Notes

What is a knowledge graphs?

A knowledge graph consists of a set of interconnected typed entities and their attributes.

(知识图由一组相互连接的类型化实体及其属性组成)

知识图谱是由一条条知识组成的，每条知识可以表示成一个SPO三元组(Subject-Predicate Object)。实体是知识图谱中的最基本元素，不同的实体间存在不同的关系。概念主要指集合、类别、对象类型、事物的种类，例如人物、地理等；属性主要指对象可能具有的属性、特征、特性、特点以及参数，例如国籍、生日等；属性值主要指对象指定属性的值，例如中国、1988-09-08等。现在的知识图谱已被用来泛指各种大规模的知识库。

知识图谱的架构：在逻辑上可以分为模式层和数据层。数据层由一系列的事实组成，而知识将以事实为单位进行存储（事实即是用实体-关系-实体或者实体-属性-属性值表示）。模式层构建在数据层之上，主要是通过本体库来规范数据层的一系列事实。

大规模知识库可以分为两类：开放链接知识库和行业知识库。其中开放链接知识库中，Freebase、Wikidata、DBpedia、YAGO这是个大规模知识库处于核心地位。行业知识库也可称为垂直型知识库，这类知识库的描述目标是特定的行业领域，通常需要依靠特定行业的数据才能构建，因此其描述范围极为有限，主要有IMDB（internet movie database）、MusicBrainz、ConceptNet等。

RDF：Resource Description Framework，是由W3C制定的，用于描述实体资源的标准数据模型。在知识图谱中使用RDF形式化的描述这种三元关系。RDF图中一共有三种类型，International Resource Identifiers(IRIs)，blank nodes 和 literals。

What is a DBpedia?

DBpedia是受欢迎的可免费获得的知识图谱（知识库）。DBpedi是由德国莱比锡大学和曼海姆大学的科研人员创建的多语言综合型知识库，在LOD（linked open data）项目中处于最核心的地位。DBpedia是从多种语言的维基百科中抽取结构化信息，并且将其以关联数据的形式发布到互联网上，提供给在线网络应用、社交网站以及其他在线知识库。由于DBpedia的直接数据来源覆盖范围广阔，所以它包含了众多领域的实体信息。截止至2014年年底，DBpedia中的事实三元组数量已经超过了30亿条。除上述优点外，DBpedia还能够自动与维基百科保持同步，覆盖多种语言。

知识图谱的关键技术：知识抽取、知识表示、知识融合、知识推理。

知识抽取：分为实体抽取、关系抽取、属性抽取。

What is a trivia?

A Trivia is any fact about an entity which is interesting due to its unusualness unexpectedness uniqueness or weirdness when compared to other entities of similar type.

（A Trivia是关于一个有趣实体的任何事实，因为与其他类似的实体相比，它具有不寻常性、意外性、独特性或者是怪异性。）

Precision,Recall,F1 score:

准确率（P值）是针对我们预测结果而言的，它表示的是预测为正的样本中有多少是真正的正样本。（预测正确的正类除以预测出来的总的正类）

召回率（R值）是针对我们原来的样本而言的，它表示的是样本中的正例有多少被预测正确了。（预测的正确的正类除以真正的正类数）

常见的F1 score=2\*P\*R/(P+R)