Ontology-Based Linking of Social, Open, and Enterprise Data for Business Intelligence

Tope Omitola[1], John Davies[2], Alistair Duke[2], Hugh Glaser[3], and Nigel Shadbolt[1]

[1] Electronics and Computer Science
University of Southampton, UK
{t.omitola, nrs}@ecs.soton.ac.uk
[2] British Telecommunicatons, UK
{john.nj.davies, alistair.duke}@bt.com
[3] Seme4 Ltd., UK
hugh.glaser@seme4.com

1 Introduction

We are at the cusp of two major revolutions impacting the world of work and commerce. These are the rise of social media and the rise of Big Data. Data is everywhere. Each phone call, email, chat request, or person-to-person interaction between a customer and a brand provides organisations with invaluable information. This wealth of data can yield valuable information, such as revealing precious insight into customers' needs and desires, allowing companies to personalise their services accordingly. A business revolves around its customers and their social connections. These social connections are valuable data that can useful for enterprises.

For a company to thrive in this new world, these data need to be used to identify opportunities in new sectors, support employees, customers, and other external partners. These data can also be used to create a more intelligent understanding of customers, and to help predict future customer behaviour. This paper (and talk) will describe how we apply Linked Data to enable BT, in particular BT Business (BTB), a division of BT, to take advantage of this Data revolution. We use Linked Data technology to integrate internal and external datasets, including structured, unstructured, and social data. We describe (and shall expand on these in the talk) how this integration allows new and insightful information to be derived.

BTB, the UK's leading provider of business communications services with over one million small and medium sized enterprise customers, would especially like to use Linked Data to solve these business challenges (amongst others):

- To manage and extract value from its disparate, isolated data,
- To take advantage of information from external, non-enterprise, and other social data in order to provide new, exciting, and useful services that create value for customers,

– To identify trends and issues that are specific to circumstances such as competitor activity, products offered, industrial sector of the customer, and profiles of members of sales teams.

2 System Operation

Ontology-based data access and management (OBDM) is a methodology that is used to access, integrate, and manage data in big enterprises. It consists of a three-level architecture constituting an ontology, the data sources, and the mapping between the two. We have applied the ideas of OBDM for our data integration process. Our systems consists of a Unified Ontology, and using this ontology to guide us to transform the datasets of four systems, into Linked Data.
a. The Unified Ontology: This consists of entities such as the concepts of a BT employee (BTEmployee), of a BT employee (EmployeeSocialMediaUser) that is also a member of a social network, of sales team members (SalesForceUser), of client companies (Account), and industrial sectors (SICCode). (We will expand on this in the talk).
b. Four different systems are involved in the process: (1) Public information of clients of BTB as derived from OpenCorporates; (2) Members' social connections as derived from LinedIn; (3) an LDAP-backed General Employee Data store; and the BTB Win-Loss system (which stores a collection of CSV files of companies that clients of BTB). We transform most of the data in these systems into RDF, linking them together using appropriate class and data instances' URIs.

Data Consumption The linked data platform that is constructed allows us to make very flexible Sparql queries, which we shall describe further in our talk at the Workshop.

3 Conclusions

This paper described how we linked social, open, and enterprise data for Business Intelligence, and how this has been applied in a telecommunications company. Some of the challenges we found include (a) the establishment of a strong business case and the availability of a 'data champion' to help bring the stakeholders together; (b) Data Discovery and Provenance. Discovering the appropriate datasets with the right provenance criteria can be time-consuming but very important; (c) Data Cleaning and Interlinking. Many of the discovered datasets may not be in the appropriate format to be useful, so they need to be "cleaned". The choice of URIs to use for interlinking datasets is application-specific. This choice should be guided by the business case; and (e) Data Modelling. A goal of integration is to have a unified view of the data from the disparate sources. Having a unified ontology helps towards this. Future data integration tasks need to think of these challenges and provide the appropriate solutions. We shall provide more details of these challenges in the talk.