

AN INTERACTION FRAMEWORK FOR BUSINESS INTELLIGENCE

Industrial thesis: CIFRE/SAP, EURECOM and EDITE doctoral school

Author: Ahmad ASSAF

Doctoral advisor: Raphaël Troncy

SAP supervisor: Aline SENART

August 26, 2013

1. Introduction

This report aims to give a brief overview of what has been done for our second year of PhD. We joined SAP Research in May 2012 in order to start a PhD thesis with SAP AG and EURECOM while being registered in the EDITE doctoral school in Paris.

Business intelligence (BI) mainly refers to computer-based techniques used in identifying, extracting, and analyzing business data, such as sales revenue. In BI, business users generally perform four types of processes: Data Selection, Data Manipulation, Data Analysis, and Data Visualization. In the early months, the objective of our research was broadly to provide a framework that enables users to have a better interaction with the different BI processes by tackling the current limitations that affect user experience. Due to several organizational changes inside of SAP, the nature of our research has now focused on the core data integration and information retrieval problems.

The main purpose for our research is to provide a framework that will help in enriching Business Intelligence with Semantics. This can be done by performing several mining techniques on the data itself, usage of annotation and semantic modeling in order to mash-up external related data. This will allow users to have better insights and analysis of their data, which will lead to more efficient and accurate decisions.

2. Second year activity

For the second year of this PhD, we have performed a literature survey spanning different fields ranging from Business Intelligence, Data Analysis and Data Integration in the Semantic Web. Doing so helps in enriching our overall knowledge about these fields which helped us in defining more our research scope.

We have participated in an internal project at SAP called **Remix**. Remix is a business intelligence solution that helps business users intuitively and quickly builds insightful reports. It enables the composition of existing BI artifacts with new data from the enterprise and from external sources. It is a self-service BI tool targeted at business users. It recommends the best course of action by leveraging content and interaction traces. Remix offers the following key features:

- Self-service BI for external data: remix enables mash-ups of enterprise data and external data with no common semantics, in particular web data feeds (e.g. OData, Linked Data Cloud).

- What you draw is what you get: With remix, business users don't need to write queries but simply graphically draw what they have in their mind. It is a direct and intuitive way to build reports without requiring the user to understand IT concepts
- Guided interaction: remix provides contextualized recommendations to the business users to increase quality when possible, suggest new insights and save them valuable report design time. Remix helps the business user to find relevant data sources, formulas, visualization, and assists in the data reconciliation process.
- Collaboration and sharing: Business users are able to share their data sets and report parts allowing other users to quickly build new reports.

The contributing teams were a five people team in Mougins, France and a four people team in Dresden, Germany. The technical development of Remix followed the Scrum methodology which is an iterative and incremental agile software development method for managing software projects and product or application development.

I was the UI/UX lead for this project with contributions in designing the data model, implementing the mash-up services in the back-end and developing the whole front-end of the product taking into consideration different user experience problems and adapting the design for mobility.

Our research has led to two papers this year:

1. SNARC: A Semantic Social News Aggregator, AI Mash-up Challenge at ESWC13, Montpellier, France 2013

Abstract:

The Internet has created a paradigm shift in how we consume and disseminate information. Data nowadays is spread over heterogeneous silos of archived and live data. People willingly share data on social media by posting news, views, presentations, pictures and videos. Social media has become a source of live and up-to-date knowledge; a property that was lacking in traditional sources. SNARC is a service that uses semantic web technology and combines services available on the web to aggregate social news. SNARC brings live and archived information to the user that is directly related to his active page. The key advantage is an instantaneous access to complementary information without the need to dig for it. Information appears when it is relevant enabling the user to focus on what is really important.

2. SNARC: An Approach for Aggregating and Recommending Contextualized Social Content, Proceedings ESWC 2013, Montpellier – France

Abstract:

The Internet has created a paradigm shift in how we consume and disseminate information. Data nowadays is spread over heterogeneous silos of archived and live data. People willingly share data on social media by posting news, views, presentations, pictures and videos. SNARC is a service that uses semantic web technology and combines services available on the web to aggregate social news. SNARC brings live and archived information to the user that is directly related to his active page. The key advantage is an instantaneous access to complementary information without the need to dig for it. Information appears when it is relevant enabling the user to focus on what is really important.

Our participation entitled “*SNARC – A Semantic Social News Aggregator*” won the first prize at the AI Mash-up challenge in ESWC13. We have also participated at the 10th Summer School on Ontology engineering and the Semantic Web (SSSW 2013) in Madrid, Spain, <http://sssw.org/2013/>.

3. Conclusion and Future Work

We have identified three main areas to focus on for this third year:

1. Enhancing Schema Matching with Linked Data (RUBIX Continuation)

We will proceed to enhance our previous publication about this topic and ensure the integration of the algorithms and design into SAP VI. It has been a great advantage to have this work identified and being pushed toward integration with an existing application. There are several engineering and research challenges that we aim to tackle:

- The integration of external Linked Data Repositories is a core issue. The reason behind that is that we don't want to keep calling REST APIs (e.g. of Freebase), but instead, we want to have the data in house whether this is going to be in HANA (AIS which is the Graph store on HANA) or Jena SDB in HANA or to have a separate triple store such as Virtuoso.
- Plug with dbpedia live and Freebase live, so that we can monitor changes in those datasets to ensure that we keep having live and fresh data.
- Building a disambiguation search function on the data imported. We aim to simulate what Freebase REST API does by sending a search term and retrieving a set of possible concepts that are relevant for this term.
- Implement the actual algorithms RUBIX uses in the HANA engine itself.

2. DCAT Crawler

The idea is to crawl current Data Catalogues services and enhance the existing DCAT representation for datasets. We aim to look up, mine and apply some of the techniques used in RUBIX to give more detailed representation of the data. Further, we plan to adapt our proposed Data Quality Framework and augment this with the DCAT, come up with metrics to measure the quality of the data and the data source.

3. Visual Mapper

We have created a visual general purpose mapping tool that eases the process of mapping items. It accepts a mapping file represented in a JSON format and provides an intuitive user friendly interface for the visualization. We still need to find out about:

- Possible evaluation methods and techniques.
- Work on real data sets (this can be done by writing some specific translation scripts to convert a mapping file to our desired JSON format).
- Investigate the integration with the Datalift project.
- As another possible use case, try to adapt this tool for automatic link discovery.

We have identified a publication plan to target the work done on the SSSW13 summer school as part of the mini project as well as the invitation we got to extend our previous at DQSMT12 for the Special Issue on Web Data Quality of the International Journal on Semantic Web and Information Systems (IJSWIS).