

跨媒体智能

彭宇新

北京大学王选计算机研究所

2020-08-28

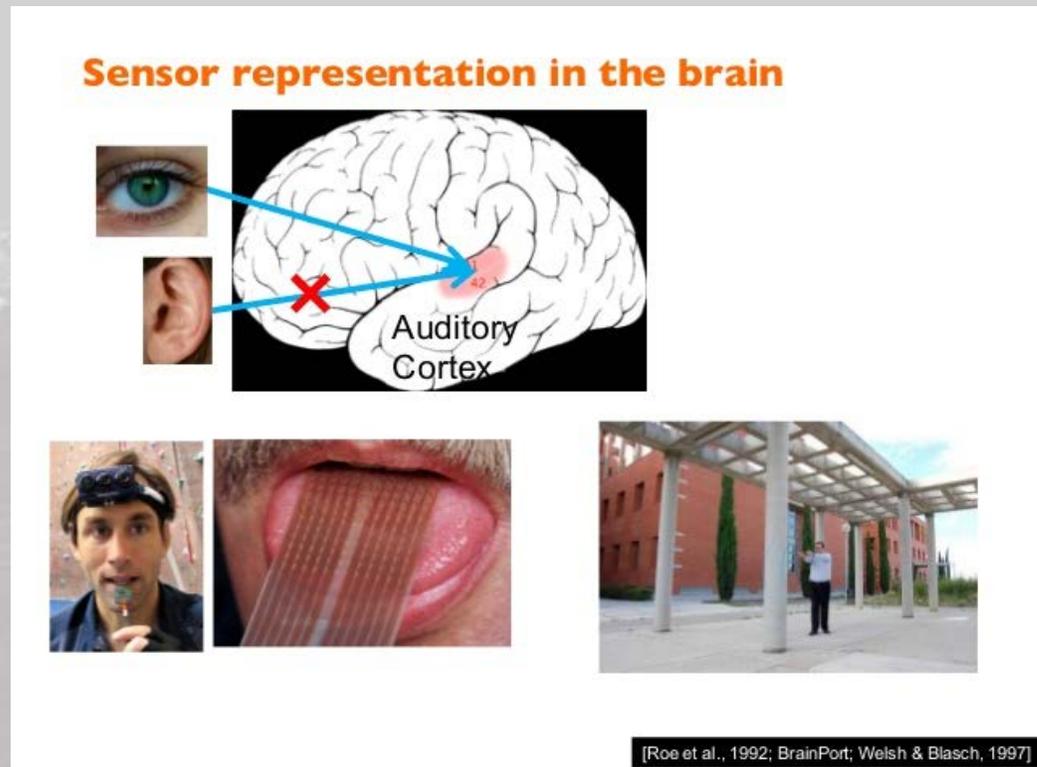
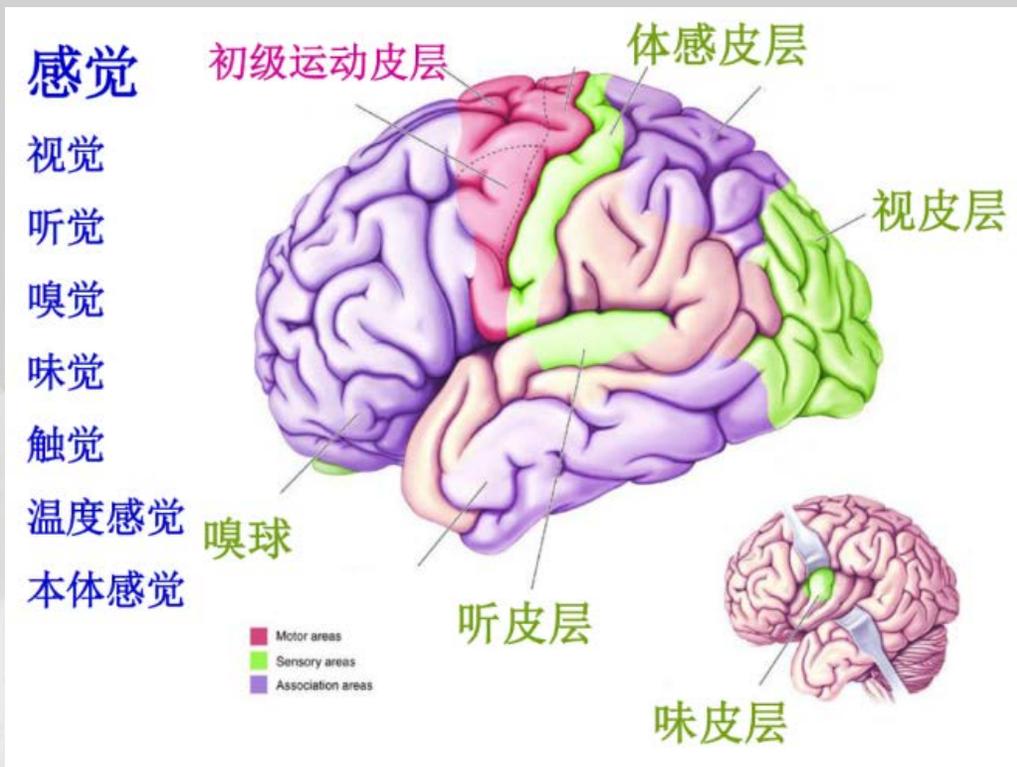
图像、视频、文本、音频等构成相互融合的**跨媒体**形态：**形式上多源异构，语义上相互关联**





