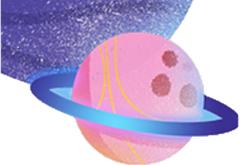


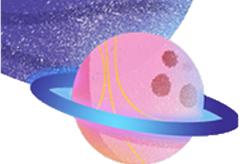
快知——多模态短视频 百科知识图谱

付瑞吉 快手 MMU知识图谱中心





- ▶ **1. 背景**
- 2. 快知体系
- 3. 构建技术
- 4. 应用与落地
- 5. 未来工作



背景：什么是知识？



What 事实知识

指客观存在的事实构成的知识。如：北京市的人口是多少？太阳和地球的距离是多少？

Why 原理知识

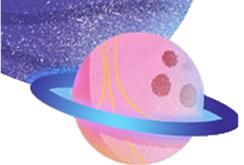
是指科学原理、自然规律等。如：地球为什么绕着太阳转？天空为什么是蓝色的？内燃机的工作原理？

How 技能知识

是指做完成某件事的技巧或能力。如：如何做鱼香肉丝？如何养多肉植物？海姆立克急救法怎么操作？

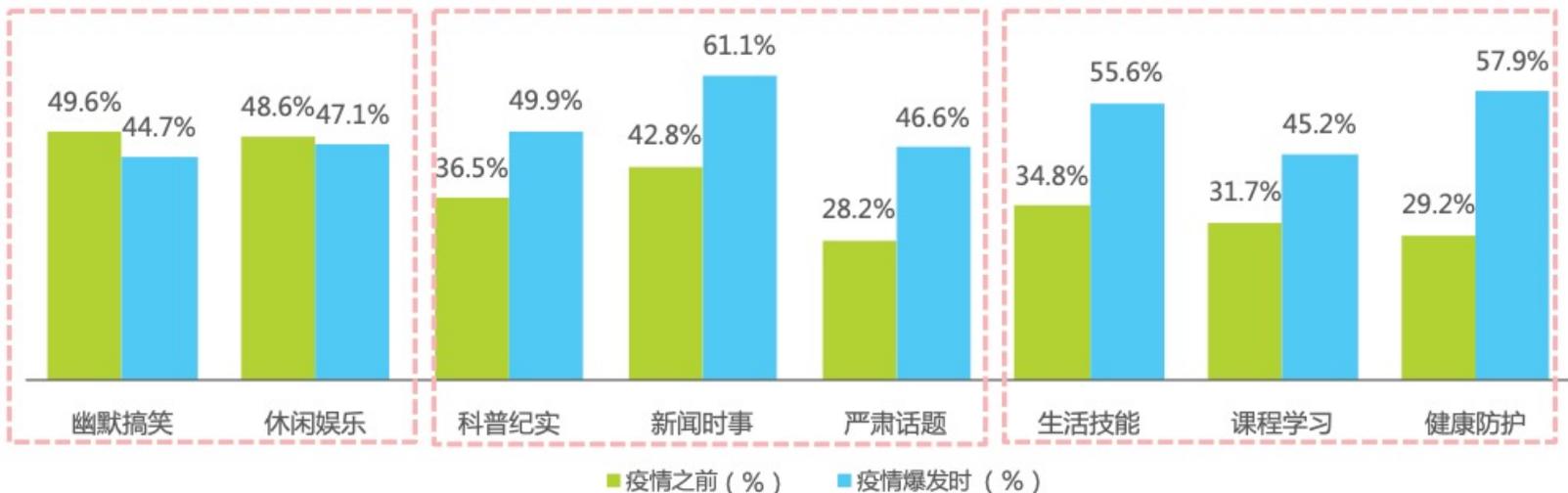
Who 人际知识

是指某个知识方向的专家（或团队）等信息。如：腰椎间盘突出治疗哪位医生比较靠谱？谁是养蜂高手？



背景：泛知识内容机会

2020年疫情前后用户内容题材的偏好变化



娱乐向

平均降幅
-3.2%

严肃向

平均增幅
16.7%

生活向

平均增幅
21.0%



