

# 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>

- ❖ 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



- ✓ 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

# Idea Management System - Definition

- What is Idea Management System?

The screenshot shows the IdeaStorm website interface. At the top, there is a navigation bar with the IdeaStorm logo and the tagline "Where Your Ideas Reign". A notification box on the right states "There are currently no Storm Sessions active. Stay tuned!". Below the navigation bar, there are four main action buttons: "View" (All posted ideas by the community), "Post" (Your idea for Dell products or services), "Vote" (Promote or demote ideas), and "See" (Your Ideas in action). A "Sort Ideas By" section includes options for "Popular Ideas", "Recent Ideas", and "Top Ideas", along with a dropdown menu set to "All".

The main content area displays a list of ideas under the "Comments" section. The first idea is titled "Pre-installed OpenOffice | alternative to MS Works & MS Office" by user "chart" on Feb 17, 2007. It has 118030 votes and includes a "Promote" button (up arrow) and a "Demote" button (down arrow). The description states: "Provide OpenOffice.org for free pre-installation alongside Microsoft Works and Microsoft Office. OpenOffice.org is more capable than Microsoft Works, and a serious competitor to Microsoft Office, at a fraction of the cost (it's free!) OpenOffice.org can open, create, edit and save Microsoft Word, Excel and PowerPoint files."

The second idea is titled "Pre-installed Linux | Ubuntu | Fedora | OpenSUSE | Multi-Boot" by user "chart" on Feb 16, 2007. It has 108450 votes and includes "Promote" and "Demote" buttons. The description states: "Offer the 3 top free Linux versions for free pre-installation on all Dell PCs."

On the right side, there is a "Login to IdeaStorm" section with a "Register Now" link, input fields for "Username:" and "Password:", a "Login" button, and a "Forgot Password?" link. Below this is a "Your Ideas in Action" section featuring a "New Idea" button and a link to "IdeaStorm Recap - 1/22/2010". The recap text reads: "Happy 2010 everyone! I know I'm a little late with the holiday greetings, but there is a lot to share on..."



- 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.



