

Auteur: Mike Henrichs / Nic Roest

Datum: oktober 2012

# Afstudeeronderwerpen Semmtech

---

Semmtech ontwikkelt momenteel een platform voor Semantic Web toepassingen gestoeld op RDF/OWL en Jena. SEMMweb™ is in essentie een generiek platform voor het beheren en ontsluiten van semantisch rijke informatie. Gebruikers van dit platform worden ondersteund bij het vastleggen van informatie middels RDF statements, hierbij worden de RDF en OWL regels nageleefd. Tevens worden gebruikers ondersteund bij het online delen van deze informatie als ook het gebruiken van deze informatie in toepassingen van derden.

De volgende onderwerpen komen voort uit de wens om allereerst modellen semantisch rijker te maken, en daarnaast uitwisseling van deze semantische modellen te bevorderen.

De onderwerpen die in dit document worden beschreven kunnen binnen Semmtech in de vorm van een afstudeeropdracht voor een WO/Master student worden uitgevoerd. De complexiteit van en looptijd moeten redelijkerwijs passen binnen de beschikbare tijd voor zo'n afstudeeropdracht.

---

## 1. SPARQL Query Builder

Dit onderwerp heeft tot doel het kunnen definiëren van SPARQL-queries middels een grafische user interface. Hierbij zullen reeds bestaande top ontologieën dienen als input - en initiële validatie - van de query. De applicatiemodule stelt gebruikers in staat om een query te bouwen die ontologieën bevraagt waarbij de volledige mogelijkheden van een SPARQL query kunnen worden benut (denk aan CONSTRUCT, FILTER en functions). Uitgangspunt hierbij is dat de gebruiker geen kennis heeft van de syntax van SPARQL-queries. Wel mag ervan uit worden gegaan dat de gebruiker bekend is met RDF/OWL en met de domeinspecifieke inhoud van de te bevragen ontologieën.

De afstudeerder wordt geacht de te definiëren SPARQL-queries te ontleden naar voor de gebruiker logische blokken in voor de gebruiker begrijpelijke terminologie. Dit vormt de basis voor het ontwikkelen van één of meerdere intuïtieve gebruikersinterfaces. Voor zover mogelijk en relevant kan hierbij gebruik worden gemaakt van semantische modelleer principes. Eventuele visuele integratie met te bevragen ontologieën lijkt wenselijk.

SPARQL-queries welke met de Query Builder zijn gemaakt moeten vervolgens ofwel direct kunnen worden geëxecuteerd – gebruik van bestaande tools is geen probleem –, dan wel kunnen worden opgeslagen voor later gebruik. Hierbij moeten queries kunnen worden uitgevoerd op lokale RDF/OWL bronnen of online via bijvoorbeeld SPARQL End-Points. De output uit direct geëxecuteerde queries moet zowel op het scherm zichtbaar kunnen zijn, als ook kunnen worden opgeslagen voor distributie en/of later gebruik.

Doelgroep zijn met name studenten die afstuderen op het gebied van computer science, en dan specifiek in de richtingen van software engineering, informatie- en gegevensbeheer of communicatiesystemen.

Mogelijke onderwerpen relevant voor dit onderzoek:

- Resource Description Framework (RDF)
- Query Language for RDF (SPARQL)
- Optioneel: SPARQL Inference notation SPIN
- Database performance technieken
- Web technieken/protocollen, zoals bijvoorbeeld SPARQL End-Points
- Visuele presentatie van queries en RDF-modellen (graph-based models)

---

## 2. Resource Density-coëfficiënt

Dit onderwerp heeft tot doel het geautomatiseerd kunnen interpreteren van de (relatieve) toegevoegde waarde van een 'resource' in een semantisch netwerk middels een zogenaamde 'density-coëfficiënt'. Met deze coëfficiënt moeten modelleers en/of beheerders van semantische modellen meer grip op het gebruiksnut van individuele resources in een semantisch netwerk krijgen. Ook kan deze coëfficiënt bij zoekacties worden gebruikt om mogelijke resultaten ten opzichte van elkaar te ranken (eg. vergelijkbaar als Google Pagerank).

De afstudeerder wordt ten eerste geacht inzicht te verschaffen in mogelijke toepassingen van de Resource Density-coëfficiënt. Rekening houdend met deze toepassingen zal een generiek mathematisch algoritme ontwikkeld moeten worden.

Vervolgens moet de afstudeerder het algoritme implementeren in een online inzetbare webcrawler. Deze webcrawler is in staat om nieuwe URI's te ontdekken uit bestaande bronnen. Ook moet er nagedacht worden over hoe (nog) niet-geïndexeerde ontologieën via reeds bekende en geanalyseerde ontologieën benaderd kunnen worden.

De webcrawler applicatie bestaat minimaal uit een beheerderinterface voor configuratie, een schedulingmechanisme, de webcrawler zelf en een beheerderinterface voor visualisatie en export van resultaten. Tevens is het een optie ten behoeve van optimalisatie van serverbelasting een mechanisme in te bouwen voor het creëren van locale kopieën van ontologieën om deze later te kunnen verwerken.

Doelgroep zijn met name studenten die afstuderen op het gebied van toegepaste wiskunde en/of computer science, en dan specifiek in de richtingen van software engineering, informatie- en gegevensbeheer of communicatiesystemen.