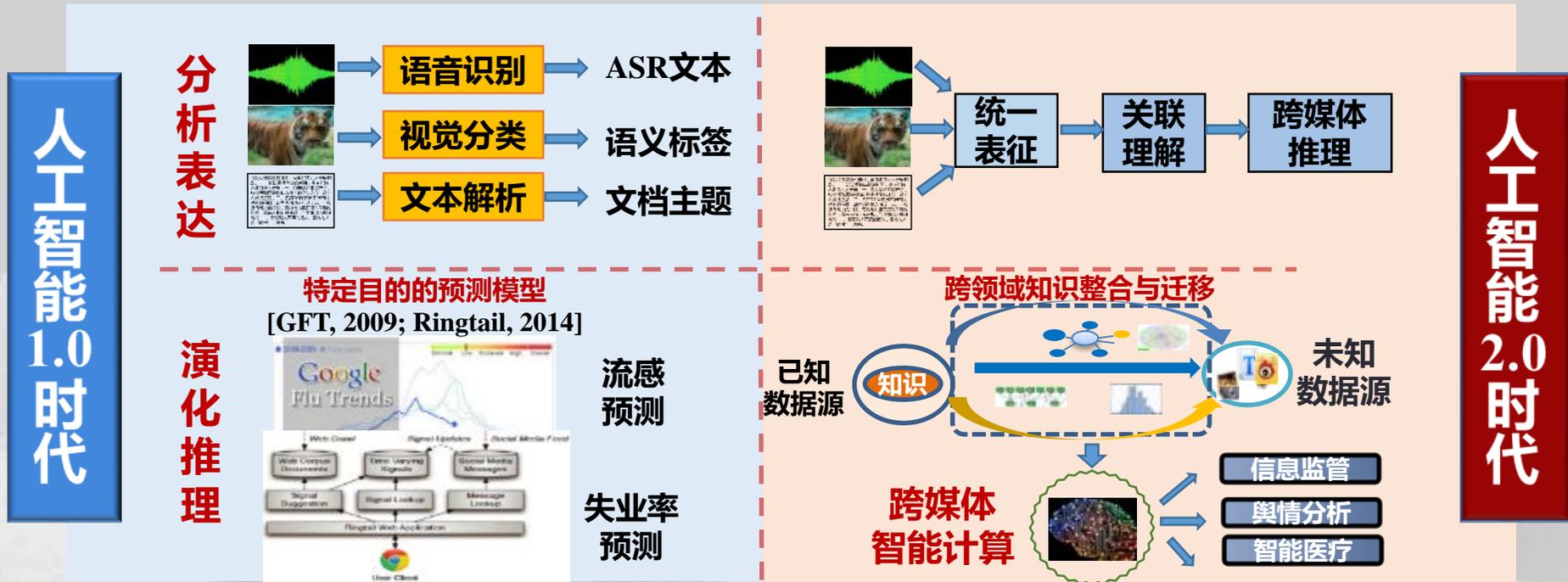
人脑认知的跨媒体特性

- 人脑组织结构决定了其对外界的认知过程是**跨越多种感官信息的融合处理**
- 传统人工智能以文本推理为主，而实际上知识及推理通常会有**语言、视觉和听觉等信息的共同参与**



