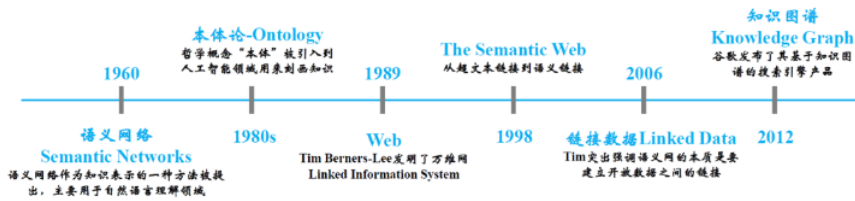


## 一、知识图谱概论



### 基础-语义网和本体论

知识图谱由文本(Texts、Documents)的链接到对象(Objects)的链接。

(知识图谱是由一些相互连接的实体和它们的属性构成的)

通过已有的规则和知识，思考、语言、推理来进行学习。**推理**

作用:

- 1、辅助搜索
- 2、辅助问答
- 3、辅助决策
- 4、辅助AI：常识推理

### KG本质

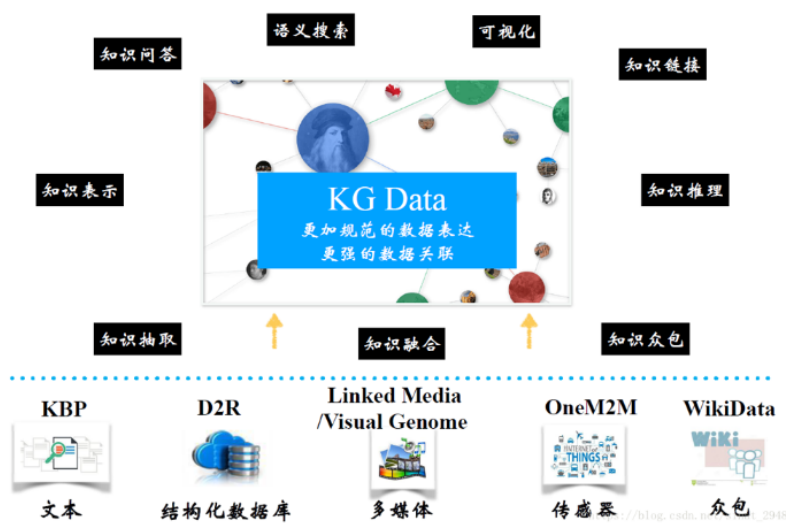
**Web视角** 像建立文本之间的超链接一样，建立数据之间的语义链接，并支持语义搜索

**NLP视角** 怎样从文本中抽取语义和结构化数据

**KR视角** 怎样利用计算机符号来表示和处理知识

**AI视角** 怎样利用知识库来辅助理解人的语言

**DB视角** 用图的方式去存储知识



### 1、知识表示

研究怎样用计算机符号来表示人脑中的知识，以及怎样通过符号之间的运算来模拟人脑的推理过程。

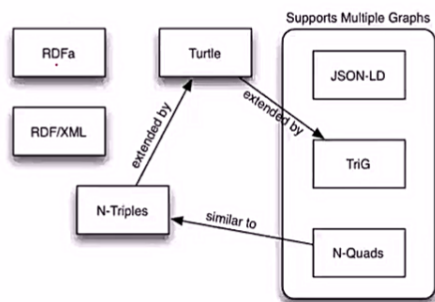
知识图谱的离散式表示：

RDF：主语--谓语-->宾语（三元组）

RDF Graph：有向标记图

RDFS：在RDF基础上定义固定vocabulary、schema：Class, subClassOf, type, Property, subPropertyOf, Domain, Range以及多了Schema

