

# **NLP and Semantic web/data technologies in Advancement of Business Intelligence**

---

**Abayarathna A.H.M.S.P (E/09/001)**

**Ilangakoon I.A.M.M.K (E/09/118)**

**Kuruppu K.A.D.I.M (E/09/198)**

<http://www.ce.pdn.ac.lk>

Department of Computer Engineering  
Faculty of Engineering  
University of Peradeniya  
Peradeniya 20400  
Sri Lanka



## **Abstract**

Business Intelligence is the future of efficient management and improves profitability in today's industry. Semantic web technologies provide much flexible knowledge gathering and processing platform on data. Data can be gathered and processed using various available technologies to find relationships base on reason. This knowledge can be stored and processed using RDF/OWL technologies and can quire by languages like SPARQL with available technology frameworks. Processing large amounts of unstructured data can be quite computation intensive process. This challenge can be overcome by parallelism and using techniques like Google Map Reduce.

Further by including Natural Language Processing the information analysis can be greatly eased out for the end user of the system. This will help the end users to make quick decisions which can lead to profitability in an organization.

## **Contents**

Abstract	i
Content	ii
Chapter 1 Introduction and Motivation	1
Chapter 2 Related Work	2
Chapter 3 Project Objective	3
Chapter 4 Methodology	4
Chapter 5 Work Plan	5
References	6

# **Chapter 1**

## **Introduction and Motivation**

By having a business intelligence system in a business organization can assist people in the organization to make decision on the fly.

Business intelligence (BI) can be defined as the process of finding, gathering, aggregating, and analyzing information for decision making .Semantic technologies of the type advocated by Semantic Web are being applied for BI in the context of the Project. We are developing a new generation of BI tools and modules based on semantic-based knowledge and natural language processing (NLP) technology to mitigate the efforts involved in analyzing information.

Data can be gathered and processed using various available technologies to find relationships base on reason. This knowledge can be stored and processed using RDF/OWL [5] technologies and can quire by languages like SPARQL [4] with available technology frameworks. Processing large amounts of unstructured data can be quite computation intensive process. This challenge can be overcome by parallelism and using techniques like Google Map Reduce.

Further by including Natural Language Processing the information analysis can be greatly eased out for the end user of the system. This will help the end users to make quick decisions which can lead to profitability in an organization.

## **Chapter 2**

### **Related Work**

One of the researches was done to Add semantics to Business Intelligence.[1] Despite the importance of analytical tools to organizations, they still lack the inference power needed to solve the requests of decision makers in a flexible way. Their approach aims at integrating business semantics into analytical tools by providing semantic descriptions of exploratory functionalities and available services.

The researches in [2] presented the Ontology-based Information Extraction for Business Intelligence. It describe the application of ontology-based extraction and merging in the context of a practical e-business application for the EU MUSING Project where the goal is to gather international company intelligence and country/region information. The results of its experiments so far are very promising and they are now in the process of building a complete end-to-end solution.

The researches in [3] presented the A Framework for Business Intelligence Application using Ontological Classification. Every business needs knowledge about their competitors to survive better. One of the information repositories is web. Retrieving Specific information from the web is challenging. An Ontological model is developed to capture specific information by using web semantics. From the Ontology model, the relations between the data are mined using decision tree. From all these a new framework is developed for Business Intelligence.

## **Chapter 3**

### **Project Objective**

The main objective of our project is to do a comprehensive analysis of semantic web and data technologies in making structured data more meaningful for an organization and also find tools and their capabilities in delivering the above requirements. Then use NLP in generating queries on semantic data space to create working software.

## **Chapter 4**

### **Methodology**

In this Project we are going to develop business Intelligence software using semantic web and Natural Language Processing. By having a business intelligence system in a business organization can assist people in the organization to make decision on the fly. In this project we are going to use Agile Methodology for developing the software. Also we have planned to maintain a blog about the work carried by each group members. First of all, we are going to study the technologies we can use in this project ( RDF , OWL , SPARQL etc).

Then we have to create a dummy relational database to use in this project as the data. Then this database is converted to Resource Description Framework (RDF). In order to this we will have to find library to convert the relational database to RDF. If not we have to develop a library to do this work. When we are able to convert the relational database to RDF we can extend this process to gather data from other sources such as web resources, email to RDF.

After creating the RDF file we are going to use SPARQL queries to get data from RDF file. In order to do that we will have to study about the SPARQL and find libraries to query RDF file using SPARQL in Java.

But the business people don't know SPARQL. So we are going to use Natural Language Processing (NLP) to convert Natural language inputs to relevant SPARQL queries and get the output.

After finishing these initial work we have plan to extend the project to use part speech to get the outputs.

## Chapter 5

### Work Plan

Activity / Week	1	2	3	4	5	6	7	8
Study About the Project								
Create the Project Proposal								
Proposal Presentation + Report								
Create Relational Database								
Study about converting Relational Database to RDF								
Implementation of Converting Relational Database to RDF								
Final Report								

Activity / Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Study about SPARQL															
Implementing SPARQL queries to get output and do additional Study															
Mid Semester Exam															
Study about the NLP and how to use it in the project															
Implementing NLP and do additional learning															
Final Report															



## Reference

- [1] Denilson Sell , Liliana Cabral , Enrico Motta , John Domingue and Roberto Pacheco Stela Group, Universidade Federal de Santa Catarina, Brazil Knowledge Media Institute, The Open University, Milton Keynes, UK INE, Universidade Federal de Santa Catarina, Brazil Adding Semantics to Business Intelligence.
- [2] Ontology-based Information Extraction for Business Intelligence. Horacio Saggion and Adam Funk and Diana Maynard and Kalina Bontcheva Department of Computer Science, University of Shffield Regent Court, 211 Portobello Street, Sheffield, S1 4DP United Kingdom
- [3] A. Martin , D.Maladhy , Dr . V . Prasanna Venkatesan: A Framework for Business Intelligence Application using Ontological Classification, 2011 IJEST
- [4] ARQ. (n.d.). A SPARQL processor for Jena. Retrieved from <http://jena.sourceforge.net/ARQ/>
- [5] Brickley, D., & Guha, R. V. (Eds.). (2004). RDF vocabulary description language 1.0: RDF schema. Retrieved from <http://www.w3.org/TR/rdf-schema/>