

专知-基于知识图谱的内容分发

方全 副研究员

中国科学院自动化所 模式识别国家重点实验室

2018-01-30

提纲



专知介绍



专知技术



专知应用

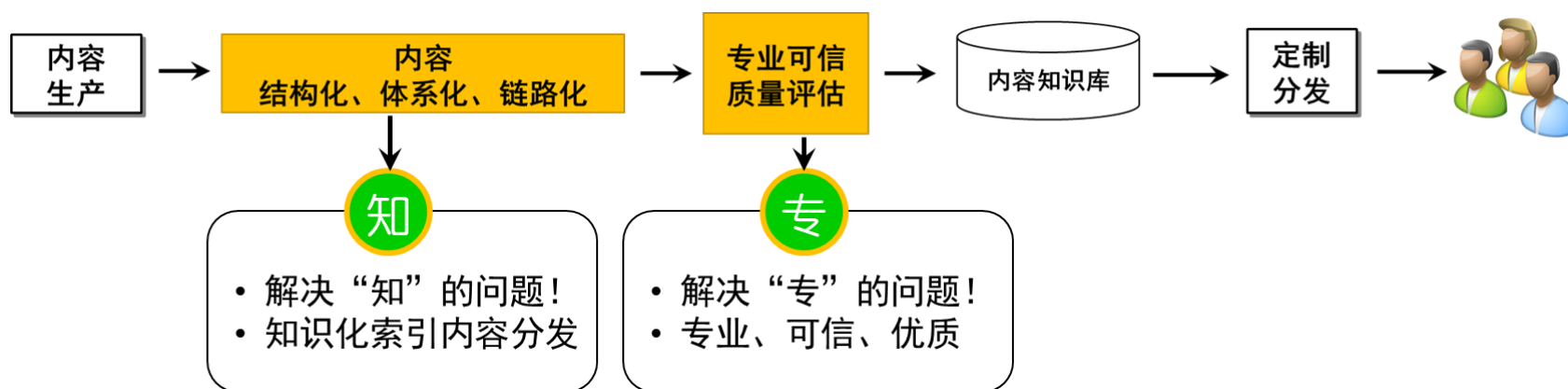
大数据时代- “专” 与 “知” 的问题

- 移动互联大数据时代，我们随时随地生产和消费内容的同时，伴随两大问题：
 - ✓ 问题-“**专**”：内容质量良莠不齐，专业可信度缺失
 - ✓ 问题-“**知**”：网络内容碎片化，不成知识体系，不利于用户学习和认知



