

Overview of Knowledge Graph

Hai Wan

School of Data and Computer Science
SUN YAT-SEN UNIVERSITY

Web: Link-Centered System

- Tim Berners-Lee
 - The father of World-Wide-Web
 - 2016 Turing Award Winner



Hypertext Links between Web pages

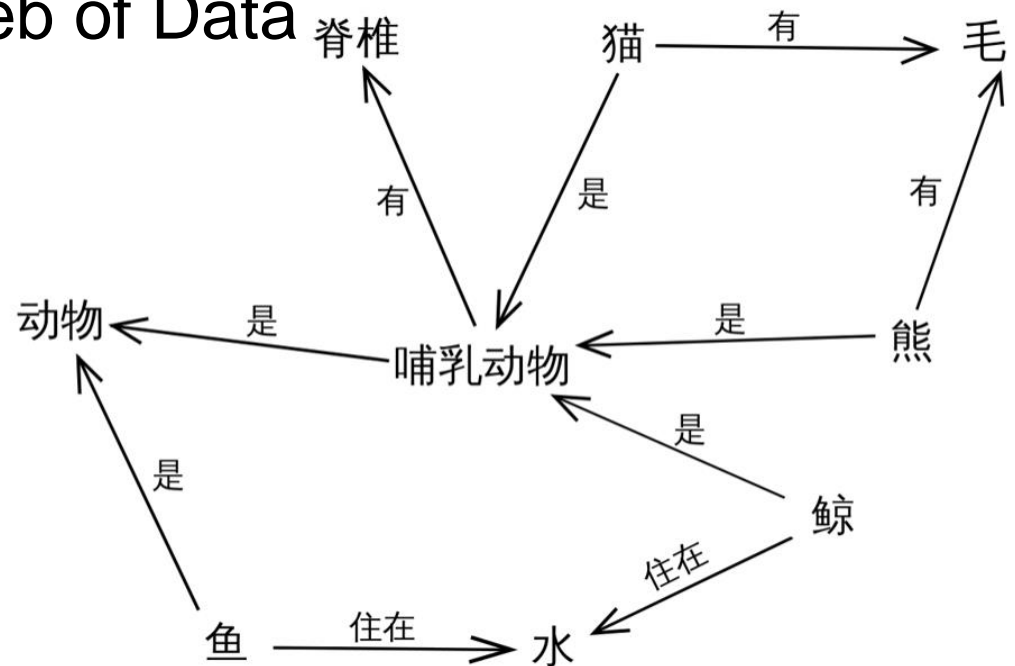
Semantic Web

- 1998 by Tim Berners-Lee
 - Web of Texts, Web of Documents



- Web of Objects, Web of Data

哺乳动物拥有脊椎这一身体结构
猫和熊都是有毛的哺乳动物
鲸也是哺乳动物，它住在水里
鱼是一种住在水里的动物



Knowledge Graph

■ 2012 by Google

- used to improve the performance of Google Search
 - understand the queries
 - more accurate results
-

Knowledge Graph

■ Google search: Tim Berners-Lee

Tim Berners-Lee - Wikipedia

https://en.wikipedia.org/wiki/Tim_Berners-Lee ▼ 翻译此页

Sir **Timothy John Berners-Lee** OM KBE FRS FREng FRSA FBCS (born 8 June 1955), also known as TimBL, is an English engineer and computer scientist, best ...

Other names: TimBL; TBL **Parent(s):** Conway Berners-Lee; Mary Lee Woods

Children: 2 **Occupation:** Professor of Computer Science

[Conway Berners-Lee](#) · [Mary Lee Woods](#) · [List of awards and honours ...](#) · [Railfan](#)

蒂姆·伯纳斯-李- 维基百科，自由的百科全书

<https://zh.wikipedia.org/zh-hans/蒂姆·伯纳斯-李> ▼

蒂莫西·约翰·伯纳斯-李爵士，OM KBE FRS FREng FRSA FBCS（英语：Sir **Timothy John Berners-Lee**，1955年6月8日－），暱稱為蒂姆·伯纳斯-李（英语：**Tim ...**

[早期生活和教育](#) · [生涯](#) · [现况](#) · [荣誉](#)

Tim Berners-Lee - World Wide Web Consortium

<https://www.w3.org/People/Berners-Lee/> ▼ 翻译此页

Sir **Tim Berners-Lee** invented the World Wide Web in 1989. He is the Director of the World Wide Web Consortium (W3C), a Web standards organization founded ...

