

# 知识图谱前沿技术研讨会



肖仰华  
复旦大学

# 人与机器

---

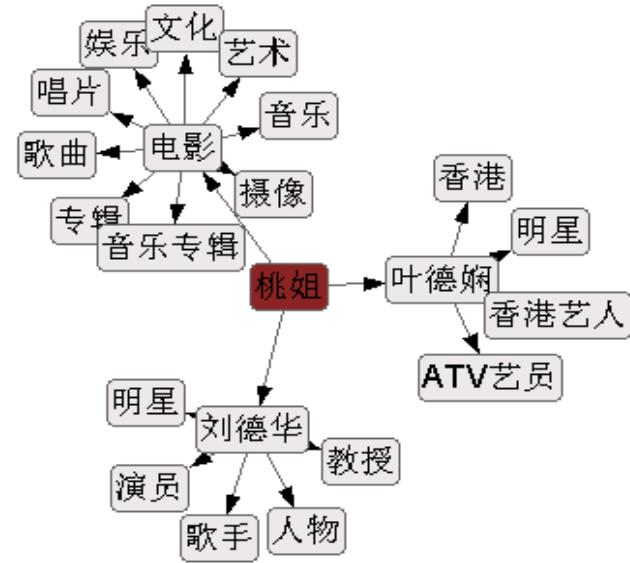


**机器语义理解是机器与人和谐共存的前提！**

# 关联是机器语义理解的核心

## 知识图谱

- Google 2012年提出知识图谱，应用于搜索
- 是大数据时代一种典型的**大规模知识**表示形式，表达了**实体及其之间的各类语义关系**，相对于本体等知识表示模型，有着较高的实体、概念覆盖率，表达了更为丰富的语义关系



## 意义1：下一代智能信息处理的关键技术

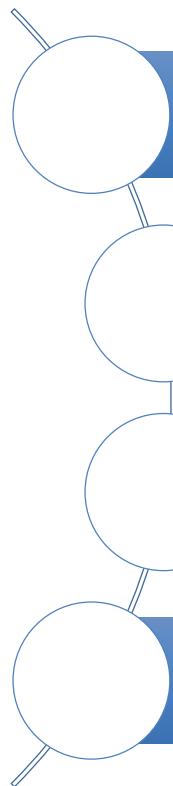
知识图谱是意图理解、关系解释、用户画像的技术基础；  
在智慧搜索、智慧软件、移动安全、智慧商务等领域有着广阔应用前景

## 意义2：实现机器智脑（语言认知）的关键技术

- 美国脑计划、欧盟（IBM）蓝脑计划、Google全球脑计划、百度大脑计划、中国脑计划
- 知识图谱为语义理解提供了丰富的**背景知识**，是探索语言以及概念认知的**数据基础**，是机器智脑实现**类人智能**的核心