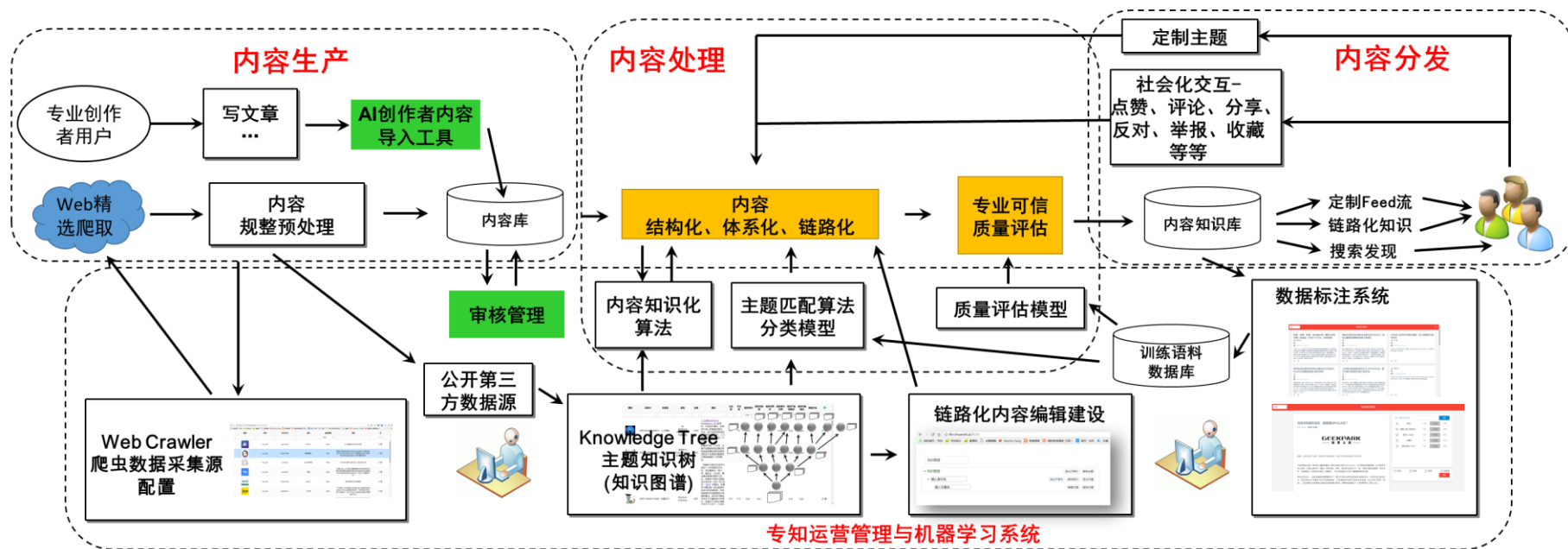
专知-基于知识图谱的内容分发

- 专知，连接人与知识，提供专业可信的知识分发服务：
 - ✓ 利用以知识图谱为核心技术解决“**专**”与“**知**”的问题，进行产品化，帮助用户又好又快找到所需知识
 - ✓ 专业可信的内容分发技术体系