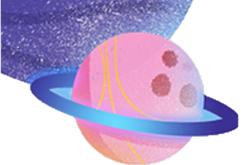
知识的关注度提升

疫情后，观众对于科普纪实、新闻事实等严肃向内容的关注，平均增幅高达**16.7%**



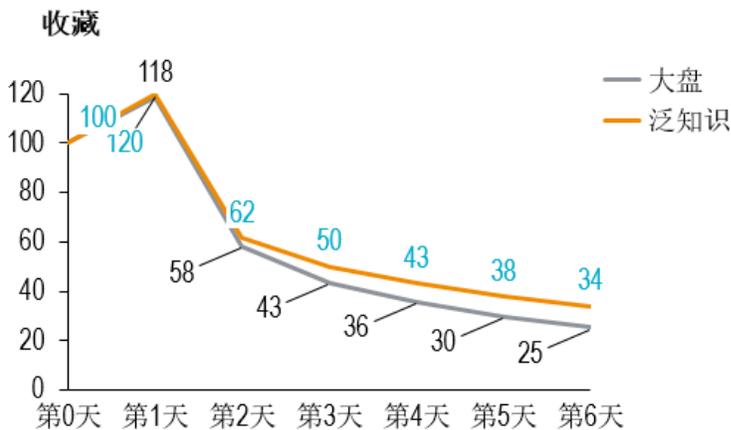
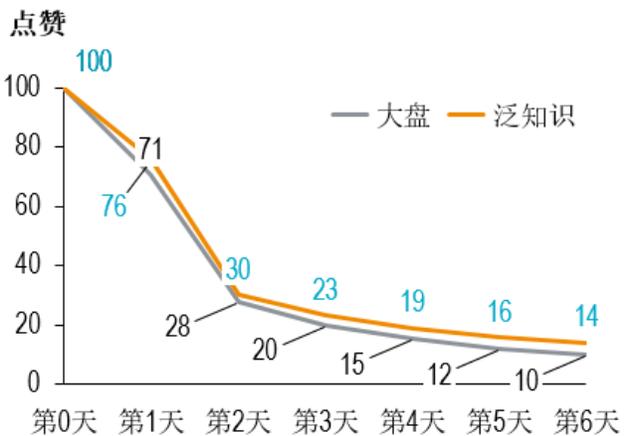
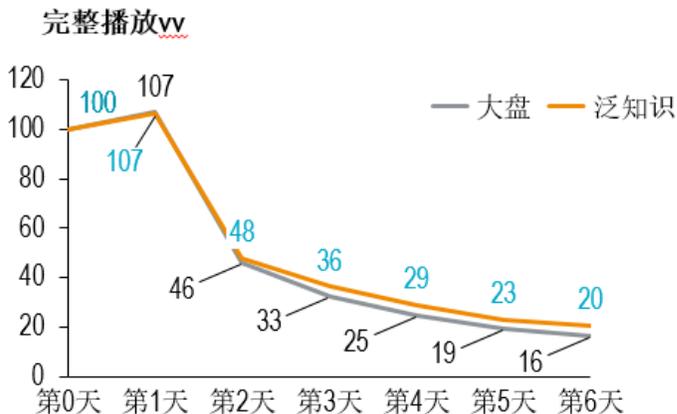
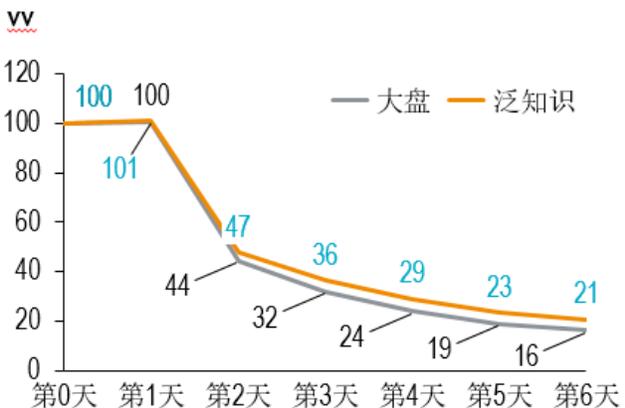
国家鼓励知识创作

2021年6月，国务院印发《全民科学素质行动规划纲要(2021-2035年)》，聚焦深化科普供给侧改革，系统提出了“科技资源科普化”以及“科普信息化”工程。



背景：泛知识内容机会

作品生命周期[以上传后第0天的消费指标为100%]



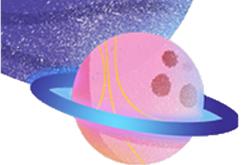
各平台布局泛知识

各大视频平台纷纷在知识类视频方向布局，充实其泛知识内容，此赛道未来的竞争将会越来越激烈。



知识类生命周期长

知识类视频，从七天内播放量、点赞等变化可以看出，比起泛娱乐内容，它的生命周期更长



背景：百科的发展



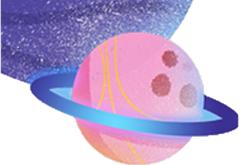
专家知识



纸质百科

图文网络百科

短视频百科?