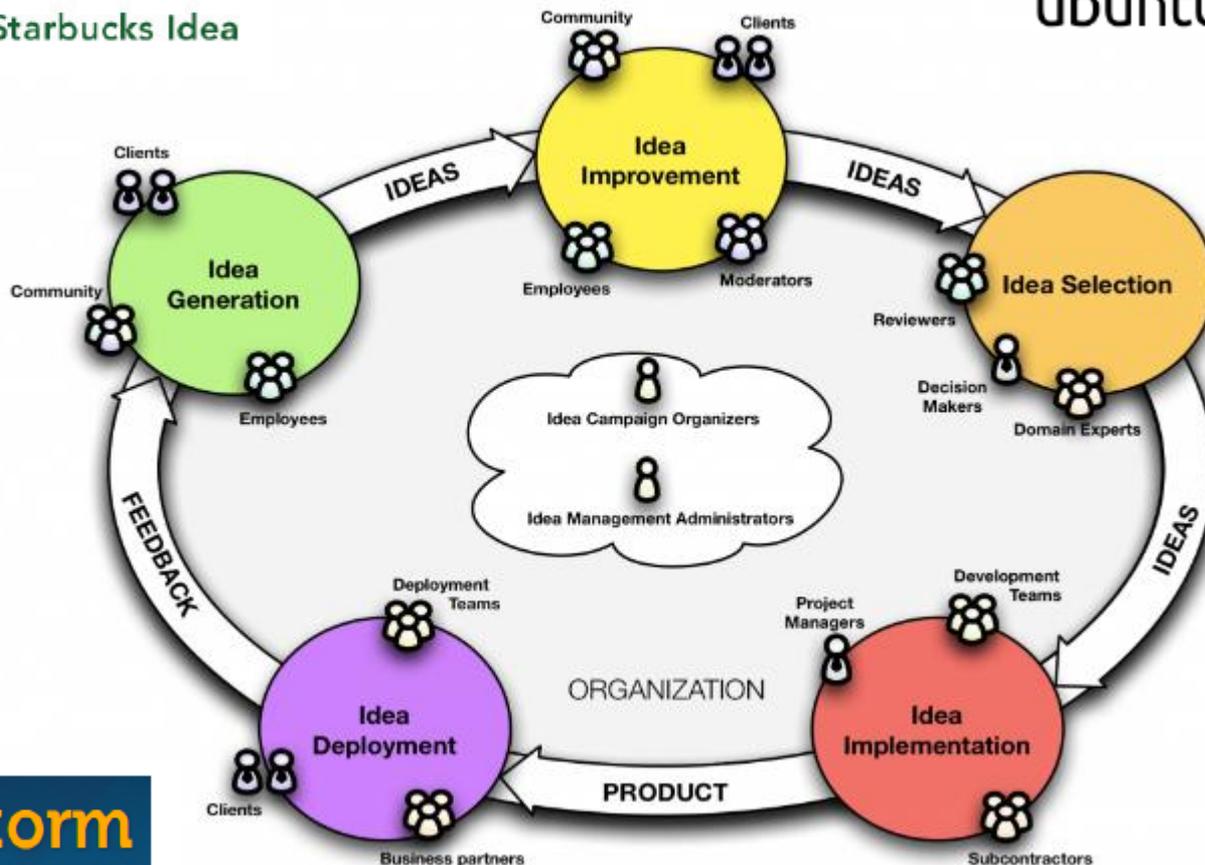
# Ideas Management Life Cycle

- Innovation through Idea Management Systems



My Starbucks Idea

ubuntu brainstorm 



IdeaStorm



- 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)

POWERED BY  
**GI2MO**



Improve semantic search model on Ideas Management System. **(This Talk!). Auto- Assisted Search**

## Our Purpose – A typical use case

### Semantic Search

on its way!



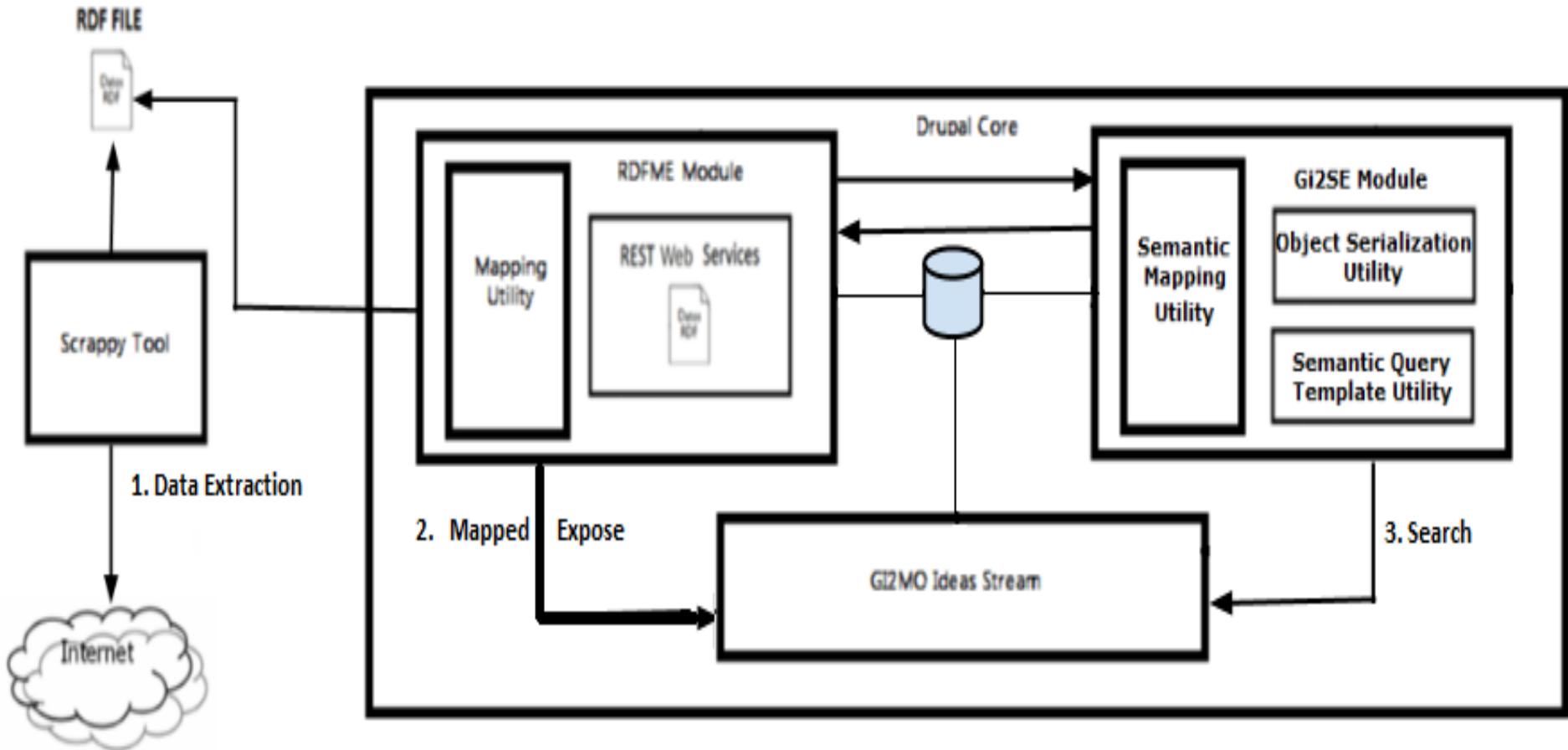
- Semantic Search approach

*Ideas that have "great" word and whose author has a number of post 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