专知技术体系架构

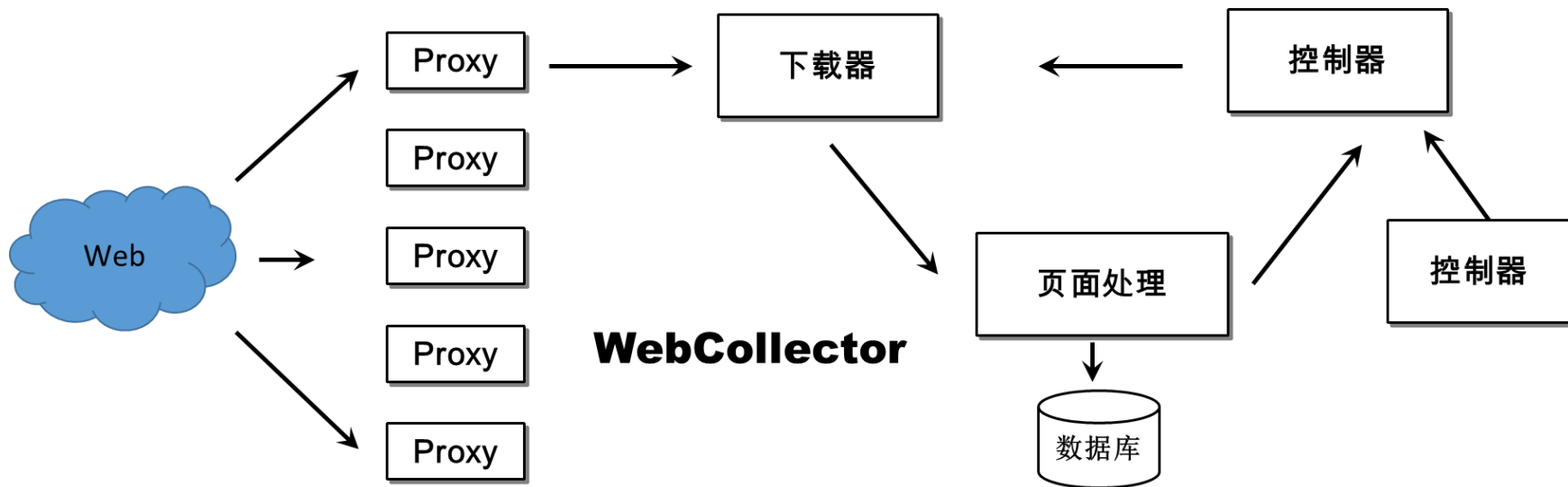
■ 内容生产、内容处理、内容分发



专知-内容生产

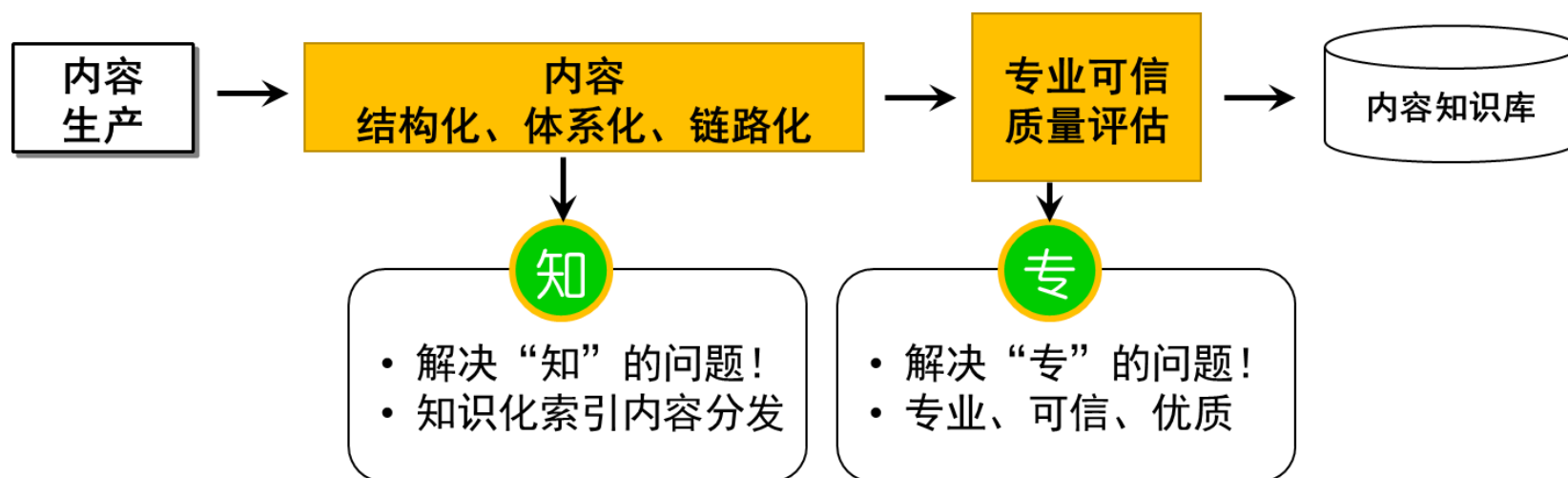
■ 实时全网定点内容采集系统

- ✓ 基于专知团队核心成员自主开发的开源爬虫WebCollector框架实现
- ✓ 采集站点源
 - UGC内容：微信公众号、知乎、简书、微博、Twitter等
 - 专业内容：arXiv、DBLP等论文



专知-内容处理

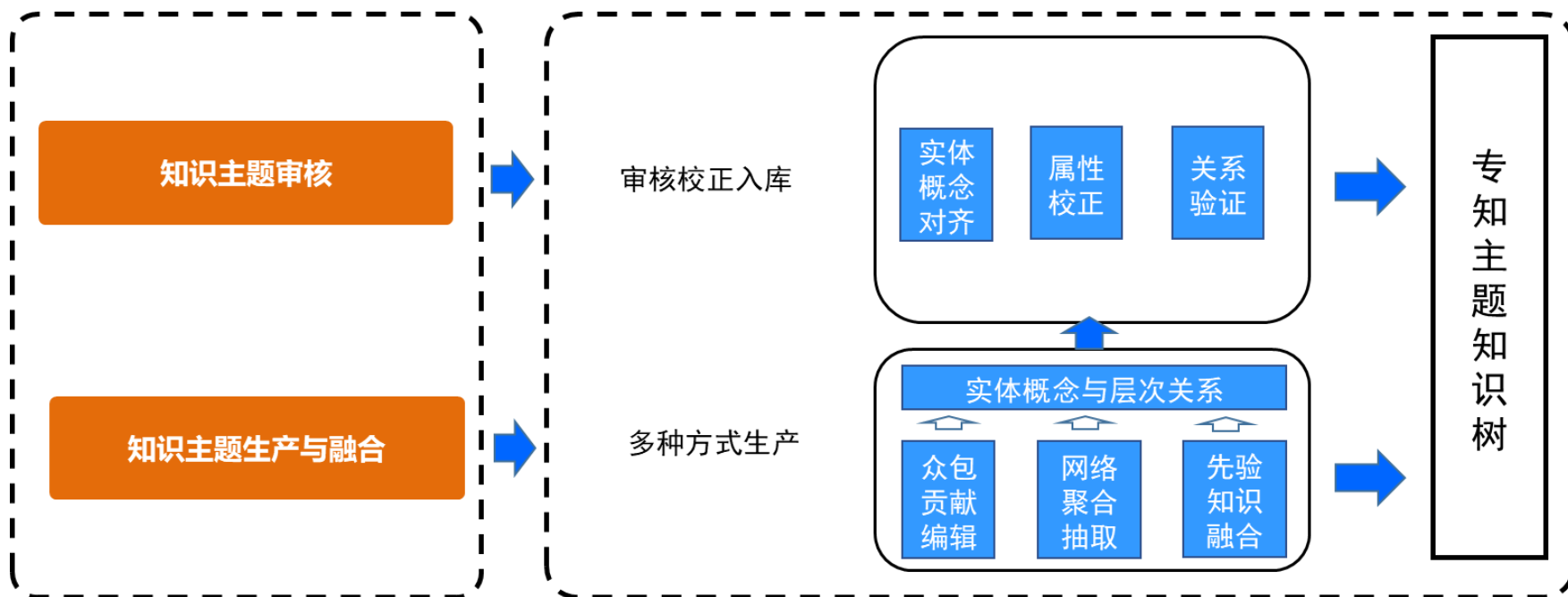
- 解决内容的“知”与“专”的问题，得到专业可信的知识
 - ✓ 基于**知识图谱**技术构建特定领域知识内容库，其主体架构是主题知识树，涵盖完整的特定领域知识体系结构
 - ✓ 以**知识计算**为核心技术算法框架，进行内容理解、用户建模和交互计算的研究与技术应用



