实现分类去重，极大降低信噪比



对每条信息提取几十个到几百个高维特征，并进行降维、相似计算、聚类、分类等处理

- 1. 日本发生7.3级地震发布海啸预警 东京震感强烈
- 2. 快讯：日本东部地区发生7.3级地震 震源深度33公里
- 3. 日本发生7.3级地震
- 4. 日本东北部发生7.3级地震 或引发约一米高海啸
- 5. 日本7.3级地震 NHK主播大呼快跑！宫城发布海啸警报



日本发生7.3级地震！东京震感强烈，宫城发布海啸预警

- 我经常幻想十来年后的有一天，一个人对我大声呵斥：“我告诉你，钱不是万能的！！”



我的目标是诺千年后。。会经常有人对我大声呵斥：钱不是万能的！！

根据用户在社交网站的社交关系和内容行为来抽取个性化信息

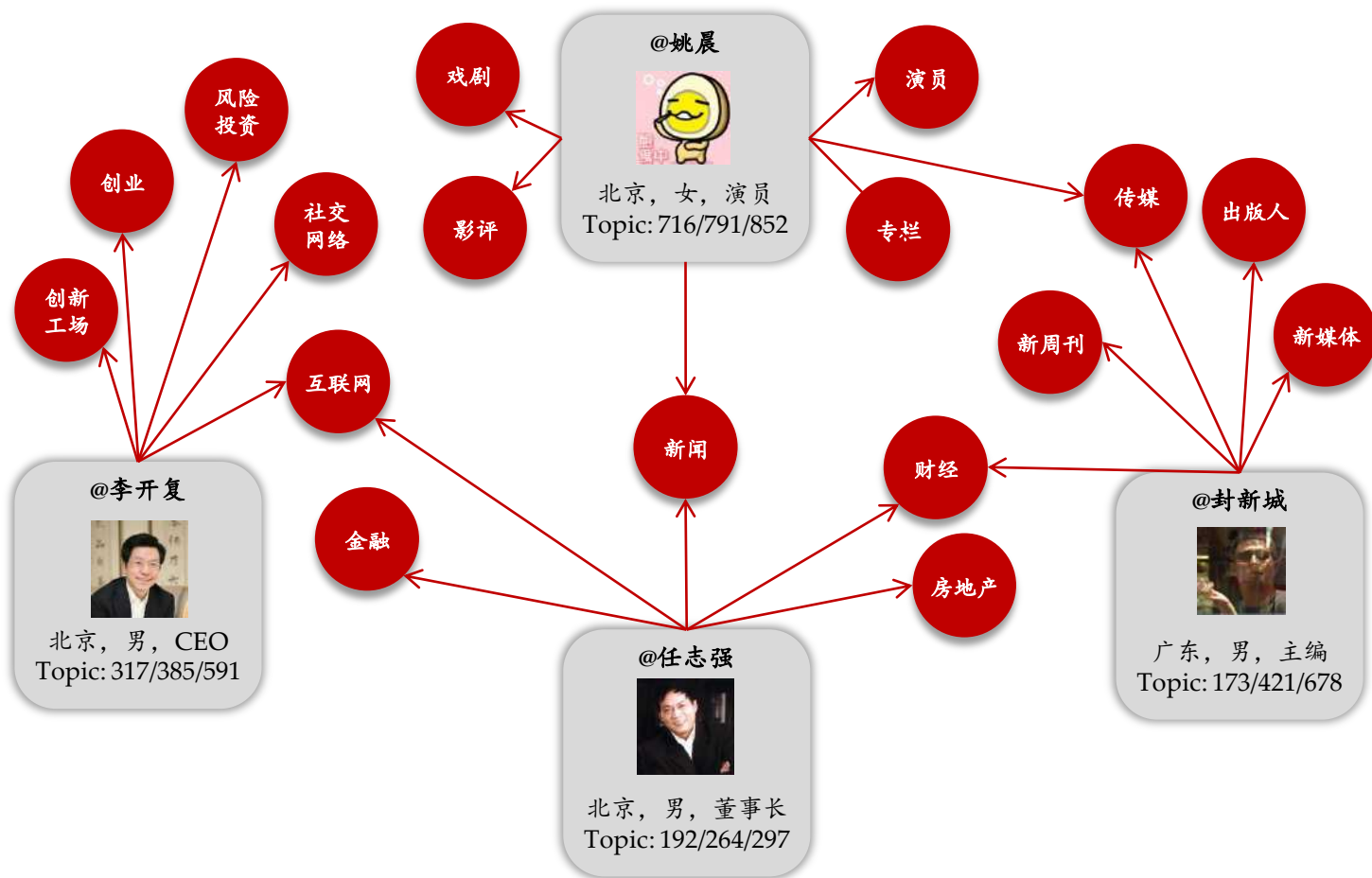
通过API分析用户的转发、评论、微博、收藏，以及用户之间的关系网络；5秒内，计算出用户几千个维度上的兴趣分布

