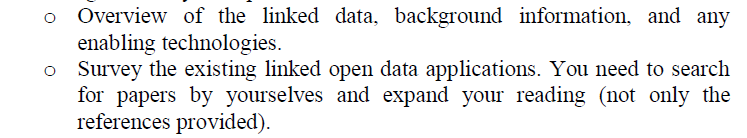
A Semantic Web Primer

The general vision of semantic web can be summarized in a single phrase: to make the web more accessible to computers.



Motivation for the semantic web (linked open data)

Where are we now?

There is rapidly maturing technology to support all phases of deployment of Semantic Web technology, and the number of substantial adoptions, both in commercial and public organizations, is growing rapidly. However, major challenges remain, such as dealing with the ever-increasing scale, lowering the barrier of adoption, and of course fighting that omnipresent bane of information systems: semantic heterogeneity.

What is the background before LOD?

A key aspect of the traditional web is the fact that its content is distributed, both in location and in ownership: web pages that link to each other often live on different web servers, and these servers are in different physical locations and owned by different parties.



In recent years, the concept of Linked Data, and the so-called Web of Linked Data, has attracted tremendous attention from both the academic world and real application world.

From Bibliographic Records to Data

Within this context, therefore, the term “Linked Data” refers to the method by which data can be displayed, published, linked and exchanged using Uniform Resource Identifier (URIs) and Resource Description Framework (RDF). Linked Data (LD) technologies make it possible to connect and enrich metadata, so that different representations of the same content can be searched and linked, thus relating resources that come from different sources and domains. Background

The Open Data initiative is closely related and linked to the conception of open government, with its “philosophy of open access to certain data without copyright restrictions” (Ferrer ; Peset; Benavent, 2011, p. 162).

It is necessary to clarify that the term Open Data does not refer to the mere availability of data on the network, i.e., Internet publication in such a way that data can be read and down­loaded. Hernández and García (2013, p. 260) claim that “to be truly open the data must be available online, preferably for download, but they must also have some sort of legal license so that they can be used, reused and redistributed, even mixed with other data, at the minimum sub­ject to “attribution” (recognition of authorship, the person who created it), or “share alike “(the norm that any exploitation made of such data, including derivative works, should maintain the same license to be disclosed).”

The *Guía breve de Linked Data* (2014) specifies that “the Semantic Web is not just about publishing data on the Web, but about doing it in such a way that they can be linked to others, so that people and machines can explore the web of data, and be able to reach related information referred from other initial data. “

The term Linked Data (LD) thus refers to a set of best practices for publishing and interlinking structured data on the Web. In the case of LD, data should be published “in accordance with the principles designed to fa­cilitate linkages among datasets, element sets, and value vocabularies” (Berners-Lee, 2011). These best practices were introduced in 2006 by Tim Berners-Lee in his Web architecture note Linked Data and they have become known as the Linked Data principles. These principles are the following:

1. Use URIs as names for things.

2. Use HTTP URIs, so that people can look up those names.

3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).

4. Include links to other URIs, so that people can discover more things.

That is, this set of best practices or recom­mendations included under this concept will allow sets of data, information, and knowledge to be exhibited, shared, and connected on the Semantic Web using URIs and RDF.”

LOD从定义上看，是将互联网打碎成无数个数据，再根据数据之间的关系相连，这样可以实现各种不同的、很有意义的应用。但是这和搜索引擎的区别在哪里呢？搜索引擎不也是通过关键词查找，能够快速从整个互联网获取资源吗？不，这并不一样，试想一下，如果我现在这个应用程序需要从互联网上获取迪士尼公司拍过的电影的名单，我该怎么办？直接Google “Movies that produced by Disneyland company” ? 显然如果机器和人一样聪明，可以从一大串字符里面，准确的提取出名单来，并且格式正确地提供给应用程序，这样当然是可行的。但是计算机没有这么聪明，互联网上也没有为每一种数据，每一串数据提供了获取数据的应用程序接口，所以我们该怎么办？答案就是LOD和语义网，我的理解是，语义网是给机器来理解的，HTML是给人来理解的，这样就相当于把网络分成两种版本，一种给机器，一种给人类，机器那种遵循着LOD的规范，这样我的应用程序就不需要通过搜索明文了，直接生成一个RDF格式的搜索命令，发到互联网上面，比如我想知道迪士尼公司拍过那些电影，我就可以先寻找Disneyland company这个数据串，再根据RDF和我想知道的数据之间的关系，进行查询，这整个过程就很直接，我通过Disneyland词条Link到Movie那里，或者可以通过Movie词条link到Disneyland那里，这就是所谓的，Linked Data概念。

Challenges and Opportunities

Due to the massive growth of available data, conventional methods of data integration are bound to fail while the complexity of processes within organizations ask for more agile options to link and mash-up data in a qualified way. Availability and matching of diverse data sources become more crucial and therefor the need for standards-based tools is growing. 传统整合数据的方式会失效，当数据复杂度急剧加快的时候

In most cases public sector information is not published in a machine-processable format that would allow data re-users from the public and private sector to automate combining public data with other public or proprietary data sources (Analysis Report of Public Sector Data and Knowledge Sources^ 2012). 这段我觉得阐述的是从方法论的层面解释Linked Data存在的必要，因为现阶段的框架并不支持语义网

The vast majority of Public Data sources do not provide the datasets in a standard format which would support true semantic enrichment and interlinking of data (such as Resource Description Framework, RDF). 承接上一段

During the last years, significant research activities have appeared that focus on industrial relevant scenarios, such as the LATC (LATC Project 2016) and the LOD2 (LOD2 Project 2016) projects that aim to contribute high-quality interlinked versions of public semantic web datasets and promoting their use in new cross-domain applications by developers across the globe (Verborgh et al. 2014). In the context of these efforts and emerging tools, while there is considerable support for linked data in other issues, such as storage (Virtuoso, Sesame), linkage (Silk Framework), discovery and publishing (SPARQL standard) and even visualization of RDF graphs (LodLive, CubeViz), there are very limited options for renovating existing data into Linked Data. Currently available solutions either support specific structured data formats, such as spreadsheets (XLWrap) and relational databases (D2R, Triplify) or provide RDF representations of data for specific sources (DBpedia). Lastly, most existing work related to exploring and visualizing RDF is limited on concrete domains and concrete datatypes and is mainly focused towards academic researchers that are familiar with the semantic web technologies.

Despite the envisioned benefits of Linked Data, a number of significant challenges and issues prevent the massive adoption of the Linked Data tools presented in the previous section by public data providers. 局限性

Linked Data-The Story So Far

With significant volumes of Linked Data being published on the Web, numerous efforts are under­way to research and build applications that exploit this Web of Data. At present these efforts can be broadly classified into three categories: Linked Data browsers, Linked Data search engines, and domain-specific Linked Data applications. 分类