# 机遇与挑战

## 新机遇



众包技术的成熟使得利用大规模人力资源构建知识图谱成为可能

完成初始积累的在线知识库正迈进爆发式增长周期

知识图谱在搜索之外应用价值日益显现

深度学习使得无监督的知识图谱构建成为可能

## 新挑战

体量更巨  
大

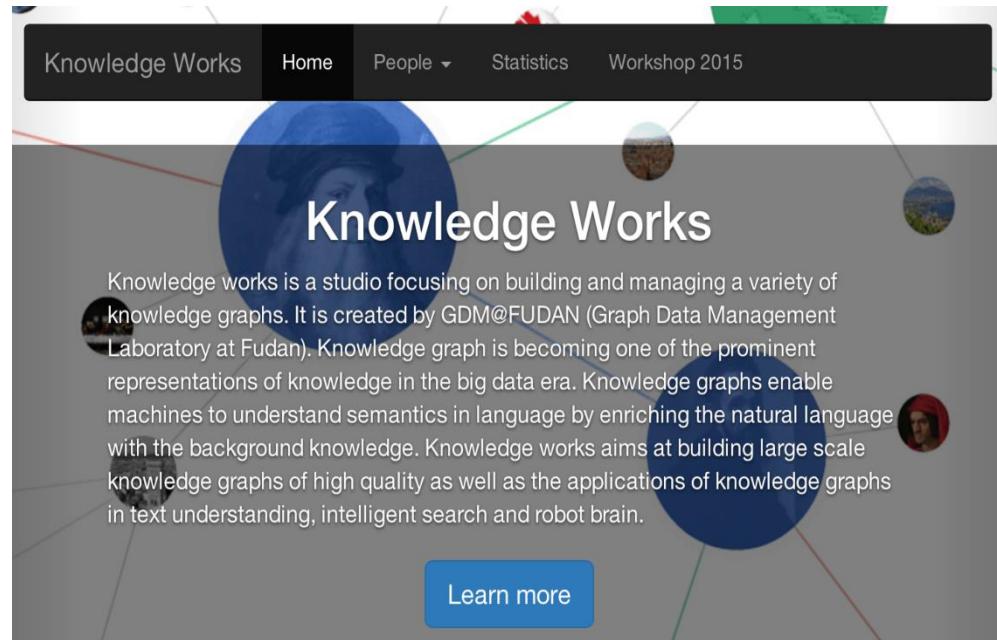
关系更多  
样

质量要求  
更苛刻

计算任务  
更复杂

# 知识工场平台

知识图谱云服务平台  
•集中发布、展示各  
类知识图谱，以及相  
关的应用系统与平台；  
•提供丰富的API访问  
接口。



<http://kw.fudan.edu.cn>

# 知识库1：CN-DBpedia



致力于从中文百科类网站（如百度百科、互动百科等）中提取结构化信息，并将其共享。目前已包含千万级实体、亿级事实。

- 支持用户进行复杂查询，
- 并将中文百科类网站数据与互联网上其他数据源数据进行连接

## 关键技术

- 基于深度学习的概念识别
- 实体链接
- 属性融合
- 跨语言类别推断
- 链接实体重组

Examples: <复旦大学, 校长, 许宁生>

中文名	刘德华
主要成就	泰山电影节亚洲最有影响力大奖
主要成就	三届香港电影金像奖最佳男主角
主要成就	两届台湾电影金马奖最佳男主角
主要成就	中国电影百年形象大使
主要成就	全港最高礼仪迎宾司仪颁奖典礼司仪
代表作品	御剑惊龙
代表作品	天若有情
代表作品	忘情水
代表作品	无间道
代表作品	暗战
代表作品	阿飞
代表作品	60KG
信仰	佛教
全球影响力	华语天王
出生地	香港新界大埔碧云村
出生日期	1961-09-27
别名	华仔
国籍	中国
外文名	Andy Lau
外文名	Lau Tak Wah
女儿	刘向蕙
妻子	朱丽倩
星座	天秤座 (七星堂)
毕业院校	可立中学
毕业院校	第十期无线艺员训练班

# 知识库2：Probbase Plus



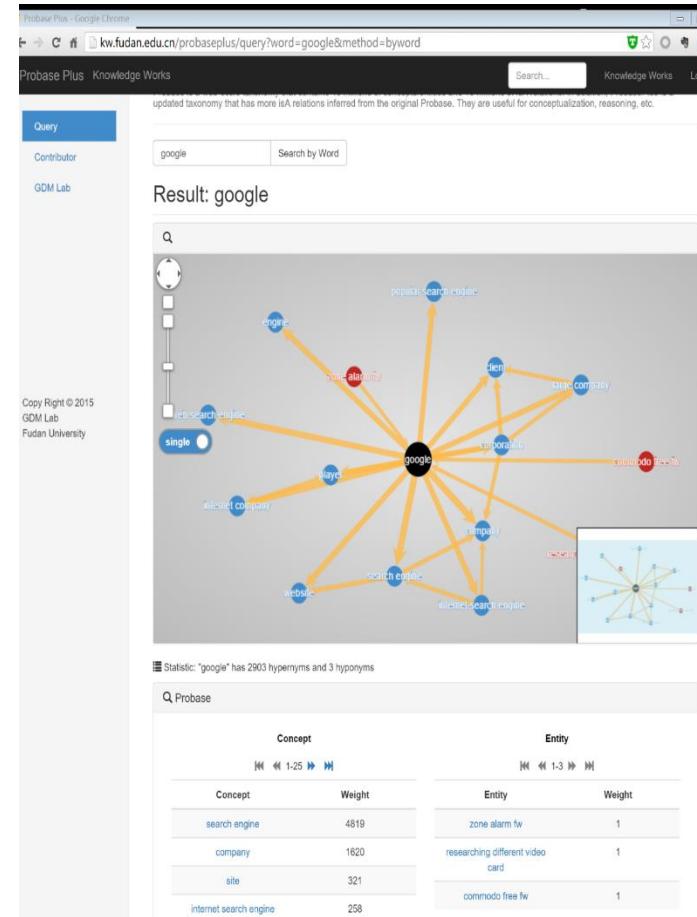
基于微软亚洲研究院大规模分类体系Probbase，通过一系列质量改进技术构建的大规模实体概念分类体系。