专知-主题知识树

■ 特定领域的完整知识体系结构

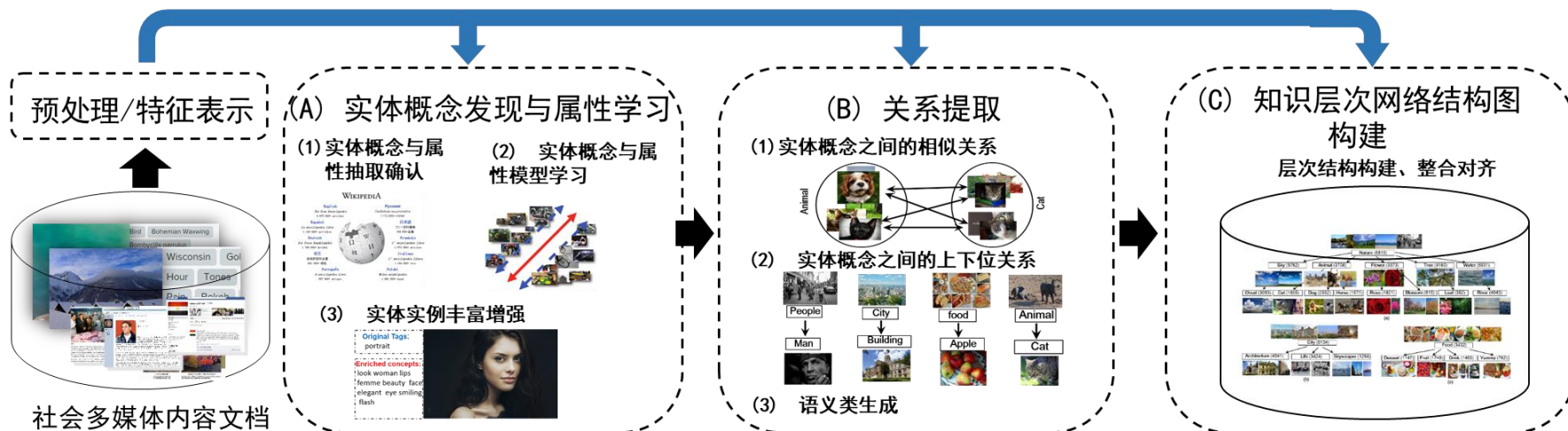
- ✓ 每个主题的描述、图标、父子上下位关系等信息
- ✓ 机器算法与人工审核协作进行