- 工业设计
- 交互设计
- 用户体验
- UI
- 视觉设计
- Design
- 界面设计
- UI设计
- 产品设计
- UX

- iPhone
- 国行
- 设计模式
- 照明系统
- 可乐
- 减重
- 互动
- 数据分析
- 国产品牌
- 篮球



为每一位用户通过多种模型计算上万个维度的兴趣分布，并且每次动作后30秒内就会更新用户的兴趣分布



将所有文章和用户之间进行模型匹配，实现个性化推荐

属性/文章		北京	创业	房地产	新媒体	影评	互联网
	[所有,Topic 716] 《泰囧》破华语片票房纪录 13天超《画皮2》	0.12	0.27	0.05	0.13	0.82	0.12
	[所有,Topic 192] 住房公积金最高与最低超30倍 缴存被指数衍	0.23	0.11	0.91	0.15	0.19	0.08
	[所有,Topic 385] 揭秘世纪佳缘创始人谢幕背后	0.08	0.90	0.03	0.12	0.04	0.86
	[女,Topic 852] 头发侧边放造型大PK 刘亦菲完胜大S	0.06	0.01	0.01	0.02	0.03	0.01
	[所有,Topic 678]主持人 2013中国新媒体产业十大趋势	0.14	0.35	0.01	0.98	0.01	0.67

海量数据处理

- 每日处理存储千万条数据，百万张图片，数万个视频
- 维护所有用户的展示、动作历史，每日存储上亿展示记录，上千万动作记录



行业内顶尖的技术团队，汇集来自微软、腾讯、百度等公司的技术精英

实时信息计算

- 每位用户授权SNS信息后，5秒内计算出兴趣分布
- 每个动作产生后30秒内更新用户兴趣
- 每条信息10秒内完成提取，挖掘，消重，分类
- 每次请求均根据每位用户的兴趣和历史，进行实时推荐和返回

高性能架构

- 使用C++和python开发的分布式系统
- 多年创业公司工程经验，基于开源和自研的高性价比和可靠性的在线系统
- 目前每日近200万DAU(含WEB)，千万条数据处理，上亿次实时请求，使用30台服务器

技术、产品、研发导向的核心成员，在搜索与数据处理，社交网络挖掘，移动产品方向经验丰富

张一鸣
创始人
CEO

- 2012年3月，创立字节跳动，专注在移动新媒体；
- 2009年9月，创立九九房，CEO，推出九九房搜索，掌上租房，掌上买房等产品；
- 2008年，加入海内/饭否管理团队，合伙人，负责后台技术；
- 2006年，加入酷讯，创业成员，搜索团队技术负责人，技术委员会主席；
- 2005年，毕业南开大学软件工程专业；

