

MineYourText: Bridging the Gap between Natural Language and Linked Data

Parma Nand[‡] and Rivindu Perera[§]

Mine Your Text Ltd.
Auckland University of Technology
Auckland, New Zealand
{pnand,rperera}@mineyourtext.com

Abstract. This paper explains the idea behind MineYourText, a spin-off startup company arising out of NLP research at Auckland University of Technology. The objective of the Mine Your Text is to provide services and tools to bridge the gap between Semantic Web and natural language. The company currently hosts two projects; RealText_{lex} and KiwiLOD.

1 Introduction

Semantic Web has become a key, and a critical component in business computing [1], scientific research [2], strategy planning [3] as well as aspects of everyday normal life [4]. In order to support the unbelievable reliance on the Semantic Web, there is a crucial need to sustain the growth in order to be able to prevent a “Semantic Bubble burst”. Data on the web is building up at a cosmic rate from the fast pace of research and accumulation from collection endpoints. Since it is impossible to process this amount of data using human intervention and if the semantic web is not sustained at the same rate, the usefulness of the data will start to taper off. Even at the current time Semantic Web is lagging behind in terms of transforming millions of web documents into data web so that it can be directly or indirectly processed by machines. Hence it is vital that the pace of Semantic Web development is sustained at the same rate as the web development.

MineYourText was formed as a spin-off targeting the sustainability of the Semantic Web based on two successful research projects fully funded by Auckland University of Technology. Our first project, RealText_{lex}, focuses on lexicalizing the Semantic Web, so that Semantic Web triples can be selected from a triple store and formulated by a machine as natural text. We have employed both Natural Language Generation (NLG) and Semantic Web technology stacks to work in parallel in the RealText_{lex} framework. This project enables one to increase the usability of the current Semantic Web by presenting its processed, encoded knowledge in a natural form to a wider audience. Our second project focuses on extracting triples from natural text to build new Linked Data resources, which

[‡]<http://people.aut.ac.nz/~pnand/>

[§]<http://people.aut.ac.nz/~rperera/>

will eventually support the transformation of the existing document web to the data oriented Semantic Web, and hence, enable it for much wider use as it will enable it be processed by a machine.

2 The overview of the Startup

MineYourText* is a registered company in New Zealand formed in 2016. We have defined the business scope as the intersection of the Semantic Web and Natural Language Processing. The core theme of MineYourText is to make the Semantic Web more sustainable by enriching its content by adding new triples and enabling it for a wide range of applications by transforming triples to natural text. Currently, we are dealing with multiple commercial projects that integrate MineYourText products for a wider and better data infrastructure.

3 Projects

In the following sections we explain two of our projects which had lead to the formation of the MineYourText as an emerging startup company in New Zealand.

3.1 Linked Data lexicalization: Overview of RealText_{lex} Project

RealText_{lex} project [5,6] was started as an academic project which later lead to significant success over two years of continuous research. The aim of the project is to generate lexicalization patterns that can transform Linked Data triples into natural text. The framework we built in this project generates patterns using four different pattern extraction modules. The generated lexicalization patterns can be used to lexicalize given set of triples from a Linked Data resource. We have received number of requests to integrate the framework in commercial projects and the most recent application looked into utilize the framework to generate movie descriptions using DBpedia triples.

3.2 New Zealand Open Data to Linked Data: Overview of KiwiLOD Project

The Kiwilod [7] project extracts triples from the New Zealand Government Open Data. We consult both semi-structured open data as well as documents to extract the triples. The triples are organized in an ontology we built based on the ecosystem specific to New Zealand. The triples are stored in a GraphDB database and we were able to reveal several important but previously hidden relationships among data.

*<https://www.nzlbusiness.com/company/registered/Mine-Your-Text-Limited>

4 Conclusion and future work

This paper presented our Linked Data start-up proposal based on two successful academic projects which will be commercialized under the new start-up. Numerous requests that we have received to integrate these framework reflect the future success of our start-up.

References

1. Hepp, M., Leymann, F., Domingue, J., Wahler, A., Fensel, D.: Semantic business process management: A vision towards using semantic web services for business process management. In: IEEE International Conference on e-Business Engineering (ICEBE'05), IEEE (2005) 535–540
2. Knublauch, H., Dameron, O., Musen, M.A.: Weaving the biomedical semantic web with the protégé owl plugin. In: In International Workshop on Formal Biomedical Knowledge Representation, Citeseer (2004)
3. Lin, H., Harding, J.A.: A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration. *Computers in Industry* **58**(5) (2007) 428–437
4. Chen, H., Finin, T., Joshi, A., Kagal, L., Perich, F., Chakraborty, D.: Intelligent agents meet the semantic web in smart spaces. *IEEE Internet Computing* **8**(6) (2004) 69–79
5. Perera, R., Nand, P.: A Multi-strategy Approach for Lexicalizing Linked Open Data. *Computational Linguistics and Intelligent Text Processing* (2015) 348–363
6. Perera, R., Nand, P., Klette, G.: RealText lex : A Lexicalization Framework for Linked Open Data. In: International Semantic Web Conference (ISWC) - Demonstration. (2015) 1–4
7. Perera, R., Nand, P.: KiwiLOD: A Framework to Transform New Zealand Open Data to Linked Open Data. Technical report, Auckland University of Technology (03 2016)