跨媒体智能 vs 传统人工智能

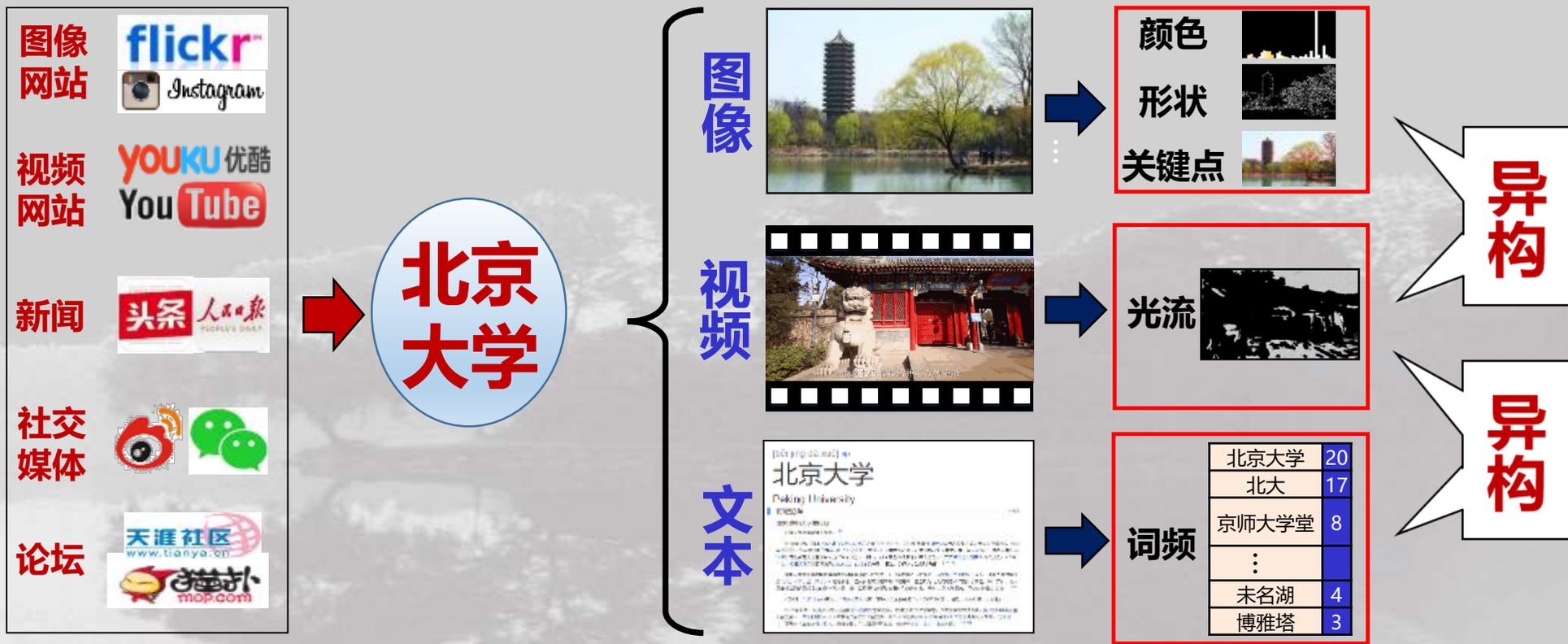
- **AI 1.0: 单媒体分析与文本推理**, 因为信息有限难以实现语义抽取, 无法应对跨媒体大数据的复杂分析和推理
- **AI 2.0: 跨越语言、视觉、听觉**等不同类型的媒体数据, 对现实世界中的知识进行更加泛化的分析与推理





