Application of Semantic Search in Idea Management Systems

Geovanny Poveda, Adam Westerski
Grupo de Sistemas Inteligentes
Universidad Politécnica de Madrid
http://www.gsi.dit.upm.es
Outline

- Background Gi2mo Project.

- The Problem
  - Users can not find specific information
  - Non-Contextualized queries

- Our Solution
  - Semantic Approach – *Semantic Search*
  - Semantic Module

- Evaluation
  - Ubuntu brainstorm - 5k ideas

- Conclusions
• Gi2mo Project

✓ Is a project which goal is to setup Semantic Web technologies in the environment of Idea Management Systems.

✓ GI2MO aims to improve current Idea Management Systems by providing data integration capabilities and additional data analysis tools though rich metadata descriptions. *Semantic annotations.*
Gi2mo Project – Technical Point of View

- GI2MO offers a test bed platform for experimentation of projects related to ideas management concept.

- Is an open-source Idea Management System based on the component architecture of a Content Management System called Drupal.

- GI2MO offers a test bed platform for experimentation of projects related to ideas management concept.
• **What is Idea Management System?**

✓ Is the process and associated discipline of facilitating ideation from organizational and management perspective.

✓ Idea Management is a promising industry sector which produces software for collecting and organizing input from people regarding proposals for innovation of products and services.
• What is Idea Management System?
Gi2mo Project – Our research so far

- Gi2mo Project

  ✓ Define Gi2MO ontology for formalizing and interlinking idea management systems (IMS).

  ✓ Connect ideas with Enterprise systems following the Linked Data model. (Rdfme Module)

  ✓ Idea Classification, filtering and analysis of idea similarities independent of idea topic.

  ✓ Integrating Opinion Mining techniques with IMS.
Ideas Management Life Cycle

• Innovation through Idea Management Systems
Research Motivation

On traditional search model, users get unnecessary information.

- Currently search models – keyboard approach
- Data analysis requires external tools
- Data interoperability – Close Data
- there is a centralized model
Gi2mo Project – Our research Goal

- Gi2mo Project (Semantic Search)

Improve semantic search model on Ideas Management System. *(This Talk!)*. **Auto-Assisted Search**
Our Purpose – A typical use case

- Semantic Search approach

Ideas that have "great" word and whose author has a number of posts greater than 10 and whose solutions have a number of reviews greater than 15 and contains some of these words: use | change | add | remove and whose comments were made by user with a rating greater than 10.

Demands for Smarter Search

- Find information faster.
- Find more specific information.
- Find Linked information.
- Enrich the search index.
Gi2se – Architecture - Picture
Gis2e Features

- **Gi2se**
  - Features.

Is a software component based on ARC2, RDF and SPARQL libraries in which users can do:

- Dynamic source updates.
- Users can execute multiple queries using multiples sources.
- Configure remote SPARQL endpoints.
- Users can configure RDF files and graph names.
Evaluation Methodology

1. Collect Ideas from Brainstorm Ubuntu
2. Import ideas RDFM E Utility
3a. Semantic Mapping Utility
3b. Object Serialization Query
3c. Building Query template
3d. Output result serialization

• Extraction -- > Semantic Scrapper

  ❖ Generalities.

  ✓ Scanners semantic based extractors using RDF to select fragments Web documents and data and build RDF graphs from unstructured information.

  ✓ Assuming that the content of a similar type are present on the web with similar visual features (CSS Style).
Mapping, export and Ideas expose

- Generalities.

The data that has been exported in RDF following the defined mappings can be also imported back into the system (this means that a number of distributed Idea Management Instances can be connected together). This functionality is enabled both via UI and as a REST service.
Mapping, export and Ideas expose

- Generalities.

✓ The data that has been mapped on Gi2MO platform is exposed via REST service, thus users can get all ideas in a single RDF resource accessible via a HTTP URL.
Third Step - A

• Semantic mapping utility
  
  ❖ Features.

  An algorithm map the information into DB (Triple store in conjunction relational data base):

  ✓ Relational database are mature and widely used.

  ✓ Triple store provides more flexibility.

  ✓ Triple store produces better performance.

  *triple store enables smart integration with relational database*
Third Step - A

- Semantic mapping utility
  - Features.

1. Each triple found is identified by a unique number.

2. Id by each group of properties found in name spaces (title, content, creator, etc)

3. Id number by each Object found = content of information

4. Each triple found is identified by a unique number.
Object Serialization Query

- How do you involve both: SQL and RDF?

Using scripts based on method chaining (integrate data in a single operation)

- PHP sentences use SPARQL parameters to instantiate SQL operations.
- We extended ARC2 functionalities.
- That is, we have defined intermediate entities between SPAQRL sentences and SQL operations.
Third Step – C - I

- **Semantic Search query template**
  - Features.

  Refers to the way in which queries are built

  ✓  An algorithm check the classes, attributes and attributes values to build automatic SPARQL sentences.

  ✓  Search entities following structure: class-> property -> value of property.

  ✓  Forces the users to create triples of information as input information.
• **Semantic Search query template**

- **Features.**
  
  ✓ We have involved prologue and form section.

  ✓ Prologue section instantiates name spaces to be used as prefix. (according properties chosen by users).

  ✓ In form section a method checks RDF properties selected by users and transform them as SPARQL select parameters.