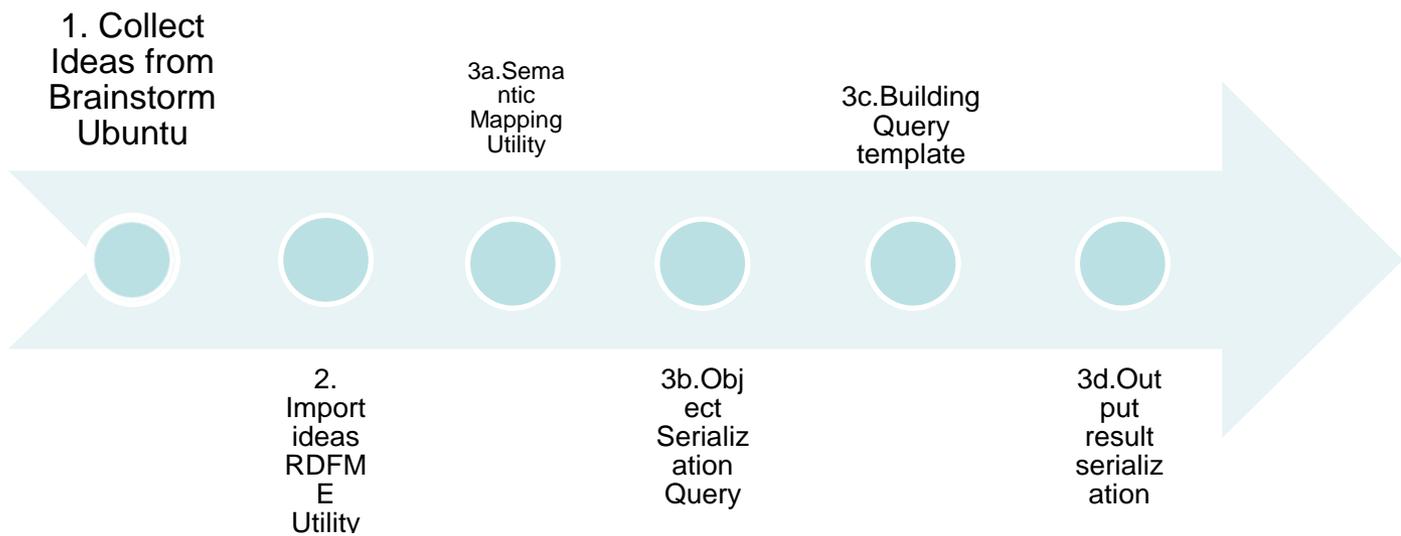
- ❖ Features.



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



Number of Ideas	Number of Triples	Crawler time (s)	Mapping time (s)
100	1795	614,67	892,16
200	13760	5582,12	4578,36
500	7843	4189,40	4093,61



Semantic scrapper, available at <http://www.gsi.dit.upm.es/index.php/en/software/details/1/2/software-scrappy.html>



## First Step

- Extraction -- > Semantic Scrapper
  - ❖ Generalities.

