

Machine Generated Content

人工智能时代的内容生产方式

CTO 沈国龙



内容消费市场到底有多火？



- 微信，截止2016年6月，近2000万公众号，每日新增认证公众号1万。近 80% 微信用户关注公众号，2015年拉动的信息消费达到952亿元。
- 新浪微博，2016年自媒体收入117亿。今年三季度，自媒体日均发布头条文章13万篇，阅读量超过1亿次，日均发布图片1.6亿张，浏览量47亿次。
- 今日头条，截至2016年10月底，累计6亿激活用户，日活超过6600万，月活1.4亿，单用户日均使用时长则超过76分钟。12万+头条号，每月增幅超过30%。
- 百度百家号CPM(千人成本)达到了3元左右，承诺100%分成。
- 一点咨询、网易新闻、搜狐新闻、凤凰新闻、知乎专栏、天天快报。

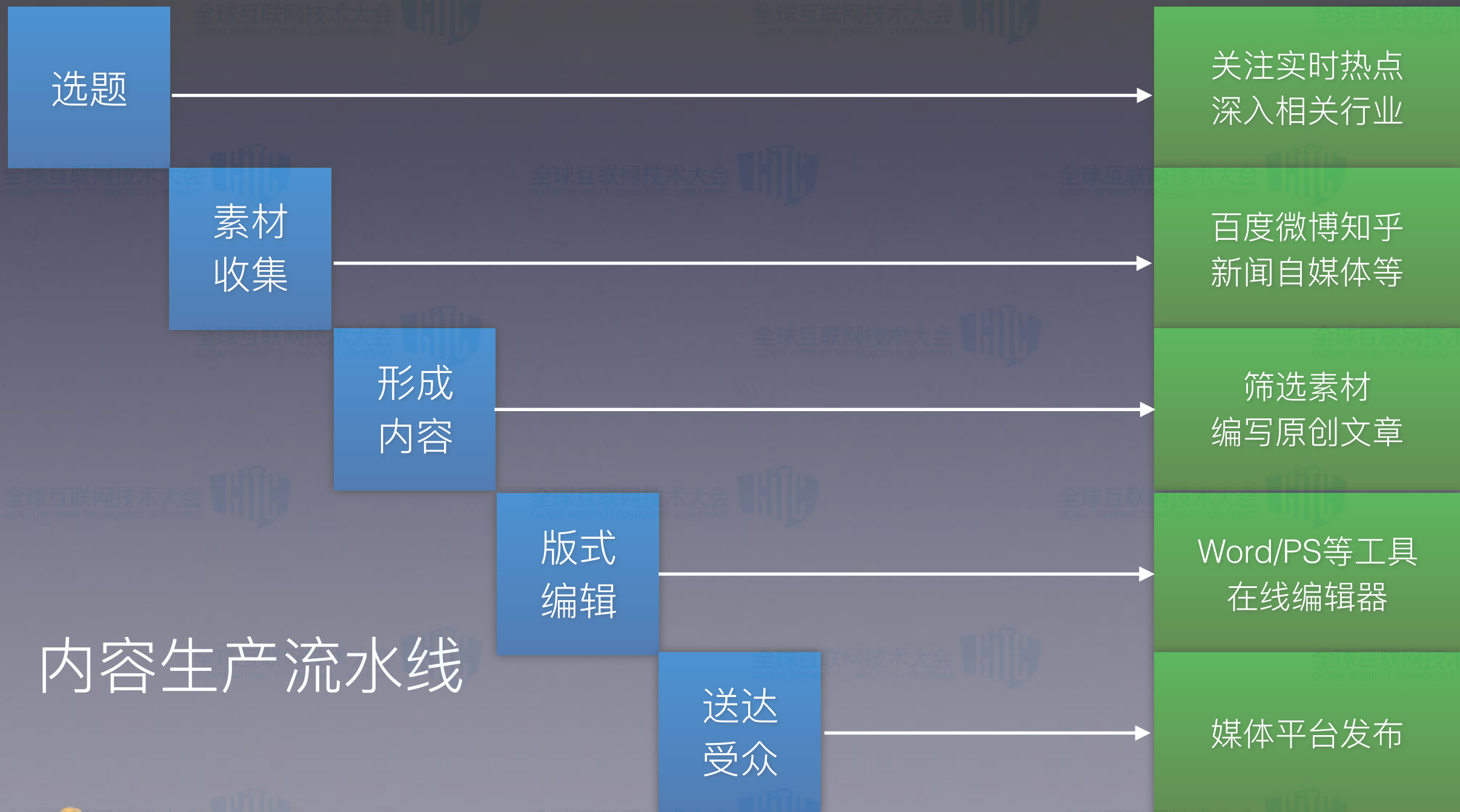
流量有了新的出口

- 碎片化时间增多，移动互联时代的红利
- 电商/游戏/搜索引擎的格局被打破
- 流量追逐优质内容，自媒体确立优势

主流内容生成方式

- UGC(User-generated Content)
 - 用户原创内容，带来了web2.0的兴起。随着信息过载和内容低质逐渐衰落。
- OGC(Occupationally-generated Content)
 - 职业生产内容，原创或购入版权。如各大新闻站点、视频网站。
- PGC(Professionally-generated Content)
 - 生产专业的内容，更多发自爱好，属于UGC的一部分，与OGC的区别在是否有报酬。

自媒体编辑是怎么工作的



每个人都需要信息漏斗

