****

Big data vs Semantic Web

Big Data Management: Coursework 2

5 April 2015

**Author:** Wanchana Ekakkharanon (H00148811)

**Course:** MSc Data Science

**Lecturer:**

## Big data Management

### Introduction

"Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" (Mauro, 2015)

Traditional Data management is all about organization, administration and governance of the data. The data then, will be analysed and can be used to adjust a direction of corporations, to plan a countermeasure against crimes, to share an important finding in the research areas and so on. To store all the data in a traditional data base system like Relational database involved a lot processed such as design a structure of table, cleansing and transforming the raw data which require a lot of times to do.

In the present day, the world is flooded with a digital information from various sources of electronic devices and many kind of software that were used on daily basis of world’s population lead to rapidly increasing of overall size of digital data (High Volume) and the rate of growing of data is incredibly fast (Velocity), also the data from different sources may have completely different structure even they were presenting the same topic.

### Objectives

Big data management is the organization, administration and governance of large volumes of both structured and [unstructured data](http://searchbusinessanalytics.techtarget.com/definition/unstructured-data).

The goal of big data management is to ensure a high level of [data quality](http://searchdatamanagement.techtarget.com/definition/data-quality) and accessibility for business intelligence and [big data analytics](http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics) applications. Corporations, government agencies and other organizations employ [big data](http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data) management strategies to help them contend with fast-growing pools of data, typically involving many [terabytes](http://searchstorage.techtarget.com/definition/terabyte) or even [petabytes](http://searchstorage.techtarget.com/definition/petabyte) of information saved in a variety of file formats. Effective big data management helps companies locate valuable information in large sets of unstructured data and semi-structured data from a variety of sources, including call detail records, system logs and social media sites.

### Technologies

Most big data environments go beyond [relational databases](http://searchsqlserver.techtarget.com/definition/relational-database) and traditional [data warehouse](http://searchsqlserver.techtarget.com/definition/data-warehouse) platforms to incorporate technologies that are suited to processing and storing non transactional forms of data. The increasing focus on collecting and analysing big data is shaping new platforms that combine the traditional [data warehouse](http://searchsqlserver.techtarget.com/definition/data-warehouse) with big data systems in a logical data warehousing architecture. As part of the process, the must decide what data must be kept for compliance reasons, what data can be disposed of and what data should be kept and analysed in order to improve current business processes or provide a business with a competitive advantage. This process requires careful [data classification](http://searchdatamanagement.techtarget.com/definition/data-classification) so that ultimately, smaller sets of data can be analysed quickly and productively.

## Semantic Web Technologies

### Introduction

### Objectives

### Technologies

## Relationship of Semantic Web and Big data

## Review

In Life Science researches, many research produced high volume of the data in very fast speed than before. It was resulted from an advancement of technologies (Computer). In the first 6 months, a project like The 1000 Genomes Project produce data more than the previous last 30 years. Many organizations (Data source) were established to store all of these kind of data, but each organization did not have the same data structure e.g. text file, xml and relational database system.

Usually, data analysis across multiple data sources and related data sources are required in order to achieve the aim of the project. However, different data sources have different data structures, even the data are in fact the same thing but they may have completely different field name which is hard to retrieve all related data, remove inconsistency data and analyse data.

Combination of Semantic Web technologies provided some solution to Big Data problems. The organization such as The Open Biological and Biomedical Ontologies Foundry established a standard ontologies suitable for biomedical domain which allowed the data from multiple sources to be shared and analysed. Some provided the tools to search a public ontologies that match research topic.

## References

**Ashutosh** A., 2012, Best Practices For Managing Big Data [online], available: <http://www.forbes.com/sites/ciocentral/2012/07/05/best-practices-for-managing-big-data/>

Mauro DA., Greco M., Grimaldi M., 2015, What is big data? A consensual definition and a review of key research topics, *AIP Conference Proceedings*