  ✓ Automatic generation of UNION operators by each created triple.

  WHERE {Graph Pattern (1) U Graph Pattern (2) U .... Graph Pattern (n) }.
• Semantic Search query template

❖ Example.

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX gi2mo: <http://purl.org/gi2mo/ns>
SELECT ?name ?symbol ?weight ?number
FROM <http://gi2mo.org/ideastream/rdfme/ws/1/ubuntuBrainstorm2012-12-12_01.rdf>
WHERE
{
  ?gi2mo gi2mo:comment "good".
  ?gi2mo dcterms:description ?ideaDescription.
  FILTER ?post > ? postValue
}
Goal: Determine if crawler and mapping process are useful for semantic search model.

We use a crawler process to get ideas and additional data which contains comments, reviews, idea statuses and other related metadata.
Results Experiment I

Crawl 5122 Ideas in 2 hours Import: 14890 triples (1500 ideas) in Gi2mo Platform in 1 hour and 45 minutes.

The time spent depends of number of connections associated by each property. In our case number of comments and solutions.
Experiment II

Goal:

Determine time spent when we try to find the following search: “Ideas that has WIFI word and whose proposed solutions contain some of these words: use | change | add | remove and whose comments were made by user with a rating greater than 10”

Using GI2se module!!.
Results Experiment II

On 1500 triples, our algorithm spent 356 (s) to find the aforementioned string. (2000 triples -> 472 s)

Time spent during the query does not depend on quantity of information that algorithm needs to process. It depends on the graph model querying implemented in the algorithm (node search).
Gi2se Features – Screenshot

- **Gi2se**
  - Features.

**Gi2se settings**

- No existia registro
- The configuration options have been saved.

**Gitonk endpoint settings**

**Configuration RDF URL and Graph Name**

You can configure the options related RDF file source, URL Endpoint and Graph Name. The SPARQL Endpoint by default is !sparql, but you can find more at !list. Some SPARQL Endpoints use local data, so to allow a custom ontology you must perform a query at the endpoint to store the data in cache.

RDF File URL:

http://localhost/gi2seFinal/modules/gi2se/test.php

Graph Name:

**Url Rdfs and Name Graphs Configured**

<table>
<thead>
<tr>
<th>URL End Point</th>
<th>Graph Name</th>
<th>Actions - Delete</th>
<th>Actions - Update</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://localhost/rdf/brainstormV2.rdf">http://localhost/rdf/brainstormV2.rdf</a></td>
<td><a href="http://gi2mo.org">http://gi2mo.org</a></td>
<td>Delete</td>
<td>Update</td>
</tr>
</tbody>
</table>

dit.upm.es
• SOC

Nuevos Conceptos.
- Alta Cohesión.
- Recursos – Representaciones
- Debilidades.
- Personalización.
- Portabilidad.
- Interoperabilidad.

Inteligencia Colectiva

MAS!

Geo 2011 UPM

dit.upm.es
Gis2e Features – Screenshot III

- Class:
  - gi2mo:#Use
    - gi2mo:#User: An instance of this class models a user account.
      - User types can be defined by creating UserGroups.
  - gi2mo:#UserGroup: An instance of this class allows to group...
Result1- Update manager tries to update even without internet connection

10 Feb 11 at 19:04

When you have no Internet connection, update manager or synaptic will still try and connect, giving you a list of warnings for each package. This can be confusing for many users. In fact, this could happen if you use your laptop at work and forget to switch off the proxy settings. So there could be several causes. A bunch of errors telling you that something could not be downloaded is unhelpful and unfriendly.

Result2- Ubuntu Software Centre add more COMMERCIAL games at paid apps

19 Feb 11 at 15:31

Installing software on Ubuntu from the repository's works great, far better, easier, and more user friendly then Windows. But for commercial Linux games... There are commercial games (ok.. not the newest or many, but still..) that you can install on Linux, but not really user friendly way. Even installing these in Windows is much easier all trough the technology (if it would be used) under Ubuntu's Software Centre, make it more easy then installing/updating Windows programs. Let´s say, Ubuntu asks some game company's (see further idea to begin) so that Ubuntu can create a commercial game PPA to sell games. Users click their desired game, pay for it, get get the cd-key per e-mail and enter the key within the installation in the software centre. Off course, it will not be a big market in the beginning, but we have to start somewhere!

Result3- Ubuntu Software Centre add more COMMERCIAL games at paid apps

19 Feb 11 at 15:31

Installing software on Ubuntu from the repository's works great, far better, easier, and more user friendly then Windows. But for commercial Linux games... There are commercial games (ok.. not the newest or many, but still..) that you can install on Linux, but not really user friendly way. Even installing these in Windows is much easier all trough the technology (if it would be used) under Ubuntu's Software Centre, make it more easy then installing/updating Windows programs. Let
Gi2mo Software

- Gi2mo.org

Software released as open source, everything available at http://www.gi2mo.org
Conclusions

✓ we presented only an approximation to real semantic business intelligence scenario, because the data contained in an IMS needs to be interpreted in a best way before executing the semantic search processes

✓ we envision to implement a module in which the obtained information during the search process, can be represented using clustering techniques
Questions?

gpoveda@gmail.com

Thanks!!

London 2012