• 异构鸿沟

- 视觉、听觉、语言等不同模态信息的特征表示**不一致**，如何实现多模态信息的**统一表征和综合利用**？



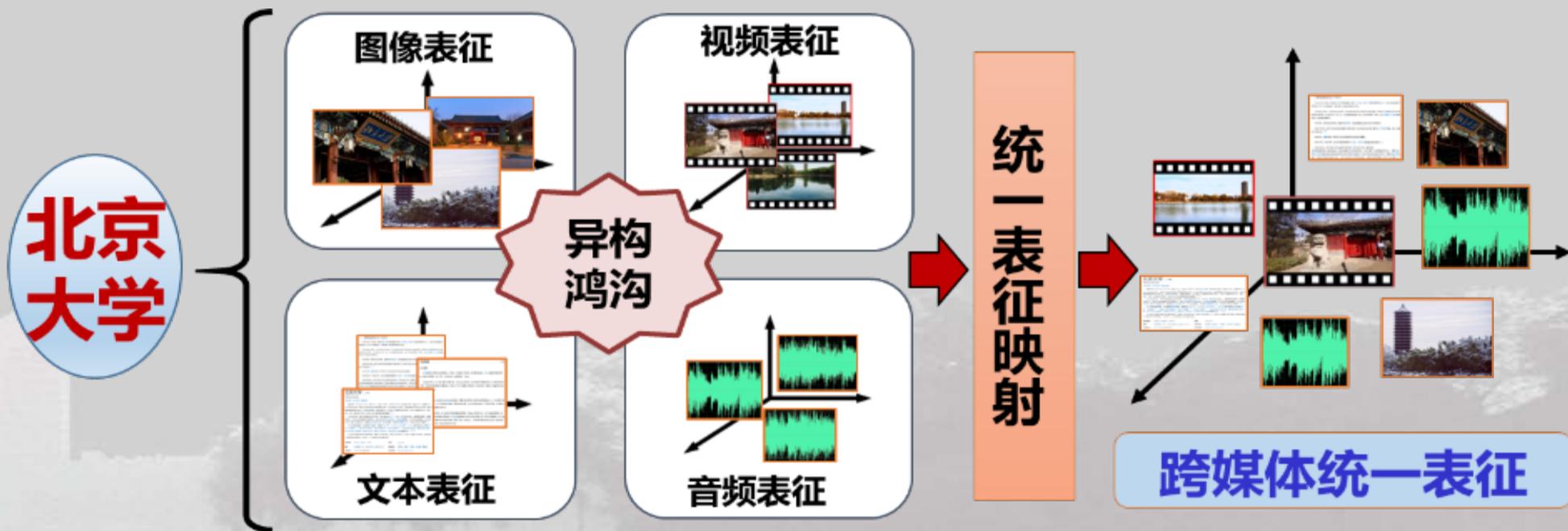


- **跨媒体智能——数据驱动与知识驱动相结合**
 - 通过**数据驱动**方法挖掘跨媒体大数据中潜在的模式与规律
 - 通过**知识驱动**的跨媒体协同推理，降低跨媒体认知决策的不确定性



