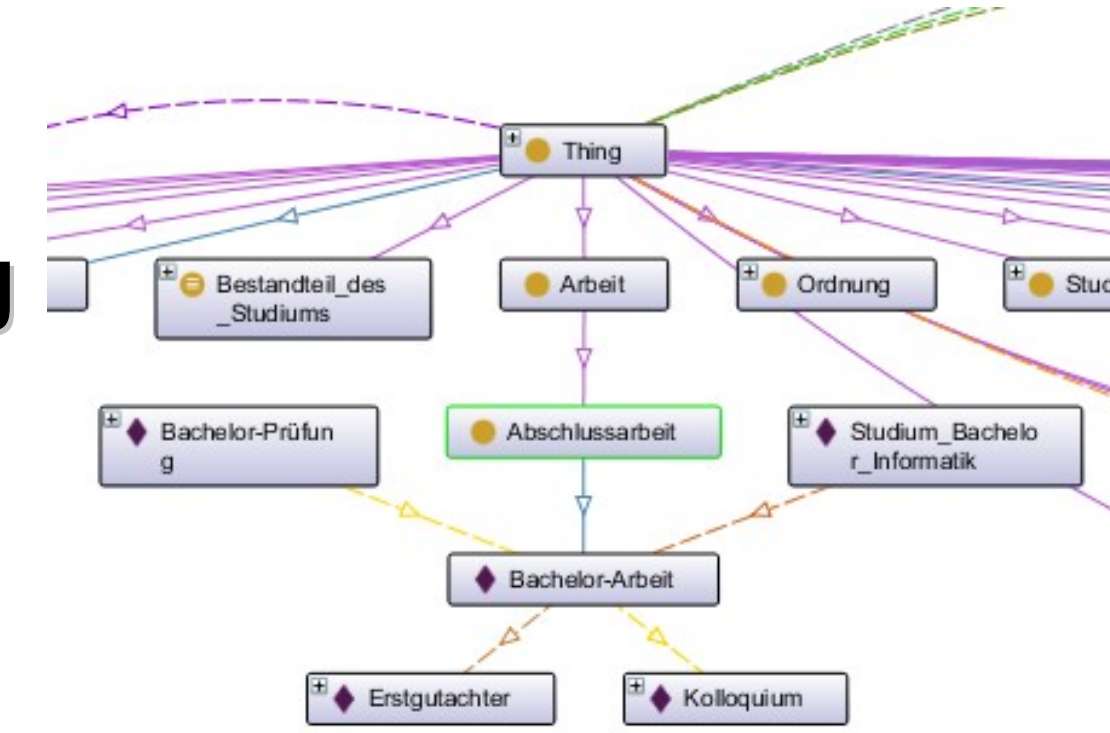


Modellierung der Studien- und Prüfungsordnung des Studiengangs Bachelor Informatik in OWL



Bachelorarbeit, vorgelegt von Marco Knietzsch

Aufgabenstellung:

In dieser Arbeit soll besagte Ordnung so aufbereitet werden, dass sie für das Semantic Web benutzbar wird. Das konkrete Ziel ist eine Ontologie die zu bestimmten Themen des Studiums befragt werden kann. Sie soll alle Fragen beantworten, die auch in der originalen Ordnung beantwortet werden. Zudem sollten Schlüsse gezogen werden können, die weiteres Wissen generieren. Fragen könnten sein: "Was sind Pflichtfächer?", "Muss ich ein Praktikum machen?", "Wie lang ist die Bearbeitungszeit für eine Bachelorarbeit?"

Darüber hinaus soll das Modell natürlich im Sinne des Semantic Webs wiederverwendbar sein.

Semantic Web:

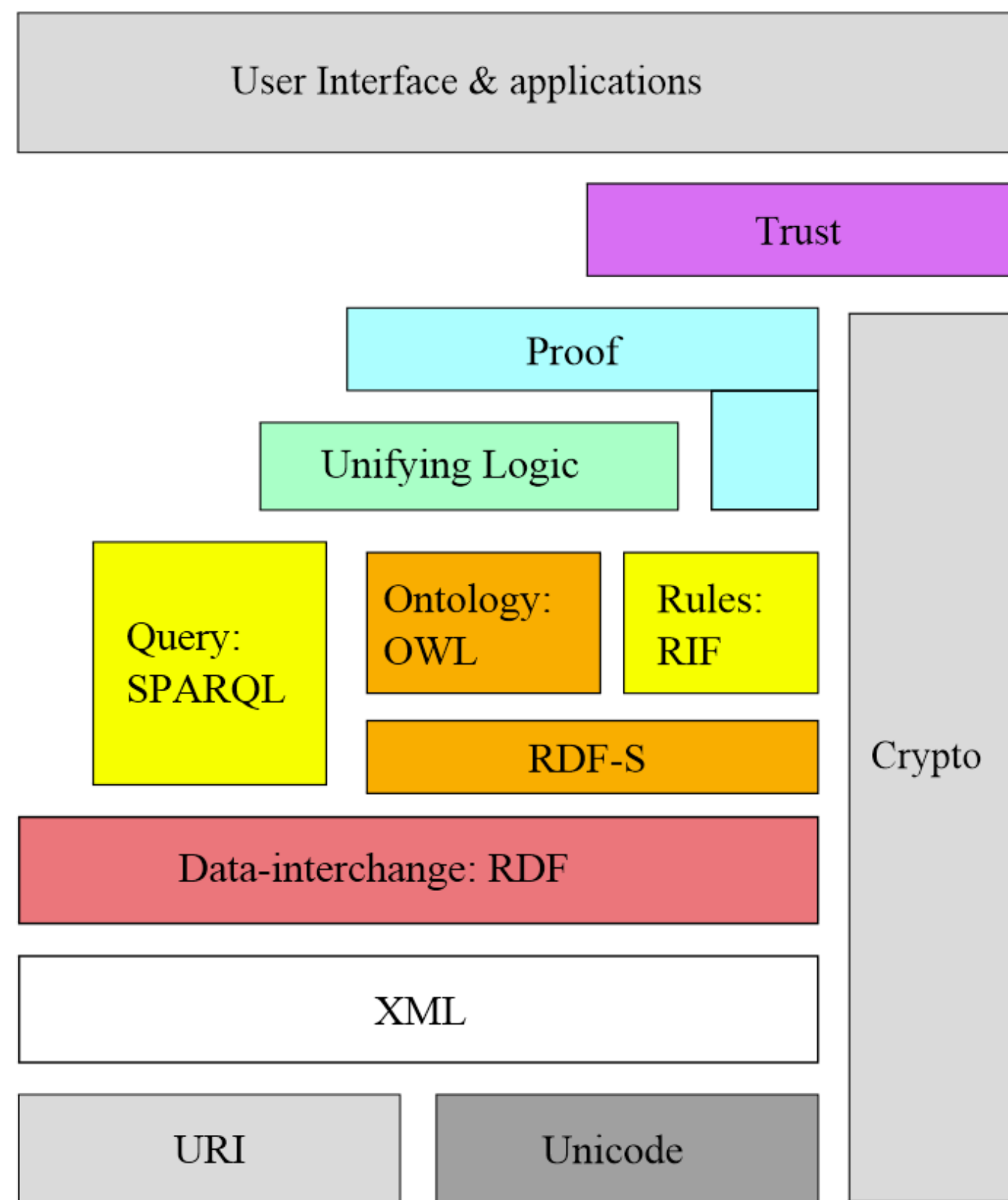
Auch bekannt als "Internet der Dinge", soll es die Weiterentwicklung des WWW sein. Ziel dessen ist es der Maschine die reale Welt "begreifbar" zu machen, so dass sie mit von Menschen zusammengetragenen Informationen umgehen und daraus Schlüsse ziehen kann.

Die Grundlage des Semantic Web ist das Resource Description Framework (RDF). RDF ist einer Datenbank ähnlich. Es unterscheidet sich dadurch, dass anstelle der Tabellen Tripel gespeichert werden. Eine Tabelle Personen hat beispielsweise für jeden Eintrag eine Zeile und für jedes Datum eine Spalte. Diese Datenbank kann nur von denen genutzt werden, die ihren Aufbau kennen. Und es ist nahezu unmöglich sie mit anderen Datenbanken zu verbinden. Zum Beispiel würde es schon zu Konflikten kommen, wenn die Spalten in einer anderen Reihenfolge wären oder die gleichen IDs vergeben wurden.

ID	Vorname	Nachname	Geboren am
0	Max	Mustermann	1.1.1960
1	Erika	Mustermann	12.8.1964

Dahingegen gibt es im RDF immer nur drei Spalten: Subjekt, Prädikat und Objekt. Jede Zelle in der Datenbank wird zu einer Zeile in RDF. Zudem bekommt jedes Subjekt, Prädikat und Objekt eine eindeutige globale ID(IRI). Somit können verschiedene Informationen über dasselbe Subjekt auf verschiedenen Servern gespeichert werden. Diese Informationen können problemlos getrennt und wieder zusammengeführt werden.

Subjekt	Prädikat	Objekt
0	hatVornamen	Max
0	hatNachnamen	Mustermann
0	geborenAm	1.1.1960
1	hatVornamen	Erika
1	hatNachnamen	Mustermann
1	geborenAm	12.8.1964



Semantic Web Layercake von I, Mhermans

OWL:

OWL erweitert RDF um Klassen und Attribute. Es stellt die Basisklasse »Thing« und die Basisproperties »topObjectProperty« und »topDataProperty« zur Verfügung. Klassen stellen Konzepte dar, die eine Reihe von Instanzen haben können. ObjectProperties verbinden Instanzen miteinander. Klassen sind selbst Instanzen von »owl:Class«, deswegen können auch sie mit ObjectProperties verbunden werden. DataProperties sind Prädikate die Instanzen mit bestimmten Werten verbinden, die aber keine Instanzen von Klassen sind.