背景：图文与短视频

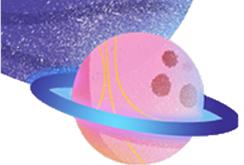
首先我们画个长方形
 然后在他上面画出他的耳朵
 这里我们拿出针管笔
 按照刚才的铅笔形状
 给他勾一下线
 勾完线准备
 米黄、赭石还有橙黄色
 拿大一点的笔
 先铺上一层清水
 然后用米黄色把它全画满
 趁他没干
 用橙黄色画出他的轮廓
 然后我们用赭石色
 从外到内加深他的颜色
 用点染的方法
 将深棕色点进去



文字

图文

短视频



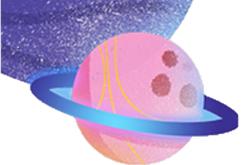
1. 背景

▶ **2. 快知体系**

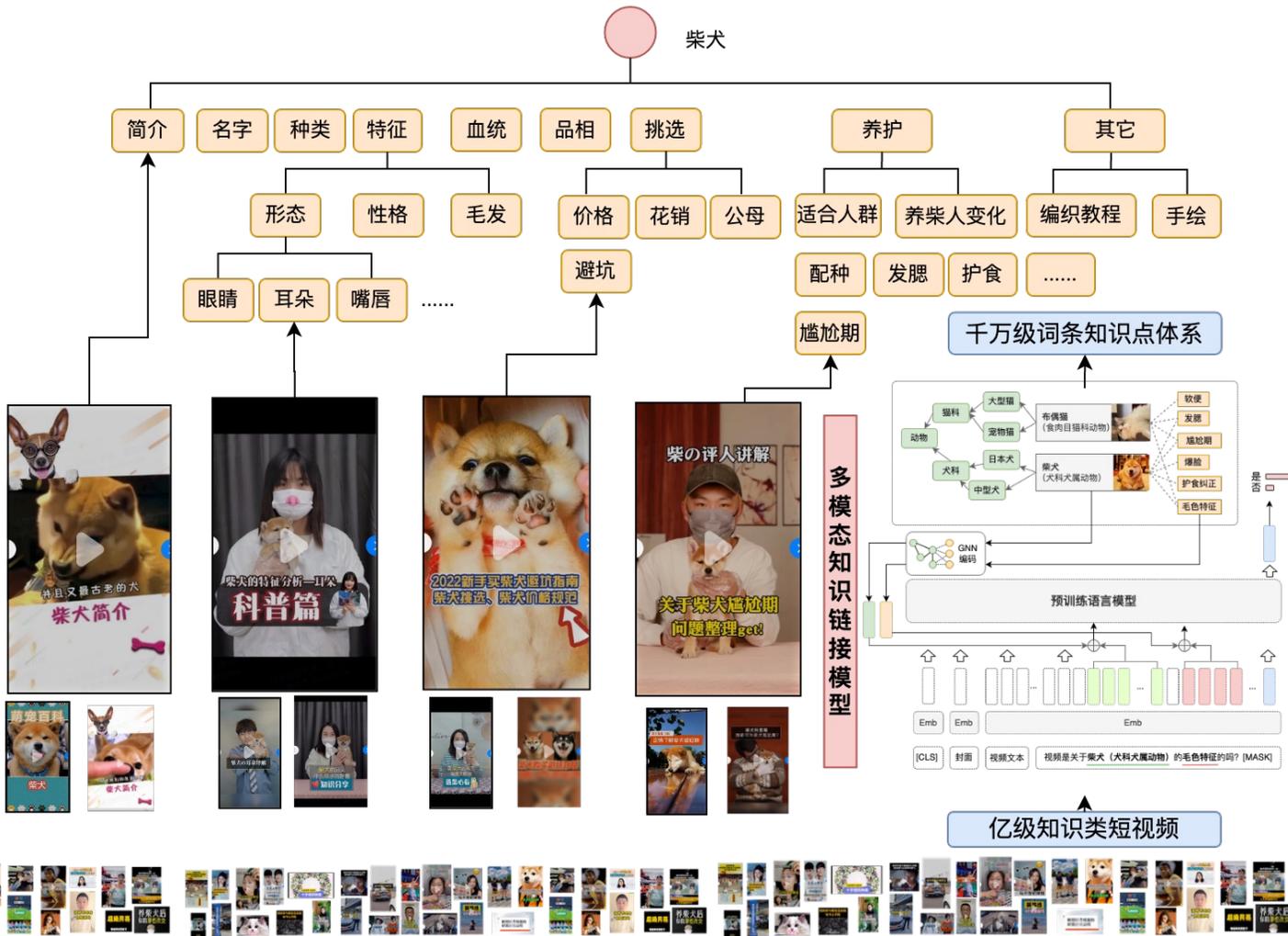
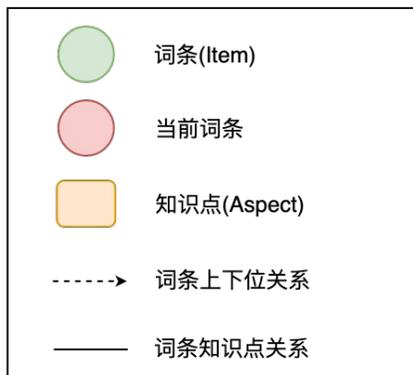
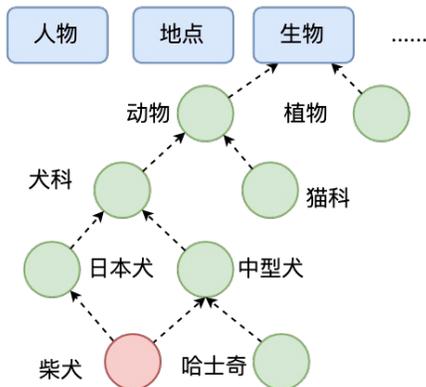
3. 构建技术