将传统信息处理和人工智能结合起来，达到**数据-知识协同驱动**的跨媒体智能分析与理解

- 跨媒体统一表征理论和模型



通过**统一表征映射**，将表达相似语义的跨媒体数据映射到同一个空间中，转换为相似的统一表征



主要任务-1

任意一种媒体样例



图像



图像



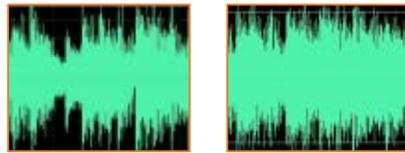
视频



文本



音频



图形



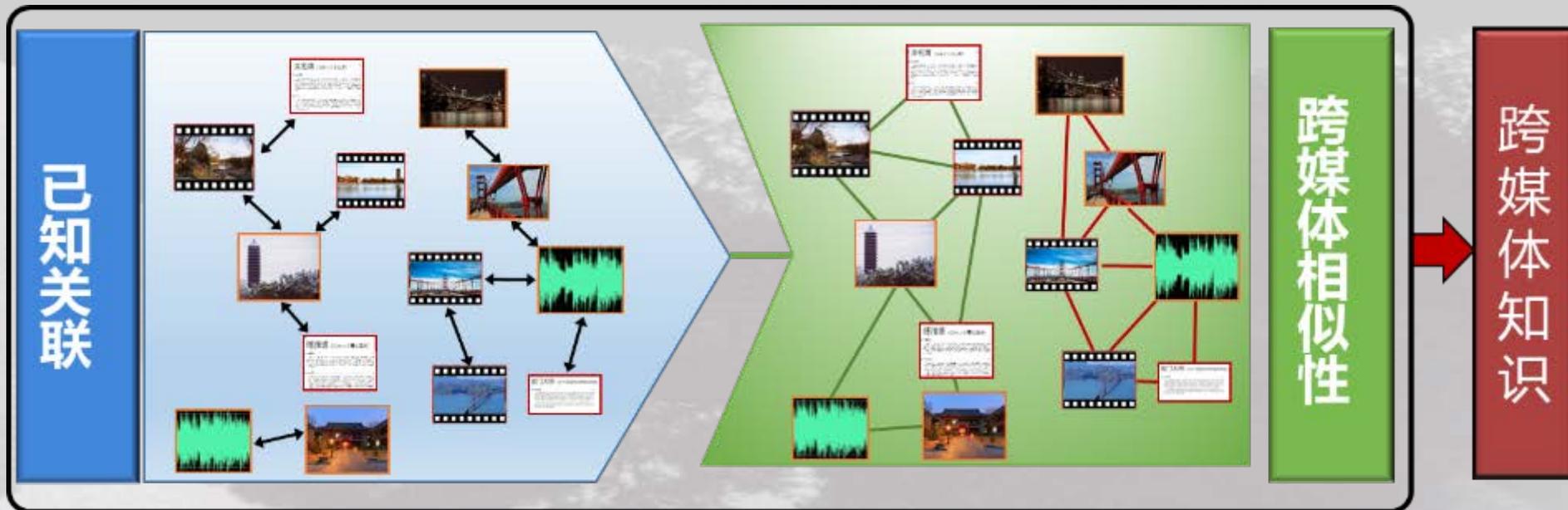
各种媒体检索结果

难点：不同媒体数据表征不一致，导致跨媒体相似性无法度量



• 跨媒体关联理解与深度挖掘

- 挖掘跨媒体知识，补充和拓展传统的基于文本的知识体系，研究跨媒体数据关联与融合方法
- 如何深入理解跨媒体数据关联，实现**相似性计算与知识挖掘**，是当前研究的重要挑战



• 跨媒体知识图谱构建与学习

- 当前跨媒体分析方法以**数据驱动**为主，可解释性与可泛化能力不足，需研究知识驱动的跨媒体分析方法
- 如何构建**跨媒体知识图谱**，形成跨媒体知识表达与学习的方法体系，为知识驱动方法提供重要依托？



跨媒体知识演化与推理

- 现有人工智能以**文本为主**，通过谓词、命题和规则等方法在充分定义前提下进行推理，无法应对跨媒体推理问题
- 如何将文本推理方式扩展到多种媒体协同，实现**跨媒体知识演化与推理**？





• 跨媒体描述与生成

- 针对视频、图像、音频、文本等数据，实现符合人类认知与表达的准确描述
- 当前的研究主要围绕**单一媒体内容的自然语言描述**，还需在人类认知优化建模等方面提高与完善，实现符合人类认知的**跨媒体内容的智能描述**



描述生成



Two men talk while standing next to a small passenger plane at an airport.



描述生成



Woman with short haircut sitting on brown couch is holding a baby while another child watches.



用户输入文本

输入文本

一只具有黄色腹部的棕色的鸟

13/200

生成



生成图像信息 [\[详细\]](#)

[Generate_0025-Queyr1-8.jpg](#)

尺寸: 224x224

大小: 10.5KB

生成所用时间: 29.2毫秒

基于用户输入文本
的生成图像

更多生成图像 [\[上一页\]](#) [\[下一页\]](#)





用户输入文本

输入文本

这只鸟的翅膀是蓝色的，有一个小的鸟喙

18/200

生成



生成图像信息 [\[详细\]](#)

[Generate_0005-Queyr1-1.jpg](#)

尺寸: 224x224

大小: 7.67KB

生成所用时间: 23.0毫秒

基于用户输入文本的生成图像

更多生成图像 [\[上一页\]](#) [\[下一页\]](#)





用户输入文本

输入文本

一朵具有黄色花瓣的花

10/200

生成



生成图像信息 [\[详细\]](#)

[Generate_4062-Queyr5-2.jpg](#)

尺寸: 224x224

大小: 10.4KB

生成所用时间: 28.5毫秒

基于用户输入文本
的生成图像

更多生成图像 [\[上一页\]](#) [\[下一页\]](#)