Grundsätze der Modellierung:

Es gibt derzeit keine Musterlösung, wie eine Ontologie zu modellieren ist. Jedoch ist der Aufbau eines Tripels intuitiv: Subjekt, Prädikat, Objekt - wie in einem natürlichen Satz. Die Annahme, dass Subjekt, Prädikat und Objekt im Satz zu einem Tripel im Modell werden, liegt also nicht fern. Jedoch können Subjekte in OWL sowohl Klassen als auch Instanzen sein. Richtlinien für die Entscheidung zwischen Instanz oder Klasse werden in »Semantic Web for the Working Ontologist« gegeben:

»The first is based on the simple observation that classes can be seen as sets of instances. If something is modeled as a class, then there should at least be a possibility that the class might have instances. If you cannot imagine what instances would be members of a proposed class, then it is a strong indication that it should not be modeled as a class at all. [...] The second guideline has to do with the properties that describe the thing to be modeled. Do you know (or could you know) specific values for those properties or just in general that there is some value?«

Zudem ist es wichtig zu wissen bis zu welchem Punkt modelliert werden soll. Es ist möglich für jede denkbare Klasse noch eine Oberklasse zu finden, die man auch noch in die Ontologie aufnehmen könnte. Dies behindert jedoch jeden, der die Ontologie wiederverwenden will, da es ihm die Freiheit nimmt selbst Klassen zu definieren. Deswegen gilt: Hat eine Klasse keine Instanzen, sollte sie vielleicht gelöscht werden.

Beispiel:

In der Ordnung heißt es: »Diese Ordnung regelt Ziel, Inhalt, Aufbau, Zugangsvoraussetzungen und zeitlichen Ablauf des Studiums in dem Bachelor-Studiengang Informatik am Fachbereich Informatik und Medien der Fachhochschule Brandenburg.«

Das erste Konstrukt ist sehr einfach: »Ordnung regelt Ziel, Inhalt...« Dies kann man als Tripel hinzufügen. Die Objekte sind nun wiederum Subjekte im Bezug auf das Studium. Intuitiv wäre es, wenn Ziel, Inhalt, Aufbau, usw. zu ObjectProperties würden, in der Form Studium hat Aufbau ... Das funktioniert jedoch nicht, da dieses Tripel kein Objekt hat. Der Zusammenhang lässt sich am besten an einem Beispiel erläutern: Wenn man einen Menschen beschreibt listet man seine Eigenschaften auf: »Er ist 20 Jahre alt, 180 cm groß, ...«. Diese entsprechen den Tripeln »Mensch1 hat Körpergröße 180 cm«, »Mensch1 hat Alter 20 Jahre«. Wenn jedoch das Konzept eines Menschen beschrieben werden soll, sind die Eigenschaften selbst die Objekte: »Ein Mensch hat eine Körpergröße und ein Alter«. Hier wird das Konzept eines Studiums beschrieben, deswegen werden Ziel, Inhalt, Aufbau... unter einer Klasse »Eigenschaft« zusammengefasst. Diese werden mit dem Prädikat »ist Eigenschaft von« verknüpft.

Das Studium könnte sowohl Klasse als auch Instanz sein. Um dies zu bestimmen werden die im Lösungsansatz genannten Fragen gestellt: Gibt es Instanzen? Im Text ist die Rede von »dem Studium«, der Singular weist darauf hin, dass es davon nur eins gibt, zumindest in dieser Ordnung. In der nächsten Ordnung wird die Rede von einem anderen Studium sein, das auch nur »Studium« heißt, aber etwas anderes meint. Das heißt es gibt eine Klasse »Studium« und unser »Studium« hier ist eine Instanz dessen. Um Verwechslungen zu vermeiden bekommt es den Namen »Studium Bachelor Informatik«, der ist noch nicht global eindeutig, aber im Namensbereich der FH Brandenburg.

Zusammenfassung:

Da die Ordnung zu einem großen Teil aus Regeln besteht ist deren Modellierung in OWL weniger sinnvoll. Das Modellieren eines Textes nach der intuitiven Methode erwies sich jedoch als brauchbar und führte zu keinen ersichtlichen logischen Fehlern. Das größte Problem ist aber die Beschränkung durch Klassen, da die selben Dinge im Text manchmal als Individuum und als Klasse auftauchen. Dies kann aber mit der Art des Textes zusammenhängen. Probleme gibt es auch mit Begriffen, die mehrere Bedeutungen haben, wie »Studium«. Da Ungenauigkeiten in der natürlichen Sprachen nicht unüblich sind, können diese Doppeldeutigkeiten schnell zur Verwirrung führen. Insgesamt lässt sich also sagen, dass es sinnvoller ist konkrete Dinge in OWL zu modellieren als deren Beschreibung.

Ausblick:

Hierauf aufbauend kann man nun den Studiengang an sich modellieren, der Daten über die Studenten, Lehrende und Module beinhaltet. Der Nutzen eines solchen Modells bestünde in Inferenzen wie »Student A kann Modul X nicht belegen, weil er Modul Y nicht belegte.« Dies würde Studenten die Möglichkeit geben ihr Studium besser zu planen, sowie die Administration des Studienganges vereinfachen. Dieses Modell hingegen erschwert eher die Suche nach Informationen über den Studiengang.