- 包含千万级的实体与概念，以及 isA 关系。
- 在概念化、摘要、搜索意图理解、用户画像等领域发挥重大价值。

## 关键技术

- 缺失边补全
- 错误边删除
- isA 关系推断
- 多源分类体系融合
- 中文分类体系补全

Examples: Bill Gates isA Billionaire



# 知识库3：Verb-base



基于概念模式的大规模动词知识库。  
通过挖掘动词短语的概念模式，构建了常见动词的语义表示形式。

- 该知识库涵盖常见2万动词，共计14万短语。
- 基于概念模式的动词知识库已经在实体概念化、词汇的语义表示等方面发挥积极的作用。

## 关键技术

基于MDL的泛化

### Examples:

Verb Base Query

Verb pattern is a probabilistic semantic representation on verbs. Verbs are important in semantic understanding of natural language. So we introduce verb patterns to represent verbs' semantics, such that each pattern corresponds to a single semantic of the verb. We constructed verb patterns with the consideration of their generality and specificity.

call

verb	phrase	pattern	frequency	type
call	call police	call department	64299	Conceptualized Pattern
call	call attention	call skill	46157	Conceptualized Pattern
call	call doctor	call professional	14223	Conceptualized Pattern
call	call cop	call job	12143	Conceptualized Pattern
call	call home	call place	8578	Conceptualized Pattern

Verb Phrase	Verb Pattern
eat apple	eat \$ <sub>C</sub> food
eat hot dog	eat \$ <sub>C</sub> food
eat breakfast	eat \$ <sub>C</sub> meal
eat lunch	eat \$ <sub>C</sub> meal
eat dinner	eat \$ <sub>C</sub> meal
eat humble pie	eat \$ <sub>I</sub> humble pie