4. 应用与落地

5. 未来工作



快知体系 (Kuaipedia)



词条 (Item)

将概念（如植物）与实体（如玫瑰花）的集合统一成为一个词条

知识点 (Aspect)

指概念或实体对应某个方面，如 玫瑰花-扦插, 植物-历史等

短视频 (Video)

在这里主要是指传达 what, how-to, why等知识的短视频

快知

快手短视频百科全书

请输入关键字搜索实体



可以试试：老虎的相关文化

清华大学

蜡笔小新

任天堂

清明节的节令食品

换一换

关于快知 (当前数据截止时间2022-08-13)

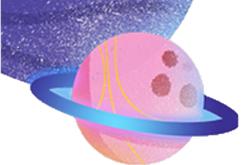
了解更多 >

知识推荐 ⓘ

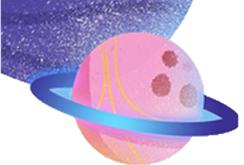
作者排行榜 ⓘ



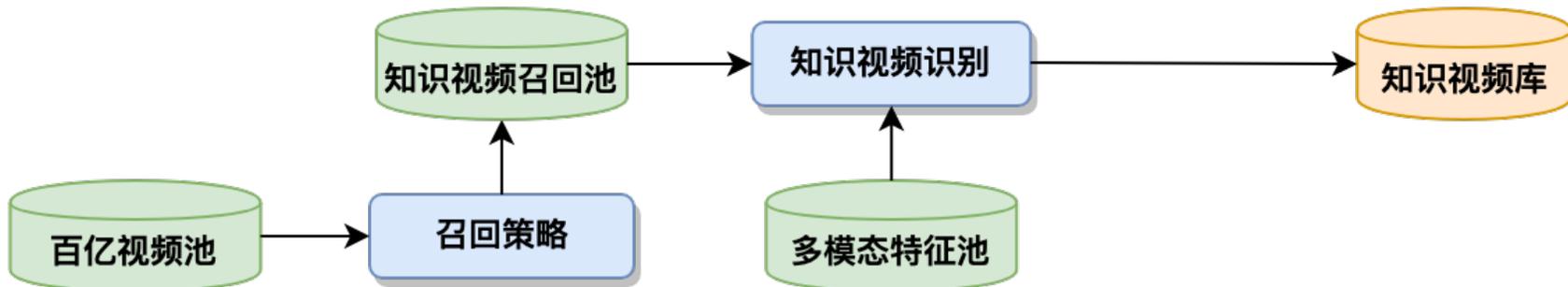
< 财经 法律 军事 科技 人文 >



1. 背景
2. 快知体系
- ▶ **3. 构建技术**
4. 应用与落地
5. 未来工作



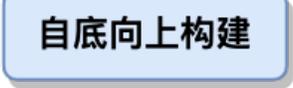