Tim Berners-Lee (@timberners_lee) | Twitter

https://twitter.com/timberners_lee ▼ 翻译此页

The latest Tweets from **Tim Berners-Lee** (@timberners_lee). Director of the World Wide Web Consortium (W3C) w3.org, the place to agree on web standards.

Sir Tim Berners-Lee – World Wide Web Foundation

<https://webfoundation.org/about/sir-tim-berners-lee/> ▼ 翻译此页

The inventor of the World Wide Web and one of Time Magazine's '100 Most Important People of the 20th Century' Sir **Tim Berners-Lee** is a scientist and



提姆·柏納-李

计算机科学家

蒂莫西·约翰·伯纳斯-李爵士，OM KBE FRS FREng FRSA FBCS，昵称为蒂姆·伯纳斯-李，英国计算机科学家。他是万维网的发明者。1990年12月25日，他成功利用互联网实现了HTTP客户端与服务器的第一次通讯。伯纳斯-李是万维网联盟的主席，为关注万维网发展而创办的组织。他也是万维网基金会的创办人。 [维基百科](#)

生于：1955年6月8日（63岁），英國倫敦

職業：計算機科學家

所获奖项：圖靈獎，皇家獎章，千禧年科技獎，更多

教育背景：牛津大學王后學院（1973年–1976年），更多

配偶：羅斯瑪麗·萊斯（结婚时间：2014年），南希·卡爾森（结婚时间：1990年–2011年）

图书



Knowledge Graph

■ 百度搜索：刘德华的女儿的妈妈

[刘德华女儿和母亲一起现身,扎着马非常可爱,身边还有两个保镖](#)

2018年9月25日 - 而就在最近[刘德华的女儿](#)和妻子朱丽倩一起现身。从网友们流露出的照片上来看,[刘德华的女儿](#)拉着朱丽倩的手,好...

<https://baijiahao.baidu.com/s?...> - 百度快照

[刘德华女儿照片曝光 大眼可爱像妈妈朱丽倩_尚之潮](#)



2017年9月26日 - 刘德华近日的生日会上,女儿终于曝光了,刘德华抱着自己的女儿,全程参加自己的生日会,从现场的照片中,可以看出来,[刘德华的女儿](#)长得非常可爱,大大的眼睛,...

www.shangc.net/news/20... - 百度快照

[隐瞒6年,刘德华女儿终曝光,长相惹争议,还好女儿长得像妈!_搜狐](#)



2018年10月4日 - 如今昔日的少女已经变身全职妈妈了,女儿今年已经6岁,看过刘向蕙的朋友都说[刘德华女儿](#)很像妈妈朱丽倩,长得...

www.sohu.com/a/2576414... - 百度快照

[刘德华女儿正面照曝光 刘德华女儿竟长这个样子难以置信_娱乐新闻_](#)



2018年6月14日 - 一直以来很多人都对[刘德华的女儿](#)感到好奇,都想看看[刘德华女儿](#)正面长什么样子,终于这回[刘德华女儿](#)的正脸照被曝光...