# 系统与平台

# CaIStayNight: 大规模分布式智能防屏蔽互联网爬虫平台

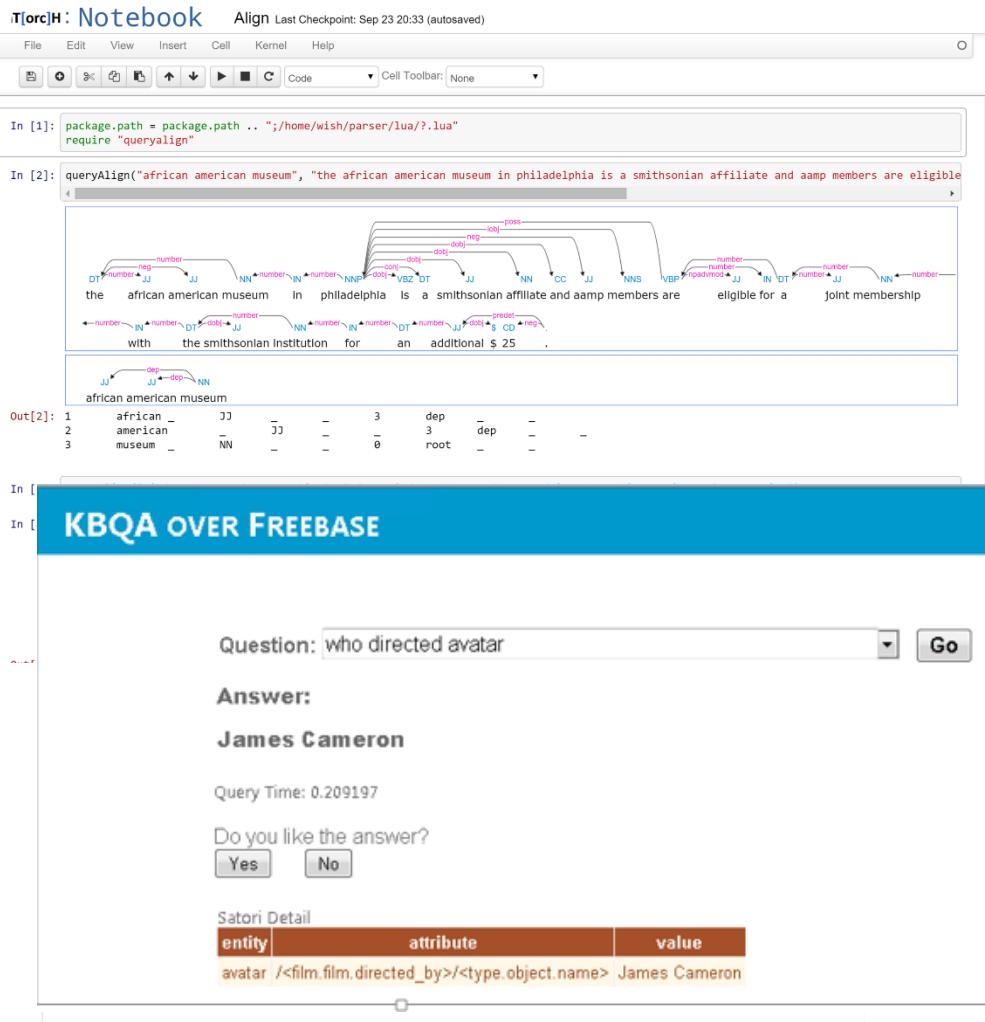
支持上百节点、分布式、防屏蔽、易部署、高可用、优先级敏感的互联网数据爬取平台。

# NLPDL: 基于Torch深度定制的深度学习平台

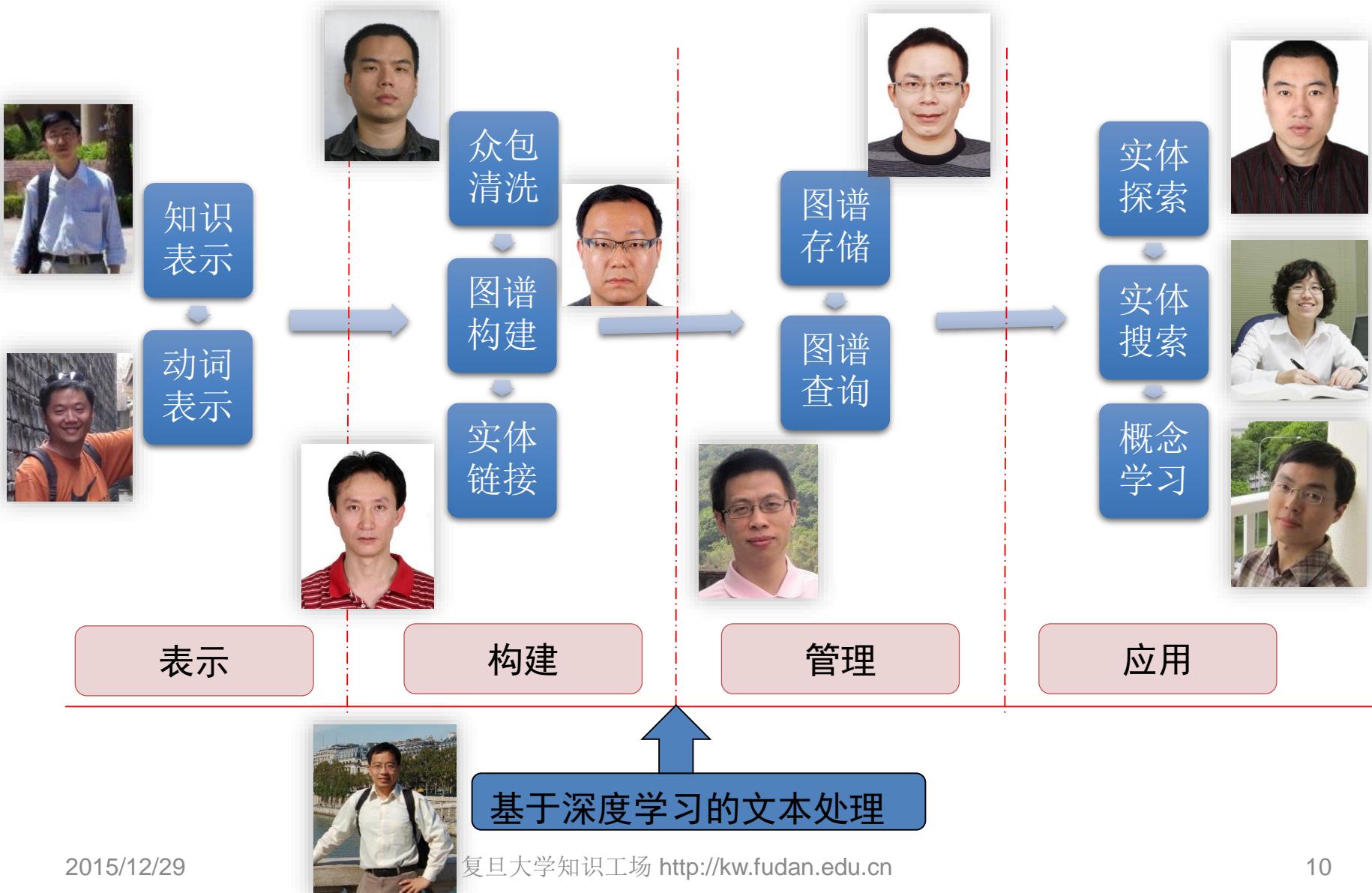
支持主流深度学习模型与常见自然语言处理任务、  
支持各类基于深度学习模型的知识抽取任务

# KBQA：基于知识图谱的自然语言问答系统

对于RDF的知识图谱直接进行自然语言回答，能够回答各类事实性问答。



# Outline



# 谢谢！