1. 知识视频识别



2. 知识体系构建

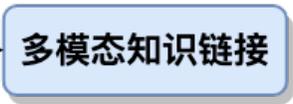


初始化

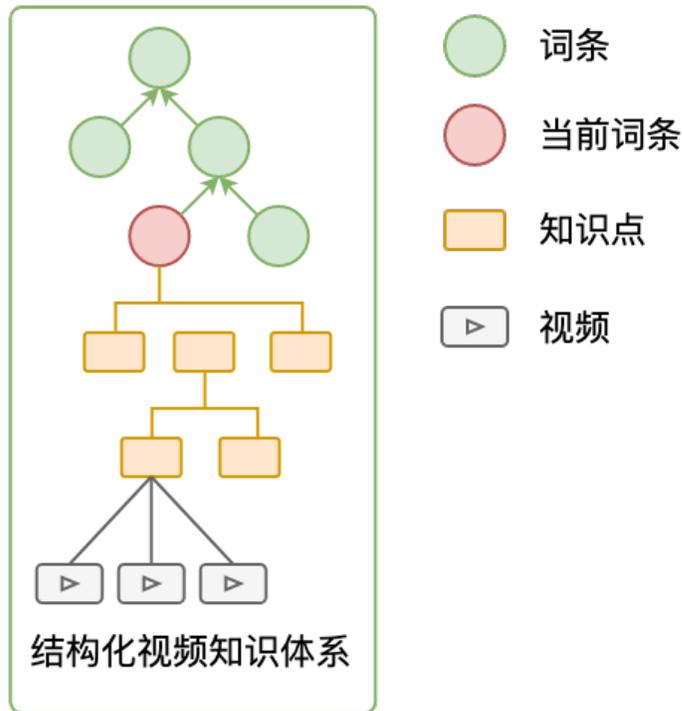


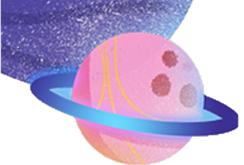
增补

3. 视频知识点挂载



挂载





技术点1：知识视频识别

挑战：知识的定义和识别难度大，考察对不同垂类不同知识形式的认知水平，而不是简单的识别能力

垂类知识示例

运动



垂类标准 (节选)

科普：运动相关的知识，包含各式各样的运动知识，如介绍运动相关人物、规则、小技巧等。

教程：视频是运动教程内容，要求体现具体做法、方案或步骤，或实操演示、亲身示范等。

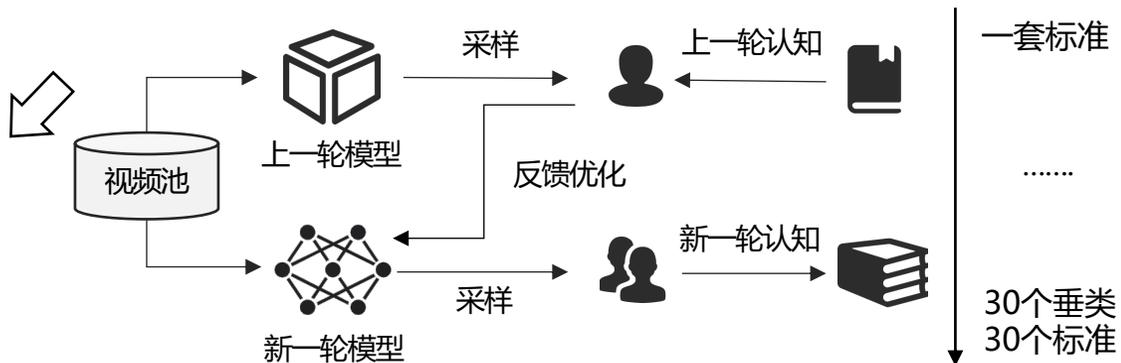
美食



科普：美食相关的知识，起源、历史、功效、搭配、冷知识等，非美食制作演示

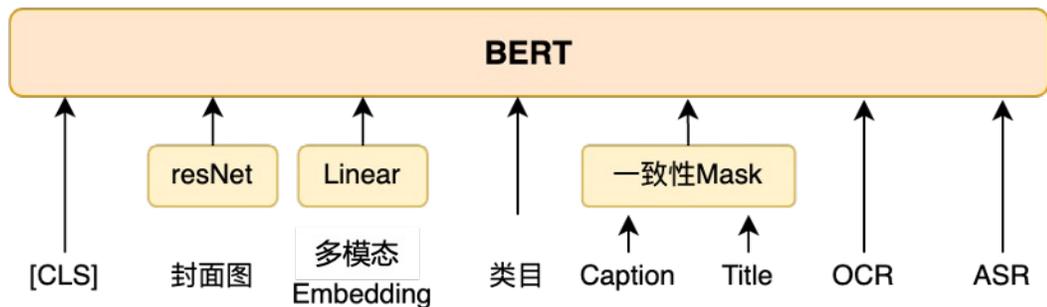
教程：烹饪教程或相关些制作教程，例如：发面；或者是食材的采摘、清洗、制作等教程。