专知-主题知识树

■ 机器算法自动构建知识库体系结构

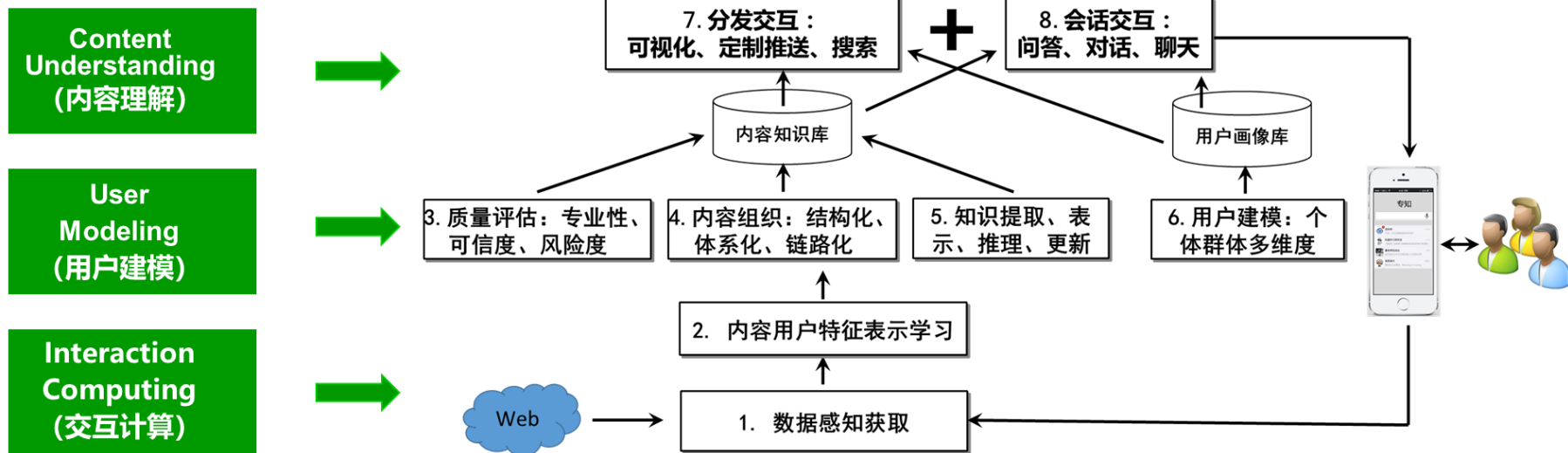
- ✓ 自动从用户产生的网络标签数据自动构建知识库体系结构
算法框架，**概念自动发现**、**概念关系提取**、**概念层次结构**
构建，有效辅助人工构建高质量可扩展的知识库