新闻/文艺作品/自媒体等

文字/图片/视频素材

海量文章

用户

在互联网用户都感觉信息过载的时代，
编辑们其实做了信息的第一步筛选。
但，编辑的信息过载如何解决？

“人工智能元年”

- AlphaGo只是一个开始，逐渐会有杀手级的智能应用上市
- 现在人工智能应用最成功的领域：语音、图像、用户画像、自然语言处理
- 智能时代的开始，首先需要借助机器去解决人做的不好或者难做好的工作，比如整理资料、死记硬背等
- “如果一个典型的人进行一项思考任务所需的时间少于 1 秒，那么也许我们现在或不远的将来就能用人工智能自动化这项任务了。” - Andrew Ng

人工智能生产内容-MGC

- Machine-generated Content, 机器生成优质内容
- 将大数据与机器学习能力应用于媒体行业
- 内容生产从农耕时代进入智能时代
 - 1.提升信息源的收集宽度、深度和速度
 - 2.储存海量素材，囊括全行业
 - 3.提升效率，帮助编辑快速完成初稿或直接生成文章，自动完成校对
 - 4.通过模型预测文章的流程度，更好的配合信息流推荐系统



MGC对媒体的影响逐渐放大

- 纽时时报Blossomblot，基于协作工具Slack软件上的一个虚拟智能机器人。
- 美联社2014年7月引入Automated Insights的WordSmith平台自动生产财经报道。
- 腾讯Dreamwriter，财经写稿机器人。根据算法第一时间自动生成稿件，瞬时输出分析和研判。
- 新华社“快笔小新”撰写体育赛事中英文稿件和财经信息稿件。
- 今日头条Xiaomingbot的AI机器人，在今年奥运会期间，完成四百多篇新闻稿，每篇只需要不到两秒钟，有一百多万的阅读。



