World Bank Open Data is a website of The World Bank which provides free and open access to global development data in many indicators such as agriculture, education, environment and health. Similar to Wikipedia, World Bank Open Data contains tremendous amount of data from various areas. However, its main data is numerical digits with labels while types of information on Wikipedia are more various. For example, indicator of “Forest area” contains data of forest area for every country and the whole world, and data could be export as XML, CSV or Excel format file. Although data on World Bank Open Data is of great importance in many fields such as academic research, financial analysis, the operation of export or visualization of data still requires human to complete, which means that computer itself is hard to understand the data behind these tables.

The main idea of this essay is to transfer data on World Bank Open Data into Linked Data and link them with corresponding data in DBpedia, so that more powerful dataset could be available for application development. As it is stated in literature review, DBpedia is the fundamental of so many applications of Linked Data technology, and the first step of these applications is to retrieve required data from DBpedia. Therefore, if data on World Bank Open Data could be extracted in RDF format and link them to DBpedia, not only the content of DBpedia will be tremendously enrich, but also applications which based on DBpedia has opportunities to enhance their functions and enrich content.

One important feature of World Bank Open Data is that it manly contains numerical data such as GDP values of countries in the world, and those numbers are authoritative since that they are the result of government statistical departments. In traditional search engine, when users try to search for numbers for a specific item such as “China GDP in 2018”, the search engine mostly just compares the keywords with context in HTML files and return bunch of different relative web pages rather accurate numbers. The reason for this result is that machine cannot understand the meaning of those numbers and if authoritative numerical data could be readable by computers, the search process would be efficient and accurate. The improvement of accuracy of this kind of problems could be helpful for academic purpose and numerical data related fields.

The concept of Link Data search engines is conceived since the birth of Linked Data and there are already some good examples of Linked Data search engine such as Sig.ma, Falcons and SWSE(Linked Data: Evolving). These Linked Data search engines provide key-word based search which is designed for human-computer interaction, same as current prevalent search engines such as Google and Bing, while the content returned by Linked Data search engines is much richer than traditional search engines. However, neither traditional nor Linked Data search engines provide powerful function for numerical data search, computer cannot understand such question: “GDP values of China from 2010 to 2018”, since that DBpedia does not contain enough numerical data for computer to understand and select accordingly.

To develop a Linked Data search engine with the ability to handle numerical requests, steps of developing a Linked Data application mentioned in literature review could be followed.

Accessing the Web of Data

Before this application could access the structured data on DBpedia, data on World Bank Open Data should be extracted firstly and then integrated into DBpedia. This process could be done easily since that data on World Bank Open Data could be exported as XML format and RDF documents are written in XML.

Vocabulary Mapping

Indicators or country names on World Bank Open Data should names after its URI in RDF and then they could be mapped to corresponding records in DBpedia. Once RDF links of World Bank Open Data and DBpedia are linked together, DBpedia is enriched and following steps (Identity Resolution, Provenance Tracking and Data Quality Assessment) are no need.

Using the Data in the Application Context

Differs from other search engines, this Linked Data search engine should firstly cache numerical data from DBpedia according to the keywords in request, then base on the predicates in request, do manipulation to cached data, finally return the result.

Increasingly web publishers upload their web pages with Linked Data and applications based on Linked data technology would be popular in the future. Adoptions of maturing technology are supporting the deployment of Semantic Web, both in commercial and public organization (Motivation for the semantic web).