Fang et al. IEEE Trans. MM 2016

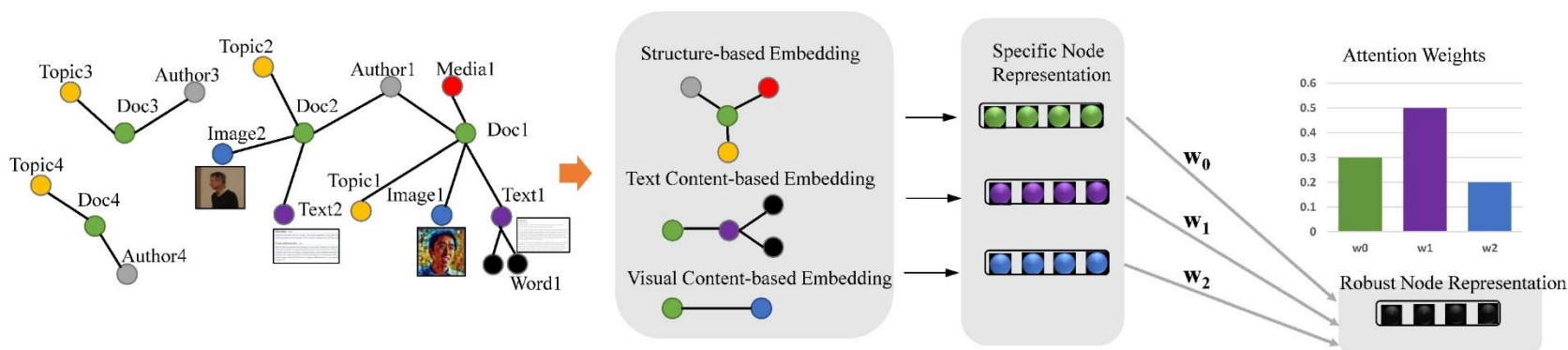
专知-知识计算

■ 专知内容处理核心算法框架



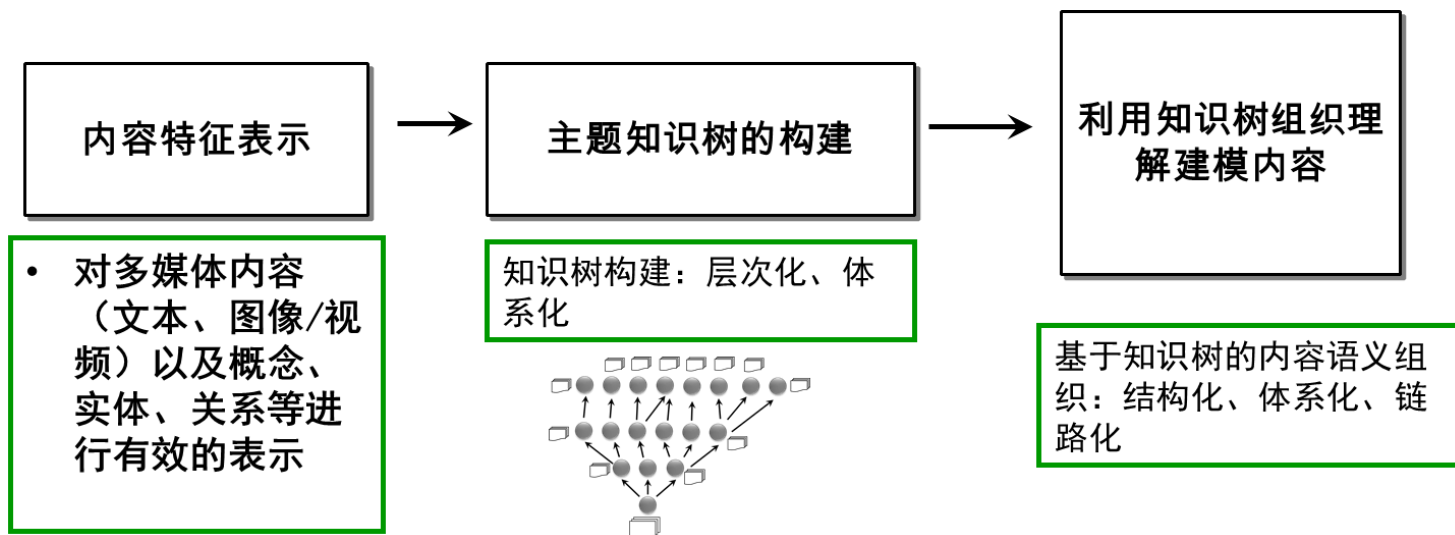
专知知识计算-用户内容特征表示

- 对内容和用户进行特征表示，支撑内容理解和用户建模
 - ✓ 符号表示法：基于字符串共现的稀疏表示
 - ✓ 分布式向量表示：低维稠密实值向量表示



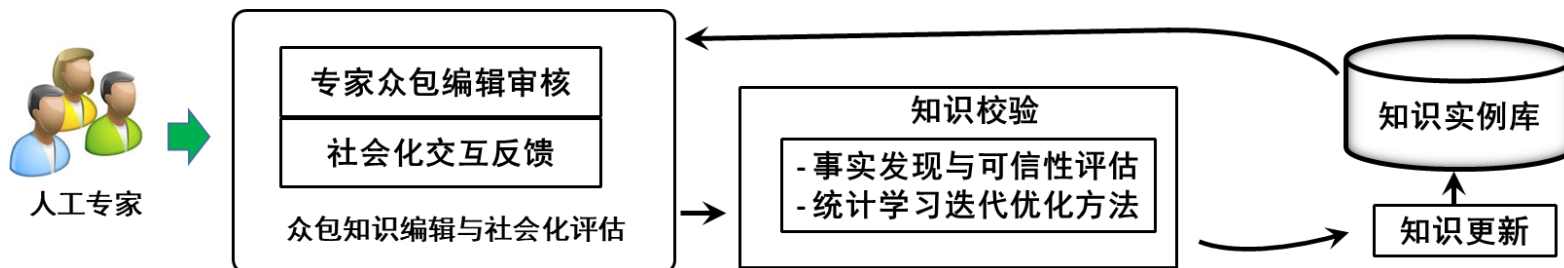
专知知识计算-内容“知”的处理

- 内容到知识: 结构化、体系化、链路化来理解分析组织内容，形成知识
 - ✓ 结构化: 文档内容语义结构性理解，产出文档主题、事件等结构性信息
 - ✓ 体系化: 对内容进行语义层次化体系性组织，知识体系化
 - ✓ 链路化: 对内容进行语义有序性组织，构成知识链路



专知知识计算-内容“专”的处理

- 内容知识专业可信性的评估
 - ✓ 专业性：在特定领域下内容知识的专业权威度
 - ✓ 可信性：内容知识的真实可信度
- 人工审核与机器算法学习协作完成
 - ✓ 通过引入人工专家的交互反馈信息，采用机器学习算法构建专业质量评估模型，来自动筛选专业可信的内容