MGC应用了哪些技术

知识图谱

概率模型

朴素贝叶斯、HMM、TF-IDF

分布式计算和存储

Elasticsearch、Spark SQL

数据收集

机器理解

知识库

语言模型

文章生成

自然语言处理

语义分析、实体识别、词性标注

深度神经网络

CNN、RNN、LSTM

主题模型

LDA、PLSA

定制Spider

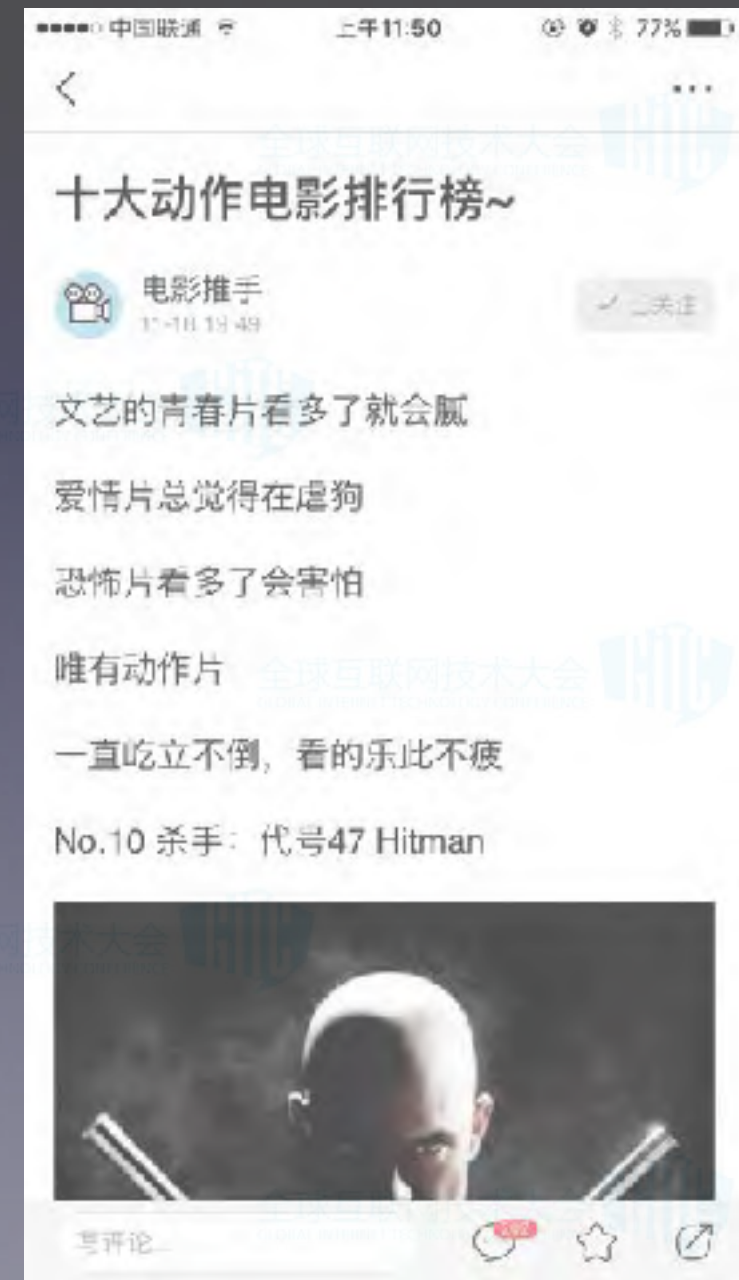
想象空间有多大

- 写作从简单到复杂可以分为五个层次：
 - 书写完整的句子
 - 组织句子构成符合逻辑的段落 — 智能编辑助手
 - 给予特定格式或模板，能够清晰传递信息表达意思 — 机器人编辑
 - 不限定格式的写作内容，达到一般人平均水平
 - 达到专业记者、作家和学者水平
- 高效智能的编辑助手可以帮助作者提升800%的写作效率
- 智能机器人编辑会最早应用在财经、体育、娱乐和商业文件，百亿级的市场规模

机器人编辑的两个例子



按照固定的格式
机器全自动生成文章
阅读可以达到10w+



吉顿特正在做什么

- 基于MGC的多个自媒体运营一个月，阅读数百万
- 智能编辑助手EditBang-爱迪帮在2016年底正式商用
- 人工智能编辑Franky在2017年中旬问世





谢谢

