Case Study: Use of Semantic Web Technologies in Natural language interface to Business Applications

C. Anantaram, Tata Consultancy Services Limited

April 2007

General Description

Tata Consultancy Services Limited (TCS) has research labs across the globe that focus on a wide range of domains of interest. One of the areas of research in TCS is to investigate natural language interfaces to business applications. A framework called NATAS has been developed, that enables users to interact with a business application in natural language by posing questions and invoking tasks.

NATAS analyses, interprets and evaluates user input and responds back appropriately. The architecture of NATAS uses a Semantic Web based ontology of the domain to aid in the retrieval of relevant data and concepts from the application. The idea is for the application to carry out a conversation with the user, in order to drill down to what the user actually wants and then to identify the task(s) that would carry out the user’s requirement. The natural language interface interprets the text and calls appropriate APIs of the application to accomplish the requested tasks. The main advantage of such a system is that the user is free to enter any information in a raw form. It is then the job of the system to process the raw information and get whatever else is required.

Benefits of Semantic Web Technology

Our approach in NATAS is based on an explicit domain ontology that is described using Semantic Web technology, namely RDF (Resource Description Framework) in the N3 representation, OWL (Web Ontology Language), and SPARQL. The ontology describes the main concepts of the domain and their inter-relationships with a <subject-predicate-object> structure for each of the concepts.

The ontology is categorized into three levels — Seed Ontology, Application Ontology and Domain Ontology. The Seed Ontology describes the basic relations that are applicable in the domain; for example “customer has a policy number”, “a customer is a person” etc. The Application Ontology is created by instantiating the Seed Ontology with the Application Data through an
Ontology Generator. The Application Database provides the actual data that is present in the business application system and forms the basis of the ontological relations (or concepts) of the domain.

![Architecture for NATAS](image)

We assume that all the instances of the objects in the domain are stored in the database. We call the data in the database the static facts. The ontology is used in conjunction with domain rules to generate derived facts based on the data of the domain. We use CWM (Closed World Machine), a description logic based tool of W3C, to do the rule evaluation and application, and thus the Domain Ontology is created. The static facts and the derived facts are then converted into RDF. This Domain Ontology is used by the NATAS system to answer questions or perform tasks.

The schema of the ontology helps define domain specific relations that are relevant and semantically meaningful for the domain. For example, we have defined schemas called “ds” and “ns” with relations that can be used to state triples such as:

```
ds:who ns:key_concept ds:eName.
```

The semantics of the relation key_concept is specific to a domain and helps the natural language engine to reason based on the semantics provided by the relation.

We use SPARQL to query the RDF data in memory and to fetch relevant data. SPARQL provides the ability and the flexibility to perform generic queries. The general structure of the query is (subject, predicate, object). We have identified seven types of queries for the subject-predicate-object (hence forth referred to as s-p-o) structure of our ontology; these are: s (only subject); p (only predicate); o (only object); s-p (subject and predicate); s-o (subject and object); p-o (predicate and object); s-p-o (subject, predicate and object specified). After the concepts are identified from the input sentence, the concepts are classified as subject, predicate or object. The actual query is formulated by binding the value of the concept raised (and classified as s-p-o) in the input sentence to the generic SPARQL query that is one of the above seven types, in order to formulate the precise query and retrieve the answer. Ontology traversal is another mechanism for obtaining answers for queries posed by the user. When the application data is transformed into the Domain Ontology, a graph structure is created in the memory. It is this graph structure that is traversed in order to obtain an answer for a query. The graph traversal figures out the node, which is directly or indirectly connecting two or more different nodes. Thus questions like “what is common between X and Y?” which are quite cumbersome to be answered by a query or through an API function, are answered easily through ontology traversal.
Key Benefits of Using Semantic Web Technology

In summary, Semantic Web Technology:

- enables distinct semantics for the various concepts in the domain, through definition of multiple schemas
- provides a crisp and simple mechanism to represent an ontology using the <s-p-o> structure of RDF
- provides mechanisms to formulate generic queries (SPARQL) and instantiate them at runtime in order to answer the queries posed.
- provides mechanisms to create parts of the ontology and query on it seamlessly, using various technologies
- provides rule evaluation and execution mechanism to create derived facts
- provides mechanisms to link in external concepts with existing concepts of the domain through simple <s-p-o> structures

The use of the Semantic Web allows the natural language system to interact “intelligently” with the user by traversing the domain ontology for both the static and derived facts. With the widespread nature of today’s domain knowledge, Semantic Web technology allows seamless integration of different resource definitions that semantically mean the same thing. This permits easier integration of domain knowledge, which in turn makes the natural language system more robust in answering queries posed by the user.

© Copyright 2007, Tata Consultancy Services Limited