RDF的序列化格式：RDFa、JSON-LD



OWL (Web Ontology Language) 网络本体语言，对本体语言进行描述

OWL extends RDF Schema

```
复杂类
:Mother owl:equivalentClass [
  rdf:type owl:Class ;
  owl:intersectionOf ( :Woman :Parent )
] .

属性约束
:Parent owl:equivalentClass [
  rdf:type owl:Restriction ;
  owl:onProperty :hasChild ;
  owl:someValuesFrom :Person
] .

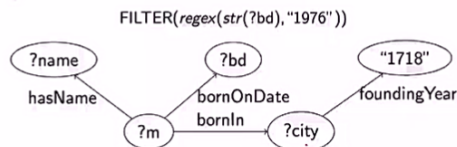
对称属性
:hasSpouse rdf:type owl:SymmetricProperty .

传递属性
:hasAncestor rdf:type owl:TransitiveProperty .

属性链
:hasGrandparent owl:propertyChainAxiom ( :hasParent :hasParent ) .
```

SPARQL：RDF查询语言 "?"

```
SELECT ?name
WHERE {
  ?m <bornIn> ?city. ?m <hasName> ?name.
  ?m<bornOnDate> ?bd. ?city <foundingYear> '1718'.
  FILTER(regex(str(?bd),'1976'))
}
```



JSON-LD：数据交换格式

JSON for Linking Data: 适用于作为程序之间做数据交换，在网页中嵌入语义数据和Restful Web Service

```
{
  "@context": "http://json-ld.org/contexts/person.jsonld",
  "@id": "http://dbpedia.org/resource/John_Lennon",
  "name": "John Lennon",
  "born": "1940-10-09",
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```

## RDFa HTML5 MicroData: 在网页中嵌入语义数据

Semantic Markup: 在网页、邮件、应用程序中嵌入语义数据

```
<p>Big (1988) is a fantasy comedy movie starting Tom Hanks and directed by Penny Marshall.</p>
```

RDFa, JSON-LD, HTML5 Microdata

```
<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>
```

搜索引擎优化

知识图谱的分布式表示:

KG Embedding: 在保留语义的同时，将KG中的实体和关系映射到连续的稠密的低维向量空间

张量分解: 多维->低维; 神经网络、距离模型

## 2、知识抽取: NLP+KR

知识抽取的主要方法:

- (1).知识工程: 正则表达式、模板匹配、规则约束
- (2).基于本体的抽取: 知识挖掘(推理)
- (3).基于模型的抽取: 模型、训练

## 3、知识存储

知识比数据的结构更加复杂，知识的存储需要综合考虑图的特点、复杂的知识结构存储、索引和查询支持推理) 的优化等问题。

典型的知识存储引擎分为基于关系数据库的存储和基于原生图的存储

在实践中，多为混合存储结构，图存储并非必须

## 4、知识问答

KBQA (Knowledge-Based Question Answering, 基于知识库的问题回答) 以直接而准确的方式回答用户自然语言提问的自动问答系统将构成下一代搜索引擎的基本形态。

## 5、知识推理

基于已知事实推出未知的事实的计算过程。

按解决方法分类：

按推理类型分类：

①基于描述逻辑的推理

①缺省推理 Default reasoning

②基于规则挖掘的推理

②连续变化推理 Continuous Change

③基于概率逻辑的推理

③空间推理 Spatial reasoning

④基于表示学习与神经网络的推理

④因果关系推理 Causality reasoning

⑤...

基于描述逻辑的推理：本体推理

描述逻辑（description logic）是一种用于知识表示的逻辑语言和以其为对象的推理方法，主要用于描述概念分类及其概念之间的关系。

描述逻辑是当前语义网发展中本体的理论基础。

基于统计规则挖掘的推理：路径排序学习方法、关联规则挖掘方法

基于表示学习的推理-Representational Learning

## 6、知识融合

知识融合是指在不同数据集中找出同一个实体的描述记录，主要目的是对不同数据源中的实体信息进行整合，形成更加全面的实体信息。（不同数据源中找到同一实体的不同表现形式）

工具：Dedupe、LIMES

## 7、知识众包

Wikibase

Schema.ORG