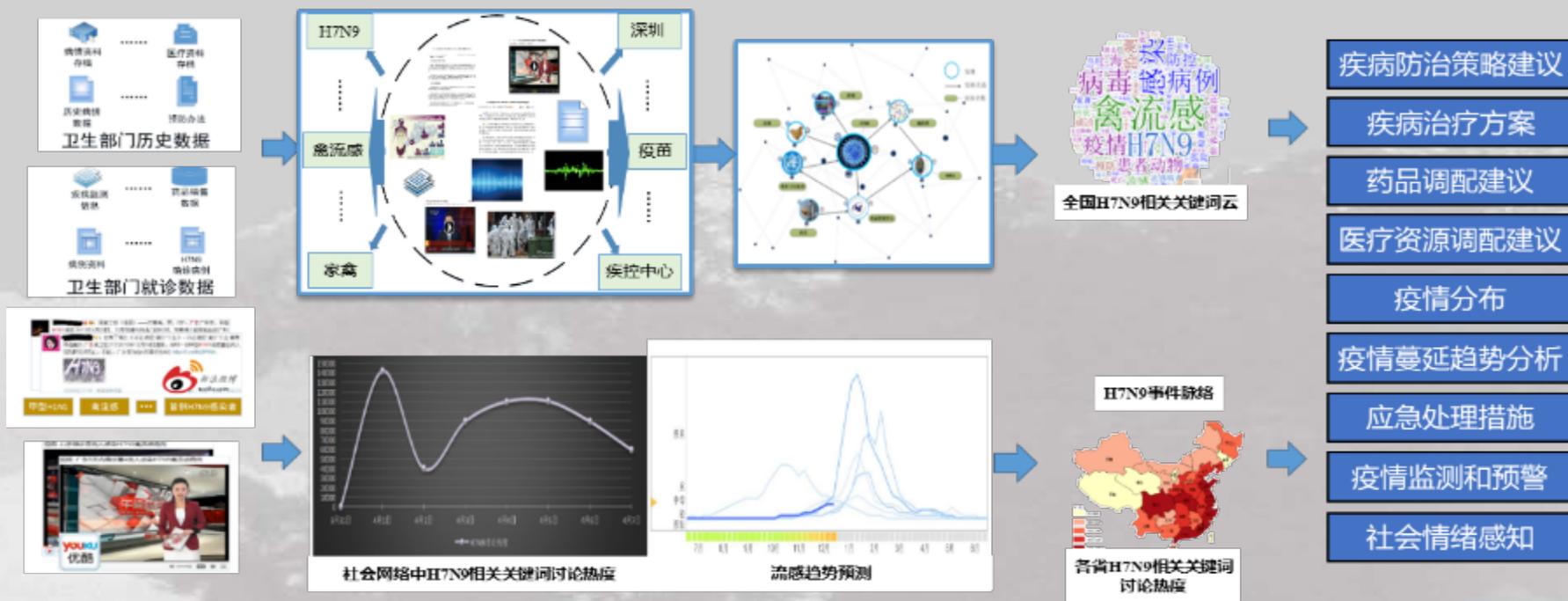
跨媒体智能引擎

- 现有系统：以**文本**为核心，以**信息检索**为主要需求，以**专用目的**为开发目标
- 如何形成高效的**通用跨媒体智能引擎**，是跨媒体智能走向应用的关键所在



跨媒体智能应用

跨媒体智能医疗：跨媒体大数据的充分利用能够支持医疗分析、临床决策、医疗管理等重要应用。跨媒体分析与理解能够为上述应用提供关键技术支撑



Yuxin Peng, Wenwu Zhu, Yao Zhao, Changsheng Xu, Qingming Huang, Hanqing Lu, Qinghua Zheng, Tiejun Huang, and Wen Gao, "Cross-media analysis and reasoning: advances and directions", *Frontiers of Information Technology & Electronic Engineering (FITEE)*, 18(1):44-57, Jan 2017.

Yuxin Peng, Xin Huang, and Yunzhen Zhao, "An Overview of Cross-media Retrieval: Concepts, Methodologies, Benchmarks and Challenges", *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 28(9): 2372-2385, Sep. 2018.

**跨媒体数据集PKU XMedia、XMediaNet、FG-Xmedia
和相关源代码下载地址，欢迎下载使用！
<http://www.wict.pku.edu.cn/mipl/XMedia>**





1. 跨媒体统一表征需要提高，目标是能实际应用
2. 跨媒体生成 (X2Y) 是新兴热点方向
3. 前沿探索：跨媒体知识图谱、数据-知识协同驱动
4. 核心：跨媒体推理（真正走向人工智能）
5. 跨媒体落地应用是跨媒体智能研究的亮点
6. 跨媒体智能与其他智能的融合与借鉴：大数据智能、混合智能等



研究室主页



GitHub主页

邮箱: pengyuxin@pku.edu.cn

电话: 010-82529699

研究室网址: <http://www.wict.pku.edu.cn/mipl/>