Mogelijke onderwerpen relevant voor dit onderzoek:

- Complexe mathematische algoritmes
- Webcrawler technologie
- Semantisch modelleren (RDF/OWL)

---

## 3. Semantische Rekenmodellen

Dit onderwerp heeft tot doel het modelleren van rekenkundige formules binnen een semantisch model. Nadat deze formules zijn vastgelegd in een model, zullen ook berekeningen met deze

formules kunnen worden uitgevoerd, dit gebruikmakend van concepten binnen dit model. Voorbeelden zijn het maken van kostencalculaties, of meetkundige berekeningen voor het vinden van oppervlakte en inhoud van concepten uit het model.

Hierbij zullen dus ook de waarden behorende bij de eigenschappen van klassen en instanties, als ook hun eenheden, in een model op een computer interpreteerbare manier moeten worden vastgelegd. Deze geconceptualiseerde waarden moeten vervolgens als input kunnen worden gebruikt tijdens de berekening van de rekenkundige formules.

Doelgroep zijn met name studenten die afstuderen binnen computer science, en dan specifiek op het gebied van formele modelleer talen en/of informatie- en gegevensbeheer.

Mogelijke onderwerpen relevant voor dit onderzoek:

- Semantisch modelleren (RDF/OWL)
- Modelleren van eigenschappen, waarden en meeteenheden
- Modelleren van rekenkundige formules en de berekeningen van deze formules
- Query Language for RDF (SPARQL)

Author: Mike Henrichs / Nic Roest

Date: October 2012

# Graduation Topics Semmtech

---

Semmtech is currently developing an application-platform for Semantic Web based on the RDF/OWL standards and Jena. SEMMweb™ is essentially a generic platform for maintaining and sharing semantically structured information. Users of this application are supported during the creation of RDF statements and resulting models, while adhering to the RDF and OWL-principles and rules. Users are also supported in online sharing of this information as also in applying this information in third-party applications.

Each of the following research topics stem from the desire to first of all be able to build semantic rich models, and secondly to be able to exchange or share these models.

The projects described in this document could be performed at Semmtech. The complexity and runtime of a single project should fit within the available time of a Master student's research project.

---

## 1. SPARQL Query Builder

The goal of this graduation project is to be able to define SPARQL-queries by means of an intuitive graphical user interface. Already existing top-ontologies will serve as input – and initial validation – of the query. The Query Builder application module shall facilitate users to define a query, covering the full extent of the SPARQL-syntax (e.g. CONSTRUCT, FILTER and functions). Targeted users will not have any knowledge in creating SPARQL-queries. It is expected though they are familiar with RDF/OWL as also with the domain specific content of the queried ontologies.

The graduation student shall decompose SPARQL-queries in for the user logical building blocks using understandable terminology. These building blocks will be the starting point for the design and development of one or more intuitive user interfaces. If relevant and of added value the SPARQL-queries maybe described using semantic modelling principles. Possible visual integration with the contents of relevant ontologies when building a SPARQL-query seems interesting.

SPARQL-queries created using the Query Builder shall be directly executable – using existing tools is fine with us –, or can be archived for later use. It should be possible to run queries on local RDF/OWL sources, or online via for example SPARQL End-Points. The output of an executed query may be visualised on screen, or can be stored for distribution and/or later use.

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:

- Resource Description Framework (RDF)

- Query Language for RDF (SPARQL)
- Optional: SPARQL Inference notation SPIN
- Database performance techniques
- Web techniques/protocols, like SPARQL End-Points
- Visual presentation of queries and RDF-models (graph-based models)

---

## 2. Resource Density-coefficient

The goal of this graduation project is to be able to rate the (relative) value of a 'resource' in a semantic model by means of a so called 'density-coefficient'. This coefficient shall ontology provide modellers and/or administrators more insight in the intensity of use of an individual resource. This coefficient may also be used to rank search results when searching for resources matching certain characteristics (like a Google Page rank).

The graduation student first of all needs to research the possible use of such a Resource Density-coefficient. Taking these use cases into mind a generic mathematical algorithm shall be developed.

As a next step the algorithm needs to be implemented in an online web crawler. The web crawler shall explore a provided set of ontologies for their included resources (identified by their URI's). It should also be possible to find and explore non-indexed ontologies via references in the already known and analysed ontologies.

The web crawler-application is build around an administrator user interface for configuration, a scheduling mechanism, the web crawler itself and an administrator interface for visualisation and export of results. For sever performance optimization it is also preferable to develop a mechanism to create local copies of ontologies for postponed indexing.

This subject is targeted at students in the field of applied mathematics and/or computer science, and more specific in the area of formal modelling languages and/or information- and knowledge management.

Some of the subjects relevant to this project are:

- Complex mathematical algorithms
- Webcrawler technology
- Semantic modelling (RDF/OWL)

---

## 3. Semantic Mathematical Models

The goal of this graduation project is to be able to model basic mathematical operations and formulas within a semantic model. Examples of the types of calculations are cost calculations of activities, or geometrical calculations for finding areas and volume of physical objects. After conceptualizing these formulas, the modeled calculations can be automatically performed, using the concepts described by the model.

Apart for formulas, the quantities of the various characteristics of classes and individuals, as well as their units of measure, need to be modelled in a computer interpretable fashion. Next these conceptualized quantities can be used as the input of the mathematical operations described by the formulas.

This subject is targeted at students in the field of computer science, and more specific in the area of formal modelling languages and/or information- and knowledge management.

Some of the possible topics:

- Semantic modelling (RDF/OWL)
- Modelling of characteristics, qualities and their units of measure
- Modelling of mathematical operations and formulas
- Query language for RDF (SPARQL)
- SPARQL Inference Notation (SPIN)