

Author: John Walker, Tim Nelissen

Date: October 2013

Thesis projects Semaku B.V.

Semaku is developing together with a client – NXP Semiconductors – a corporate light weight Semantic Platform. This Semantic solution enables corporations to unleash their content silo's in a flexible way and allow many meaningful (new) data relations. High quality standards are set for data entry into the central data repository. The quality check will be done via data dictionary, provenance and process validation. For the central repository Semaku has chosen for RDF statements (in Dydra) and OWL-principles and for it's application framework SPARQL queries and the Graphity framework.

The described thesis subjects arise from the need to further improve data quality and reliability within the complete process flow. Therefore these subjects are seen as crucial for further development of the corporate Semantic Framework. They also will be based on a real use case (NXP Semiconductors).

In summary the three defined problem statements for the thesis cover areas like: data transformation, data modeling plus validation and process optimization. These subjects are described in detail in the next section. The subjects, scope and complexity of the problem statements will fit within the given Master research project period.

1. Data transformation

The goal of this thesis is to define an optimal standard data transformation process. The aim is to use as much as possible a “standardized” transformation mechanism - data connector and preferably a model driven change propagation process. In the electronic industry and even beyond corporations are using more or less the same domain solutions. The scope of this thesis will include one or more of the following source environments:

1. a Product Lifecycle Management (PLM) - Enovia (Dassault)
2. a Content Management System (CMS) – Alfresco
3. IBM Rational DOORS (Dynamic Object Oriented Requirements System)

The data transformation process will be added as a standard service to the Semantic platform. It will become as such part of the core Base Module functionality.

The primary audience for this research project is computer science students, with an interest in software engineering, information and knowledge management and/ or distributed systems. Some of the subjects relevant to this project are:

- Data analyses and integration
- Resource Description Framework (RDF)
- Query language for RDF (SPARQL)
- Web services (SOAP and REST)
- JSON and or xslt
- Java, Maven and Spring
- Bootstrap framework (html5 and CSS3)

2. Data modeling

The goal of this thesis subject is to define an optimal mix for data quality validation when transporting data from source environments to the meta data “cloud” / RDF Triple store. These validations are planned to be added as a core functionality to the Base Module and will be presented in a Dashboard.

The graduation student will perform a profound theoretical and practical analysis of best practice solutions for:

- Dictionary versus target data validation methods
- Data provenance validation methods

The Dictionary source is in our use case stored in a Product Information Management environment named eptos (Paradine). The idea is that before loading the source information into the Triple store the data will be validated via a light weight diff functionality against the dictionary. The results should be presented in a Dashboard. Next to proper terminology management the data quality in the Triple needs to be measured against the provenance information. We are looking for a best practice light weight provenance data validation diff and temporal reasoning solution. The results should be presented in a Dashboard.

The primary audience for this research project is the group of computer science students, with an interest in software engineering, information and knowledge management and/ or distributed systems. Some of the subjects relevant to this project are:

- Data analyses and integration
- Data ontologies e.g. OntoMI, OWL
- Resource Description Framework (RDF)
- Java, Maven and Spring

- Bootstrap framework (html5 and CSS3)

3. Process optimization

The goal of this thesis subject is to define a generic strategy for modeling and conversion of data into RDF. What are the pro's and con's for positioning the modeling process at source or destination location or even in between in the enterprise services environment.

The content source environments which are looked at are:

1. the Product Lifecycle Management (PLM) environment - Enovia (Dassault)
2. the Content Management System (CMS) – Alfresco
3. IBM Rational DOORS (Dynamic Object Oriented Requirements System)

The idea is that the modeling process is flexible and can be added as a standard service, component to the Base Module framework. Besides that the idea is that the model should also drive the visual representation of the data in the search and relational browser area of the Base Module.

The primary audience for this research project is computer science students, with an interest in software engineering, information and knowledge management and/ or distributed systems. Some of the subjects relevant to this project are:

- Data analyses and integration
- Data ontologies e.g. OntoMI, OWL
- Resource Description Framework (RDF)
- Java, Maven and Spring
- Enterprise services environment (Tibco, etc..)
- Bootstrap framework (html5 and CSS3)

About Semaku

Semaku B.V. is a startup company offering Product Information Management (software) solutions and services. Our ambition is to become a market leader in integrated Product Information Management Services and Solutions. Our ambitious and enthusiastic team will support you during the thesis period.

For your thesis you will also have direct access to highly skilled people and organizations in the area of Semantic Web and Product Information Management services. Short and informal communication lines are common practice within our organization.

Practical information

Semaku will offer a competitive placement fee.

You will be located at our Strijp-S office in Eindhoven and will also regularly visit NXP Semiconductors at the High Tech Campus in Eindhoven.

As we are looking for talent in the earlier described area's there might be a good job opportunity after finalization of the thesis.