www.hxnews.com/news/yl... - 百度快照



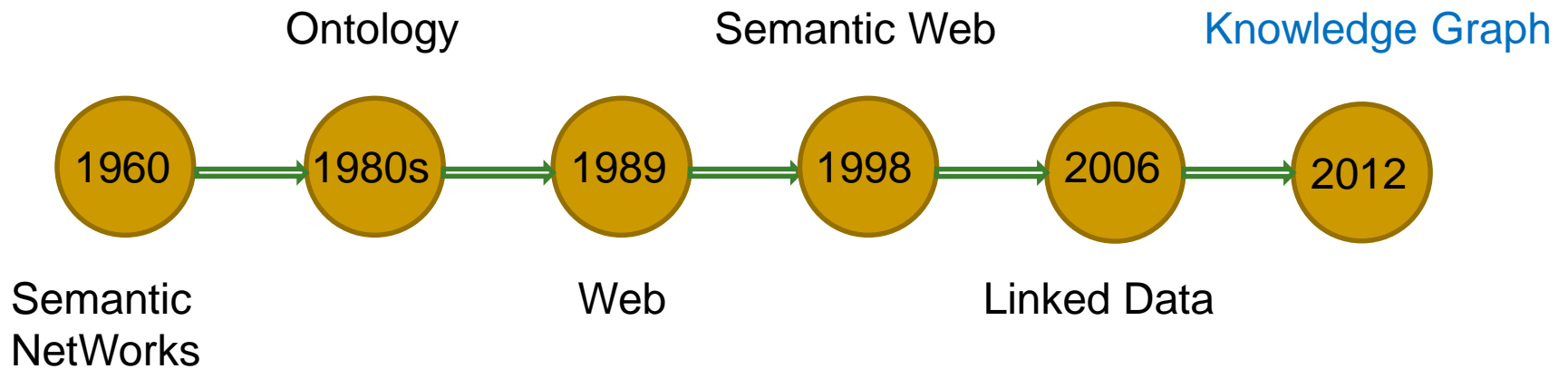
刘向蕙母亲:

朱丽倩

朱丽倩 (Carol), 1966年4月6日出生于马来西亚檳城。曾当选为马来西亚选美小姐,后做了平面模特。1984年,她和姐妹参加马来西亚檳城的“新潮小姐”选美获得季军,之后赴... [详情>>](#)

来自百度百科

Knowledge Graph



Knowledge Graph

 Freebase

 LINKING OPEN DATA
W3C SWEO Community Project

WordNet
A lexical database for English

 2HISH.me

PKUBASE

NELL

schema.org

....the new SEO?