黄河
产品

- 9年经验，九九房移动产品总监，自我创业，北京工业大学；

梁汝波
研发

- 8年经验，艾克赛利主任工程师，九九房研发总监，南开大学；

屠锋锋
商务拓展

- 11年经验，Oracle美国、微软中国高级产品经理，海豚浏览器VP，北京大学、德州大学；

充满挑战与机会的推荐媒体方向和优秀的团队不断吸引一流的人才加入

产品

- 陈林： 5年经验，微软、腾讯工程师，Global Apps创始人，北京大学

技术

- 胡殿伟：5年经验，曾任职微软、腾讯，Global Apps创始人，清华大学
- 李秋良：5年经验，九九房 Android 高级工程师，北航
- 周晶锦：3年经验，九九房高级工程师、研发经理，清华大学
- 银平： 6年经验，酷讯高级工程师，乐投科技构架师，北京大学
- 马志刚：7年经验，腾讯研究院高级工程师，北京大学博士
- 张永华：6年经验，微软研究院、腾讯研究院高级工程师，天津大学
- 陈韬： 7年经验，华为，伟速科技创始人，南开大学
- 李飞： 4年经验，酷讯网、赶集网高级工程师，研发经理、南开大学
- 刘增禄：6年经验，墨者在线手机之家、兴长信达高级工程师、构架师，南京大学
- 白庆奇：3年经验，百度，西安交通大学

行政/人力

- 肖金梅：6年经验，九九房HR经理，人民大学

在扎实的技术基础上，2013年产品会不断突破，加强进入壁垒，并尝试国际化和商业化

扩充信息类型和来源

- 扩充数据量，来源，类型：覆盖商品、电影、消费等数据，将更多数据类型融合到头条，数据更丰富，优化重要单款的数据质量，加强UGC的功能
- 数据处理系统升级：加强文本、多媒体的处理能力，让信息质量更高

深挖用户特征，升级推荐系统

- 交互改进：让用户更容易产生动作，积累更多用户行为
- 推荐系统升级：开发更多推荐模型，以及模型的组合，以及模型的测试对比系统
- 用户特征：挖掘用户更多不同类型的兴趣，增加不同的推荐维度

强化社区互动

- 构建兴趣社区：用户搜索，推荐，榜单等功能，让用户更容易发现感兴趣的人
- 加强互动：私信，群组，让用户之间互动交流更流畅
- 打通SNS：强化分享功能和插件，回馈更多数据到SNS和网站，并吸引更多的人加入

国际化和商业化尝试

- 国际化：推出和Facebook、Twitter结合的英文版，获取英语国家用户
- 商业化：推出广告系统，基于用户兴趣和场景的推荐广告

巨大的广告价值

高覆盖渗透、极具黏性



- 照顾用户对信息的多种需求，高频使用，很高的启动次数和时长
- 全信息类型，覆盖不同层次用户群，面向所有移动互联网用户

数据精准丰富



- 通过社交网络和行为数据能够采集丰富的兴趣爱好数据
- 强大的数据收集和分析能力，有能力从数据中抽取商业价值
- 移动设备的唯一性和定位的特点，能够提供地理位置属性

广阔的变现潜力

广告即内容的最佳载体



- 信息流的展现形式，用户专注的阅读，是广告即内容的最佳载体
- 能够提取用户兴趣爱好，是将广告转换为内容的前提

数据输出



- 强大的信息处理能力，能够与媒体合作构建深入的合作关系，提供评论组件、信息聚合等服务
- 依赖丰富精准的数据，可以为媒体提供更加精准的基于用户的广告投放，通过数据变现

本轮计划融资**1250万美元**，支持未来**2年**的发展

	2012	2013	2014
移动DAU	150万	750万	1,500万
人员数量	38	69	100
人员及办公支出(\$ 000)	950	1,785	2,588
市场费用(\$ 000)	250	3,960	5,376
服务器支出(\$ 000)	250	508	1,004
支出总计(\$ 000)	1450	6,254	8,968
收入(\$ 000)			6,000

- 人员增长：
 - 技术：工程构架，算法模型
 - 商业和媒体经验人才
- 市场费用主要用于App推广和品牌建设

附录