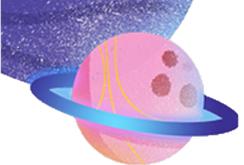
人机结合-认知迭代



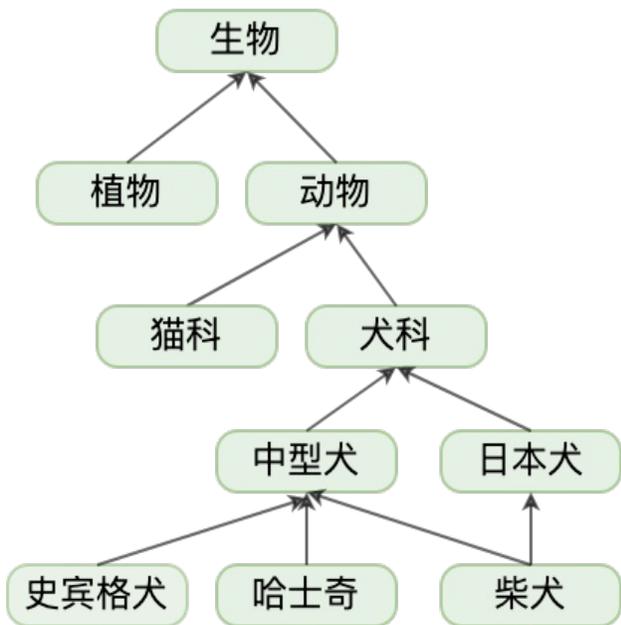
多模态知识视频识别模型

非知识 科普 教程





技术点2：知识体系构建 (自顶向下)



目录页数据发现



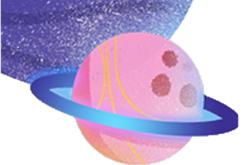
柴犬是体型中等并且又最古老的日本犬。柴犬能够应付陡峭的丘陵和山脉的斜坡，拥有灵敏的感官，使得柴犬屡次成为上乘的狩猎犬。柴犬性格活泼、好动。对自己喜欢的玩具...

体型 中型犬

别名 柴犬, doge

分类 [脊索动物门](#)^o、[哺乳纲](#)^o、[食肉目](#)^o、[犬科](#)^o、[犬属](#)^o

- 发展历史
- 名字由来
- 形态特征
- 性格特征



技术点2：知识体系构建 (自底向上)

泛知识内容结构化难度高：通用的知识体系，不完全适合快手特有的知识

我与我家柴犬 小柴毛色
 柴犬怎么纠正护食 柴犬毛色
 柴犬护食怎么办 柴犬毛发特征
 晒晒我家柴犬 阴阳师柴犬
 养柴千日用柴一时

输入：原始Query文本



小柴毛色
 柴犬怎么纠正护食 柴犬怎么毛
 柴犬护食怎么办 柴犬毛发特征

1. 知识意图Query识别



柴犬怎么纠正护食	柴犬	纠正护食
小柴毛色	小柴	毛色
柴犬护食怎么办	柴犬	护食
柴犬怎么毛	柴犬	怎么毛
柴犬毛发特征	柴犬	毛发特征

2. 词条与知识点识别



柴犬 (犬科犬属动物)	纠正护食
	毛色
	护食
	怎么毛
	毛发特征

3. 词条实体消歧



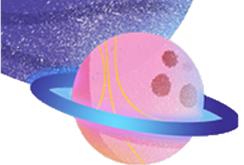
简介	名字由来	性格特征
毛色特征	毛发特征	毛色
护食	纠正护食	怎么毛

4. 知识点消重



柴犬 (犬科犬属动物)		
简介	名字由来	性格特征
毛色特征	护食	怎么毛

输出：合理&相关的知识点



技术点2：知识体系构建

蜜蜂

昆虫纲膜翅目蜜蜂科动物的统称

编辑

蜜蜂是一种群居昆虫，由蜂王、雄蜂、工蜂等个体组成。

知识点	○ 主要分布	○ 管理*	○ 过冬*	○ 幼虫*
	○ 分工特点	○ 采蜜*	○ 死亡*	○ 盗蜂*
	○ 信息交流	○ 过箱*	○ 合群*	○ 蜂巢*
	○ 繁殖方法	○ 春繁*	○ 打架*	○ 蜂药*
	○ 养殖技术	○ 种类*	○ 喂水*	○ 换王*
	○ 分蜂*	○ 造脾*	○ 筑巢*	○ 寿命*

含“*”知识点来源于快手的特色知识点

情感

健康

三农

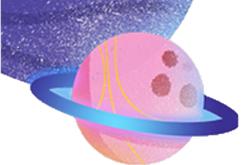
美食

游戏

美妆

财经

.....