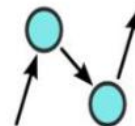
 DBpedia

 XLORE

 WIKIDATA

WEB CHILD

CN-DBpedia



ConceptNet

An open, multilingual knowledge graph

Herbnet

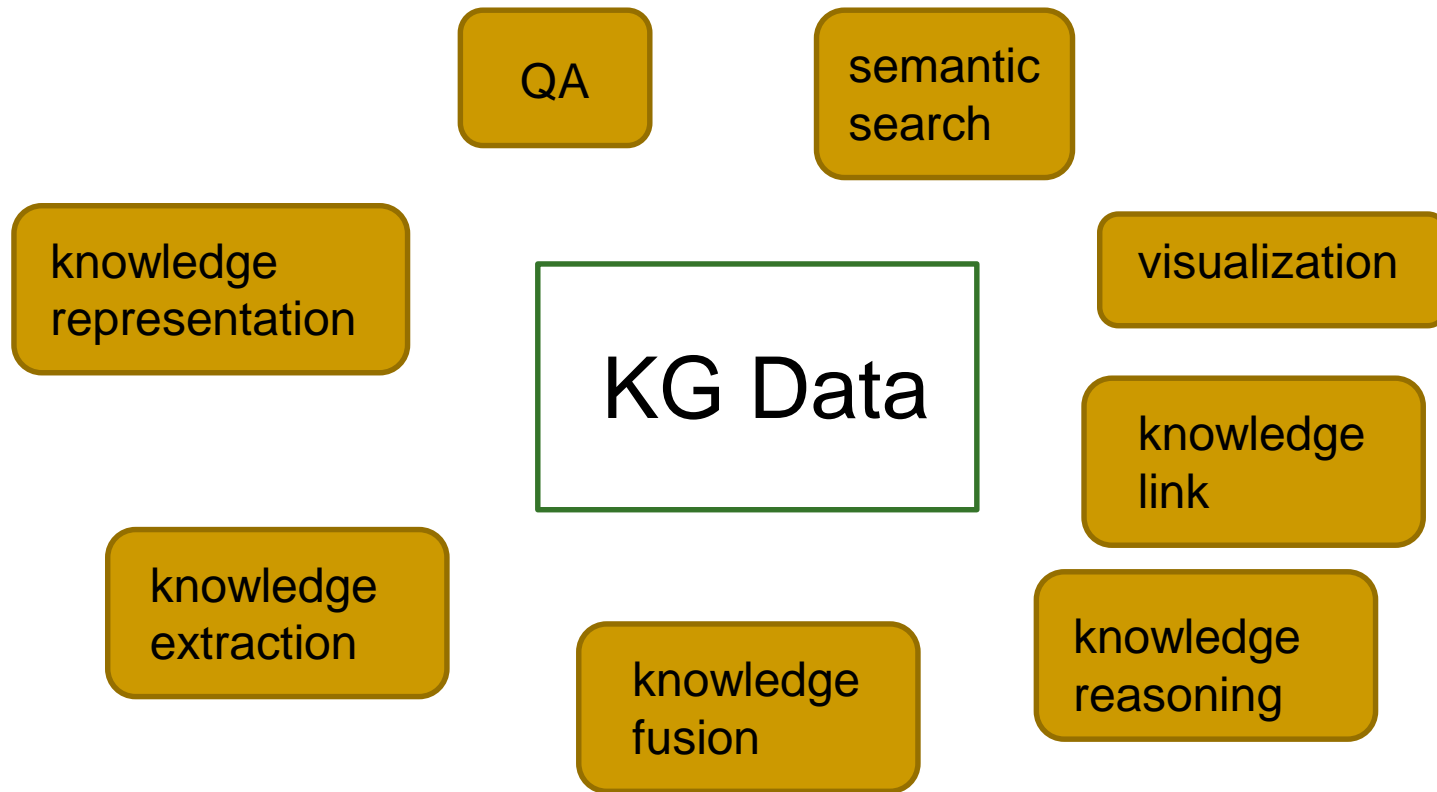
 yago
select knowledge

 LinkedGeoData.org

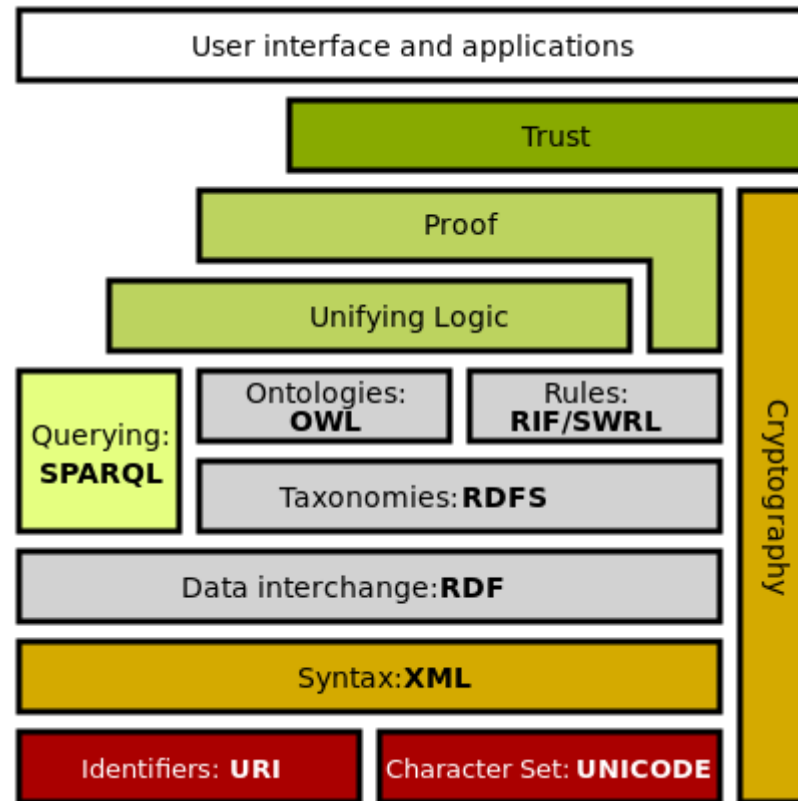
WEBKB

linked life data 

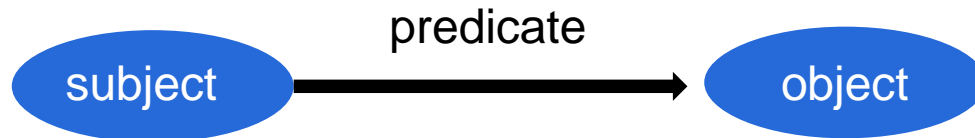
Knowledge Graph



Knowledge Representation



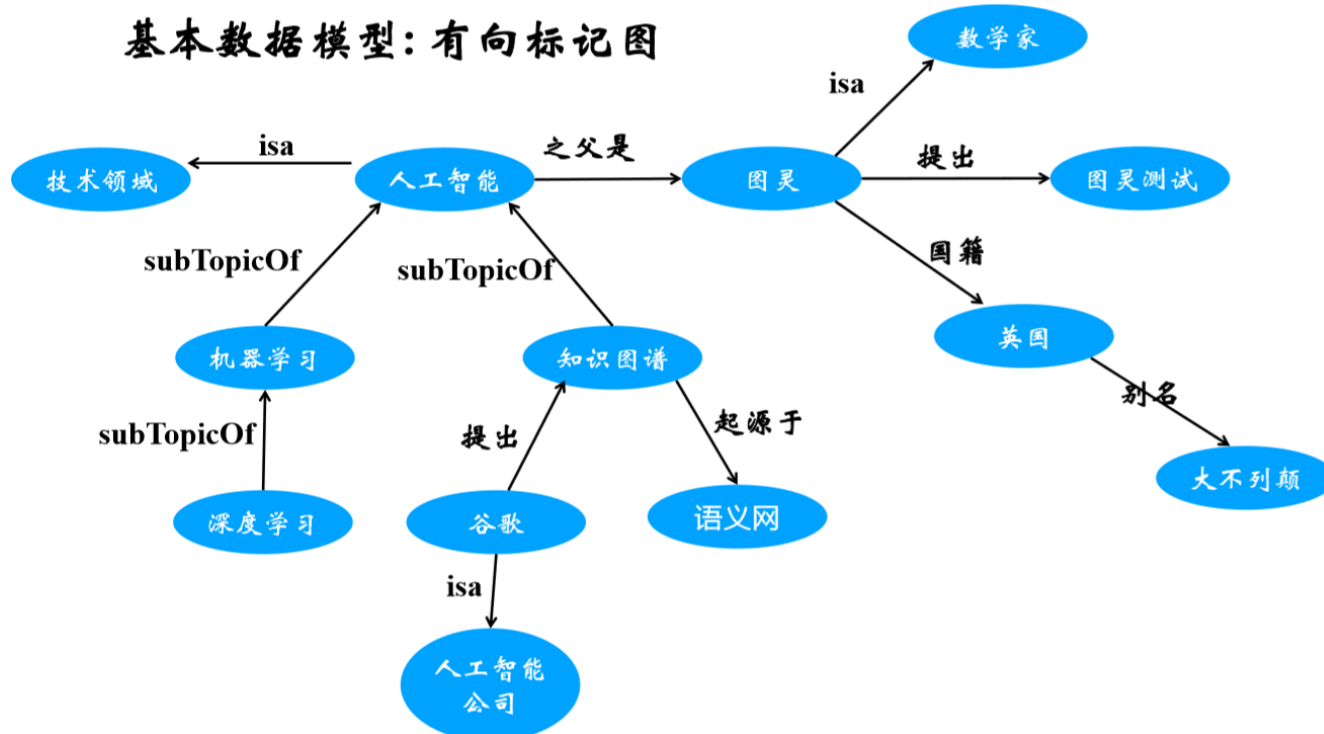