专知-专业可信的知识分发

主题知识库



主题知识树
主题知识链路
主题知识荟萃
主题动态内容

主题链路



系统性知识



精品



专业



可信

主题定制



个性化定制资讯



媒体服务



个人用户

搜索发现



精准搜索 发现更多



个人用户



企业用户

交互评论



社会交互 可信评论



收藏



评论



点赞

创作分发



专业创作 高效分发



个人用户



合作团体



AI机构

提纲



专知介绍



专知技术



专知应用

军事资讯内容分发-专知<http://military.zhuanzhi.ai>

■ 专知

- ✓ **主题知识树**：构建了上千个主题
- ✓ **内容生产**：基于WebCollector构建了全网关于军事的站点做了精选采集
- ✓ **内容处理**：对军事内容进行结构化和专业质量评估
- ✓ **内容分发**：提供定制推荐、搜索发现的功能



军事资讯内容分发-专知<http://military.zhuanzhi.ai>

专知

首页

主题

发现

搜索



我的关注

全部关注

武器

枪

战争

朝鲜战争

现代战争

第一次世界大战

坦克 · 陆军 · 主战坦克 · 武器 · 骑兵 · 今天 12:13

泰国军方1月2次“炫耀”VT4坦克 高度评价其性能



凤凰网军事频道



凤凰网军事频道，是凤凰军事唯一官方公众号。掌握最新的防务新闻，解读最透彻的军事情报，为读者展现全方位的国...

泰国率先放下了此前国际上度中国顶级装备的成见，结果收获了一款性能出乎意料的优秀坦克。（资料图）泰国陆军26日



点赞 | 0

阅读 0+

评论 0+



枪 · 霰弹枪 · 特种部队 · 枪械 · 火箭筒 · 今天 12:13

独家|土耳其版战狼9.7高分 结果转眼就被现实惨重伤亡打脸



凤凰网军事频道



凤凰网军事频道，是凤凰军事唯一官方公众号。掌握最新的防务新闻，解读最透彻的军事情报，为读者展现全方位的国...

不少人将《山2》称为土耳其版的《战狼2》，但与《战狼2》不同的是《山2》很快被残酷的现实打脸。（资料图）凤凰



点赞 | 0

阅读 0+

评论 0+





■ 专知，专业可信的人工智能知识分发服务

- ✓ 针对AI知识资源分布不均衡状况和各行业AI人才需求，聚焦在人工智能领域，为**人工智能从业者提供专业可信的AI知识分发服务**
- ✓ **人工智能主题知识树**：构建了包含5千多个人工智能主题的知识库，涵盖人工智能、大数据分析、系统架构、编程语言等类目
- ✓ **内容生产**：基于WebCollector构建了实时定点精选采集内容生产系统，采集微信、微博、垂直站点、Arxiv等网站
- ✓ **内容处理**：对AI内容进行结构化和专业质量评估
- ✓ **内容分发**：为用户提供**定制推荐、搜索发现、链路学习**的服务



■ 专知，专业可信的人工智能知识分发服务

- ✓ 针对AI知识资源分布不均衡状况和各行业AI人才需求，聚焦在人工智能领域，为**人工智能从业者提供专业可信的AI知识分发服务**
- ✓ **人工智能主题知识树**：构建了包含5千多个人工智能主题的知识库，涵盖人工智能、大数据分析、系统架构、编程语言等类目
- ✓ **内容生产**：基于WebCollector构建了实时定点精选采集内容生产系统，采集微信、微博、垂直站点、Arxiv等网站
- ✓ **内容处理**：对AI内容进行结构化和专业质量评估
- ✓ **内容分发**：为用户提供**定制推荐、搜索发现、链路学习**的服务

人工智能知识分发-专知 www.zhuanzhi.ai



■ 专知，专业可信的人工智能知识分发服务



产品于2017年9月15号上线！

人工智能知识分发-专知 www.zhuanzhi.ai



■ 截止2018年1月30日



专知

专知平台
使用用户数

5400+

专知平台
网页日访问量级

2K+



专知微信
公众号用户数

19000+

专知微信
文章日阅读量级

1W

人工智能知识分发-专知 www.zhuanzhi.ai

■ 专知使用



PC访问 www.zhuanzhi.ai



微信扫一扫使用专知



人工智能知识分发-专知 www.zhuanzhi.ai



■ 专知公众号-第一时间获取人工智能资讯知识



人工智能知识分发-专知 www.zhuanzhi.ai



■ 定制主题，获取最新资讯知识

1、浏览器输入专知网址 www.zhuanzhi.ai

为人工智能从业者服务!

专知 zhuanzhi.ai，致力于为AI从业者提供一个优质纯粹、专业可信的AI知识分发服务平台！

注册 · 登录

邮箱/手机

fangquanyi@gmail.com 获取验证码

.....