技术点3：视频知识点挂载



挑战：视频知识点挂载对词条的消歧要求高，且要筛选出最核心词条和知识点

多模态知识链接

将短视频链接到（词条，知识点）

- 1. 实体识别** 识别候选实体（词条）Mention词
- 2. 实体消歧** 根据上下文对实体（词条）进行消歧
- 3. 知识点排序** 对视频对应的知识点排序链接

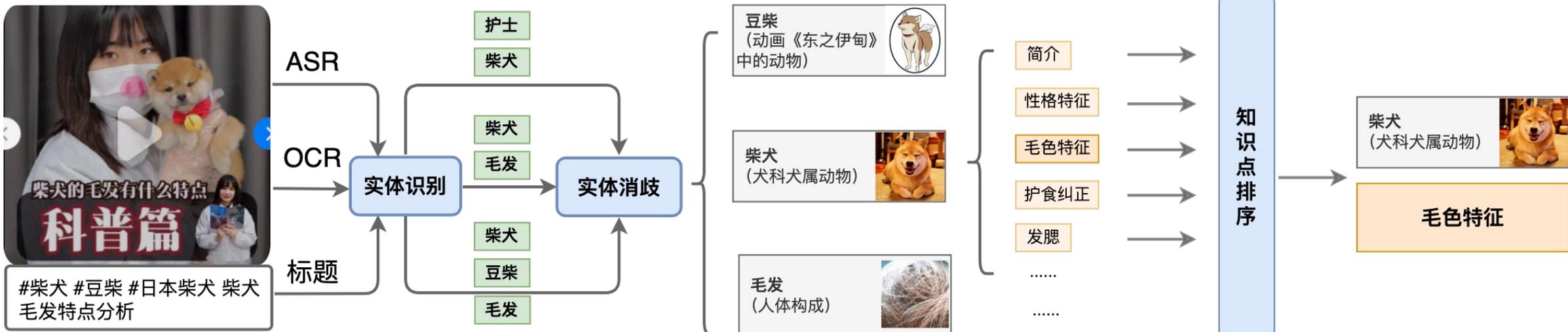


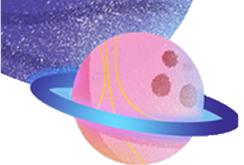
(1) 消歧要求高



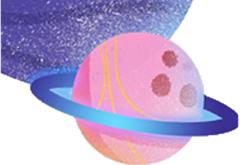
- 四轮定位
- 步骤
- 车轮定位
- 四轮定位
- 烟台
- 修车
- ...
- 50个词条
上千知识点

(2) 核心知识点排序难





1. 背景
2. 快知体系
3. 构建方式
- ▶ 4. **应用与落地**
5. 未来工作



应用与落地



快手医典

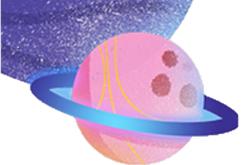


改善前



改善后

内容分发质量优化



快手搜索 卧室榻榻米 Q

卧室做榻榻米 一定要牢记这几点

涨知识 上快手
随时随地 发现新知识

点击下载

而且收拾卫生不藏灰尘

通过优质知识内容吸引站外的兴趣用户

MMU

迷你视频知乎

用知识型视频回答问题

新能源汽车的优缺点

按回车键搜索

「新能源汽车的优缺点」

新能源汽车的
优点和缺点

第一个续航焦虑
上车就开右眼看

视频id: 49017838879
科普: 汽车

新能源汽车缺点

技术改变人生
北方威能梦想

视频id: 52096537874
科普: 汽车

新能源汽车分三类

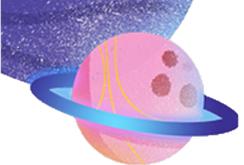
了解特点 买车不会花冤枉钱

视频id: 71472022577
科普: 汽车

新能源车值得买吗

视频id: 68854813577
科普: None

多模态知识问答



对下游任务的帮助

Entity Typing

原始输入: [CLS] 这家店里有一只柴犬, 毛发特别有质感 [SEP]

增强输入: [CLS] 这家店里有一只柴犬, 毛发特别有质感 [SEP] 养护[MASK] [SEP] 毛色特征[MASK] [SEP]...

输出 : 生物/动物

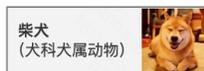
Task	Input	P	R	F1
Entity Typing	Vanilla	97.79	97.07	97.43
	Enhanced	98.8	97.44	98.12
Entity Linking	Vanilla	74.82	83.77	79.04
	Enhanced	78.91	83.67	81.22

Entity Linking

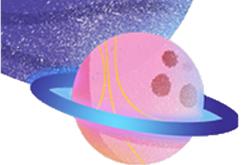
原始输入: [CLS] 这家店里有一只柴犬, 毛发特别有质感 [SEP] 柴犬 (犬科犬属动物), 日本犬种之一, 属于中小型犬种...

增强输入: [CLS] 这家店里有一只柴犬, 毛发特别有质感 [SEP] 柴犬 (犬科犬属动物), 日本犬种之一, 属于中小型犬种... [SEP] 养护 [MASK] [SEP] 毛色特征 [MASK] [SEP] ...