RDF: Triple-based Assertion Model



use SPARQL to
do some search

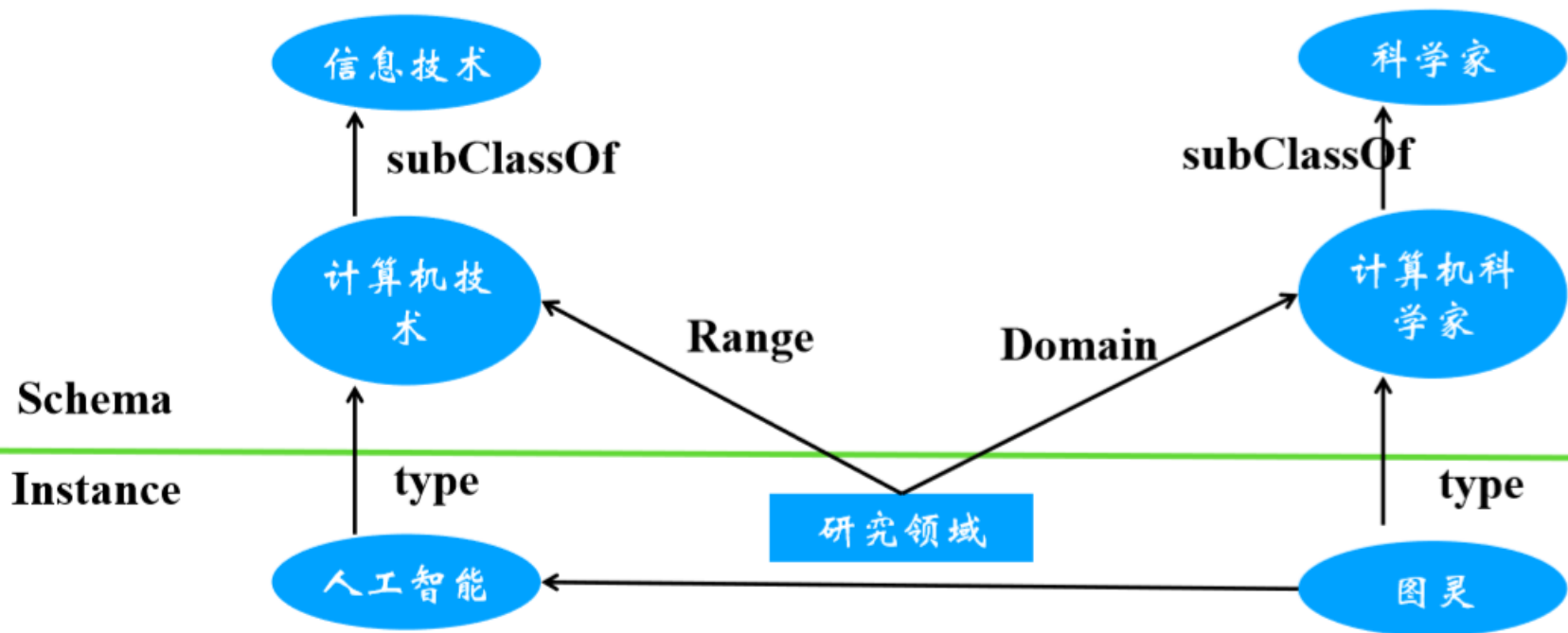
RDF Graph: Directed Labeled Graph

基本数据模型: 有向标记图



RDFS: RDF Schema

similar to the design of table and field in database



OWL: Web Ontology Language

Ontology in Philosophy

Ontology is the philosophical study of the nature of being, becoming, existence or reality, as well as the basic categories of being and their relations.

—— Merriam-Webster

Ontology in Computer Science and Artificial Intelligence

An ontology is a description (like a formal specification of a program) of the concepts and relationships that can formally exist for an agent or a community of agents.

—— Tom Gruber, Founder of Siri

Web Ontologies

Ontologies based on web standards such as RDFS/OWL. OWL is based on Description Logic, a very very long history of research in Artificial Intelligence.