注册 已有账号? 登录

2、注册登录专知

精选文章 推荐 人工智能 深度学习 机器学习 计算机视觉 自然语言处理

1、LDA基本介绍
★ 4+ jianwei wang, shengsheng qian · 最近

图像分类的概念
★ 2+ 专知出品 · 最近

本周荟萃主题

信息推荐
信息推荐，是指根据用户的习惯、...

机器翻译
机器翻译，又称为自动翻译，是利...

人工智能知识分发-专知 www.zhuanzhi.ai

■ 定制主题，获取最新资讯知识

www.zhuanzhi.ai/#/topics

3、关注定制你感兴趣的 主题

关注主题
请关注至少3个主题

个性化信息流
个性化信息流展示

开始探索优质内容

导引 链路 最新 最热 已关注 4 搜索主题

 人工智能

 大数据分析

 编程语言

 系统架构



人工智能

人工智能(Artificial Intelligence, AI)是研究、开发用于模拟、...

已关注

1241+ 人关注 9620+ 篇文章



自然语言处理

自然语言处理(Natural Language Processing, NLP)

已关注

1133+ 人关注 655+ 篇文章



计算机视觉

计算机视觉是一门研究如何使机器“看”的科学，更进一步的说...

已关注

1138+ 人关注 587+ 篇文章

人工智能知识分发-专知 www.zhuanzhi.ai

■ 定制主题，获取最新资讯知识

The screenshot displays the Zhuanzhi website interface. At the top, the URL www.zhuanzhi.ai/#/timeline is shown in the browser address bar. The navigation bar includes the '专知' logo, a '首页' (Home) button, and links for '主题' (Topics), '发现' (Discover), and '消息' (Messages). A search bar is located on the right side of the navigation bar.

On the left side, there is a sidebar titled '我的关注' (My关注的). It contains a list of topics with corresponding icons: '全部关注' (All关注的), '机器学习' (Machine Learning), '人工智能' (Artificial Intelligence), '计算机视觉' (Computer Vision), and '自然语言处理' (Natural Language Processing).

The main content area displays a timeline of articles. The first article is titled '52 个有用的机器学习与预测API' (52 Useful Machine Learning and Prediction APIs) by 北京思腾合力科技有限公司 (Beijing Siteng HeLi Technology Co., Ltd.). The article text mentions that with the application of AI and machine learning, many APIs with similar functions have emerged. The article has 0 likes, 0+ views, and 0+ comments.

The second article is titled '今年双 11 有点不同，「技术 SHOW」或成最大看点' (This year's Double 11 is a bit different, 'Tech SHOW' may be the biggest highlight) by EGO Networks. The article text mentions that EGO is a high-end technology community under the EGO brand, serving CTOs and technical VPs. The article has 0 likes, 0+ views, and 0+ comments.

4、获取最新资讯知识

人工智能知识分发-专知 www.zhuanzhi.ai

■ 定制主题，获取最新论文

专知

首页

主题

发现

搜索

资讯 ☒ 论文

我的关注

☐ 全部关注

机器学习

人工智能

深度学习

TensorFlow

计算机视觉

自然语言处理

机器学习 · 12月15日

Variance-based regularization with convex objectives

John Duchi, Hongseok Namkoong

We develop an approach to risk minimization and stochastic optimization that provides a convex surrogate for variance, allowing near-optimal and computationally efficient trading between approximation and estimation error. Our approach builds off of techniques for distributionally robust optimization and Owen's empirical likelihood, and we provide a number of finite-sample and asymptotic results characterizing the theoretical performance of the estimato...

点赞 | 0

下载PDF

阅读 2+

评论 0+

反向传播 · 机器学习 · 12月15日

Monotonic Chunkwise Attention

Chung-Cheng Chiu, Colin Raffel

nce-to-sequence models with soft attention have been successfully d to a wide variety of problems, but their decoding process incurs a atic time and space cost and is inapplicable to real-time sequence ransduction. To address these issues, we propose Monotonic Chunkwise Attention (MoChA), which adaptively splits the input sequence into small chunks over which soft attention is computed. We show that models utilizing...

点赞 | 1

下载PDF

阅读 1+

评论 0+

4、获取最新论文

www.zhuanzhi.ai



■ 主题链路，学习系统性知识



人工智能知识分发-专知 www.zhuanzhi.ai

■ 搜索发现，精准获取AI知识资料



谢谢各位专家和领导！



www.zhuanzhi.ai