输出 : 链接概率 0.88



* [MASK] 位置的Embedding将由该知识点TOP1视频的embedding表示



对下游任务的帮助

- 快知中的知识作为Prompt, 帮助LLM回答事实性问题

<p>Prompt:</p> <p>Can all bees sting?</p>	<p>Prompt: <u>Bees are social insects, composed of individuals such as queen bees, male bees, and worker bees.</u> Can all bees sting?</p>	<p>Prompt: <u>Bees are social insects, composed of individuals such as queen bees, male bees, and worker bees. There are some aspects of "bee" as follows, Biodistribution, ecology communication, division of labor, selection, types, to-hive, prevent fights, water feeding, swarming, prevent stealing, exchange queen, spring fertility, making honey, lifespan, bee larva, inspection of a box, overwintering, sting, bee larva, bee medicine, nest building, death.</u> Can all bees sting?</p>	<p>Prompt: <u>Bees are social insects, composed of individuals such as queen bees, male bees, and worker bees. Some frames of a short video on the aspect "sting" of item "bee" is as follows.</u> Can all bees sting?</p> 
<p>Answer (0/10):</p> <p>✓ -</p> <p>✗ Yes, all bees can sting.</p> <p>✗ Yes, all bees have a stinger and can sting.</p>	<p>Answer (1/10):</p> <p>✓ Yes, all female bees can sting, but male bees do not have the ability to sting.</p> <p>✗ Yes, all bees can sting.</p>	<p>Answer (2/10):</p> <p>✓ Yes, only female bees (worker bees and queen bees) can sting. Male bees do not have a stinger.</p> <p>✗ Yes, all bees have a stinger and are capable of stinging</p>	<p>Answer (10/10):</p> <p>✓ No, not all bees can sting. Only adult female worker bees can sting.</p> <p>✓ No, not all bees can sting. Only adult female worker bees have the ability to sting and use it to defend their hive against threats.</p> <p>✗ -</p>

(a). Simple prompt

(b). + item description

(c). + description & aspects of item

(d). + description & frames of a video

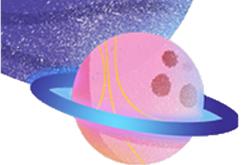
- 快知中的知识作为Prompt, 帮助文生图模型生成更符合事实的图片

<p>Prompt: Rose cuttage</p>	<p>Prompt: Rose cuttage. One of a videos tells <u>how to propagate rose cuttings using a simple method. Cut rose branches into pieces, smooth the bottom with a blade. Dip rose branches into sap and spread them flat on a towel. Cover with water and place in a warm, sunny location. After about a month, roots will grow and can be planted in a pot.</u></p>	<p>Prompt: Rose cuttage</p> 
		

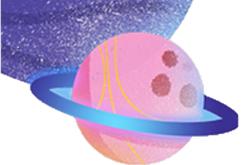
(a). Only Item-aspect "Rose Cuttage"

(b). + ASR text of a video of "Rose Cuttage"

(c). + frames of a video of "Rose Cuttage"



1. 背景
2. 快知体系
3. 构建方式
4. 应用与落地
- ▶ 5. 未来工作

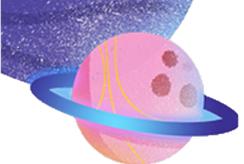


知识体系的完善与更新

视频切片粒度知识链接

知识权威度的判别





快知已开源，欢迎大家关注



Kuaipedia Public

main 1 branch 0 tags

ScarletPan Update README.md cce5159 on Jun 20 19 commits

- data Update README.md 2 months ago
- images release a subset of kuaipedia 2 months ago
- LICENSE release a subset of kuaipedia 2 months ago
- README.md release a subset of kuaipedia 2 months ago

README.md



Kuaipedia is developed by [Kuaishou Technology](#), collaborating with HIT and HKUST. It is the world's first large-scale multi-modal short-video encyclopedia where the primitive units are items, aspects, and short videos.



Github



Arxiv



知乎

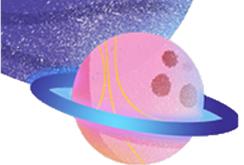
开放快知百科部分数据

	全集	已开源
词条数	> 26 million	51,702
知识点数	> 2.5 million	1,074,539
视频数	> 200 million	769,096

开放评测集和模型结果

Model	Item P	Item R	Item-Aspect P	Item-Aspect R
Random	87.7	49.8	36.4	49.6
LR	90.4	68.3	55.1	2.7
T5-small	93.7	76.1	79.3	58.5
BERT-base	94.3	77.8	81.5	62.7
GPT-3.5	90.5	86.4	41.8	95.7
Ours	94.7	79.7	83.0	65.7

<https://github.com/Kuaipedia/Kuaipedia>



感谢您的观看

 Kuaipedia |  快手