Other Representation

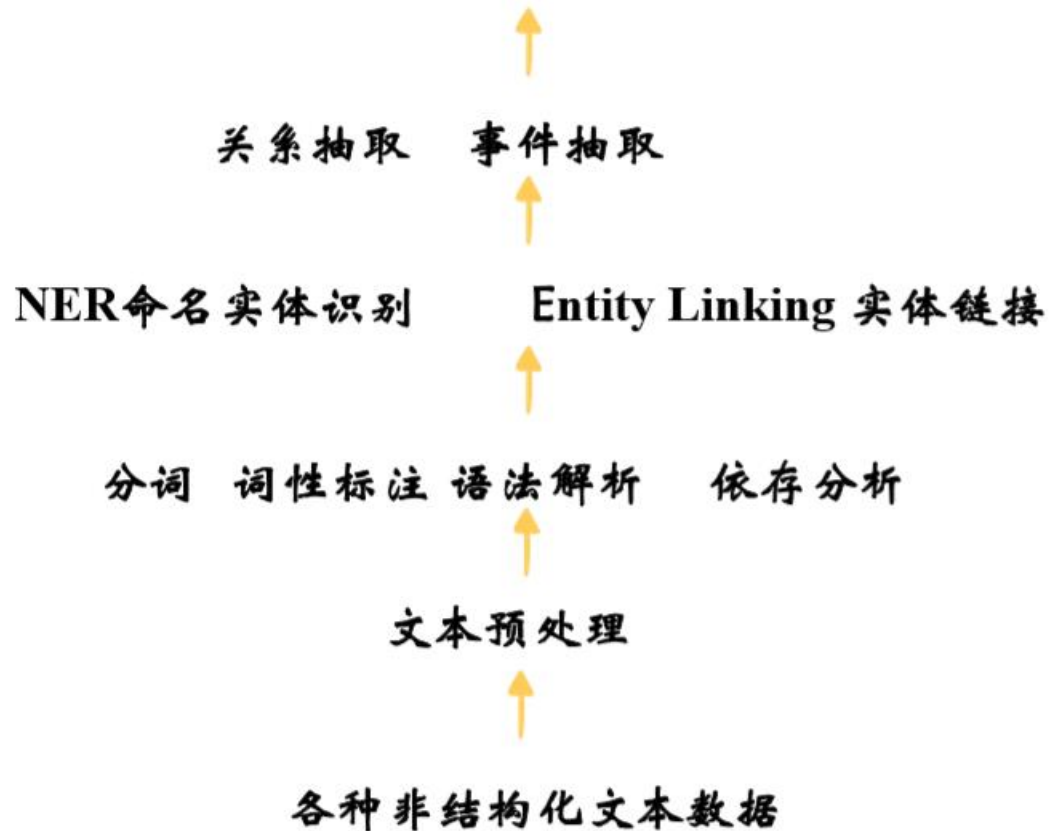
- JSON-LD: JSON for linking data
- RDFa
- HTML5 Microdata

structured data in Web pages

```
<div itemscope itemtype="http://schema.org/Movie"><p>  
  <span itemprop="name">Big </span> (1988) is a  
  <span itemprop="genre">fantasy comedy</span> movie starting  
  <span itemprop="actor" itemscope itemtype="http://schema.org/Person">  
    <span itemprop="name">Tom Hanks</span></span> and directed by  
  <span itemprop="director" itemscope itemtype="http://schema.org/person">  
    <span itemprop="name">Penny Marshall</span></span>.</p></div>
```

Knowledge Extraction

KR: 三元组、多元关系、模态知识.....



Knowledge Extraction

- Knowledge Engineering
 - regular expression
 - template matching
 - BootStrap
 - rule constraint
 - POS、NER、 etc.
- Ontology-based Extraction
 - knowledge mining / reasoning
 - PRA: graph-based extraction
 - TransE: embedding-based extraction

Knowledge Extraction

■ Model-based Extraction

□ Model

- SVM、 Logistic Model
- CRF
- LSTM etc.

□ Training

- supervised learning
 - unsupervised clustering
 - distant supervision
-

Knowledge Storage

■ DataBase

□ Relational Database

- MySQL
- PostgreSQL

□ Non-Relational Database

- MongoDB
- Redis
- Graph Database
 - neo4j
 - Jena TDB

Knowledge Reasoning

example: 刘德华的女儿的妈妈的老公是谁

- reasoning based on description logic
 - Ontology reasoning
- reasoning based on statistical rule mining

$$isPartOf(A, B) \Leftrightarrow hasPart(B, A)$$

$$hasFriend(A, B) \Leftrightarrow hasFriend(B, A)$$

- reasoning based on representational learning
 - word to vector > vector computing

Knowledge Fusion

Find data of the same entity in different datasets, and the main purpose is to integrate the entity data in different data sources, so we can get a more comprehensive entity information

■ Tools

- ❑ Dedupe
- ❑ LINES
- ❑ you can find them in OpenKG