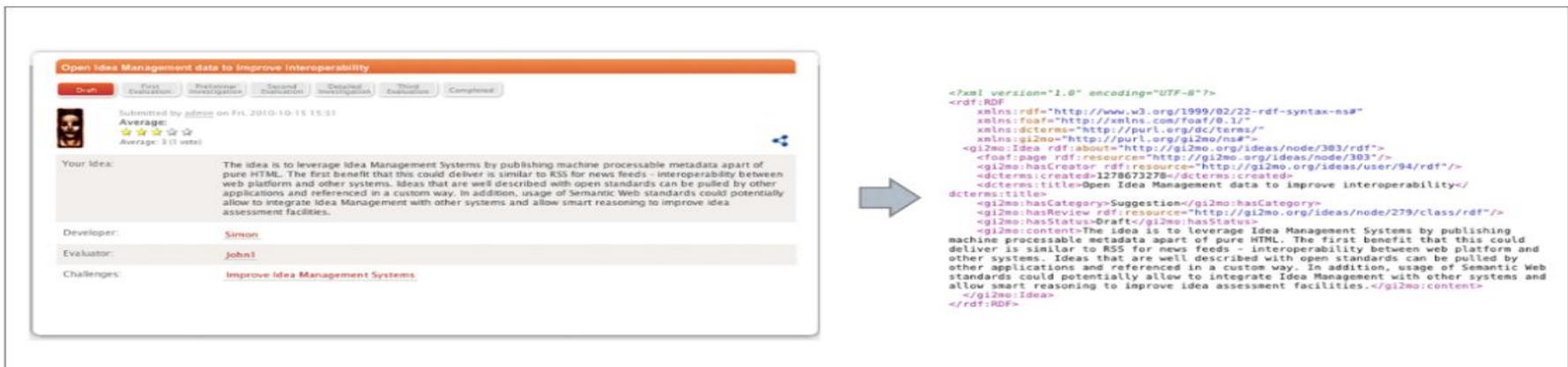


- ✓ Scrappers semantic based extractors using RDF to select fragments Web documents and data and build RDF graphs from unstructured information.
- ✓ . Assuming that the content of a similar type are present on the web with similar visual features (CSS Style).

## Second Step

- Mapping, export and Ideas expose
  - ❖ Generalities.

# (RDFme)



The image shows a screenshot of a web application interface for idea management. The interface has a header "Open Idea Management data to improve interoperability" and a navigation bar with buttons: "Draft", "First Evaluation", "Peer-review", "Second Evaluation", "Detailed", "Third Evaluation", and "Completed". Below the navigation bar, there is a profile picture and a submission timestamp: "Submitted by admin on Fri, 2010-10-15 15:31". There is an "Average" rating section with five stars and the text "Average: 3 (1 vote)". The main content area is titled "Your idea:" and contains the text: "The idea is to leverage Idea Management Systems by publishing machine processable metadata apart of pure HTML. The first benefit that this could deliver is similar to RSS for news feeds - interoperability between web platforms and other systems. Ideas that are well described with open standards can be pulled by other applications and referenced in a custom way. In addition, usage of Semantic Web standards could potentially allow to integrate Idea Management with other systems and allow smart reasoning to improve idea assessment facilities." Below this text, there are fields for "Developer:" (Simon), "Evaluator:" (John), and "Challenges:" (Improve Idea Management Systems). An arrow points from the "Your idea:" section to a block of RDF/XML code on the right. The code is an XML document with the following structure: <?xml version="1.0" encoding="UTF-8"?><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:gi2mo="http://purl.org/gi2mo/ns#"><gi2mo:Idea rdf:about="http://gi2mo.org/ideas/node/303/rdf"><foaf:page rdf:resource="http://gi2mo.org/ideas/node/303"/><gi2mo:hasCreator rdf:resource="http://gi2mo.org/ideas/user/94/rdf"/><dcterms:created=1278673278/><dcterms:title=Open Idea Management data to improve interoperability/><dcterms:hasCategory=Suggestion/><gi2mo:hasReview rdf:resource="http://gi2mo.org/ideas/node/279/class/rdf"/><gi2mo:hasStatus=Draft/><gi2mo:hasStatus=><gi2mo:content=The idea is to leverage Idea Management Systems by publishing machine processable metadata apart of pure HTML. The first benefit that this could deliver is similar to RSS for news feeds - interoperability between web platform and other systems. Ideas that are well described with open standards can be pulled by other applications and referenced in a custom way. In addition, usage of Semantic Web standards could potentially allow to integrate Idea Management with other systems and allow smart reasoning to improve idea assessment facilities.</gi2mo:content=></gi2mo:Idea=></rdf:RDF=>

Example of metadata annotations export for a single idea.

close or Esc Key

- ✓ 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.

## Second Step

- Mapping, export and Ideas expose
  - ❖ Generalities.

(RDFme)

Idea - (idea)

Class: gi2mo:Idea		
Properties Blacklist: language, changed, body, name, comment, last_comment_timestamp, teaser, promote, moderate		
ID	RDF property: foaf:page nid	Pattern: http://pgi2.atosorigin.es/node/i @@@value@@@ will be changed for the value of this field
Creator ID	RDF property: gi2mo:hasCreator uid	Pattern: http://pgi2.atosorigin.es/user/@ @@@value@@@ will be changed for the value of this field
Created	RDF property: dcterms:created created	Pattern: @@@value@@@ will be changed for the value of this field
Comment	RDF property: gi2mo:hasComment map_comment	Pattern: http://pgi2.atosorigin.es/node/i @@@value@@@ will be changed for the value of this field
Title	RDF property: dcterms:title title	Pattern: @@@value@@@ will be changed for the value of this field

- ✓ 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

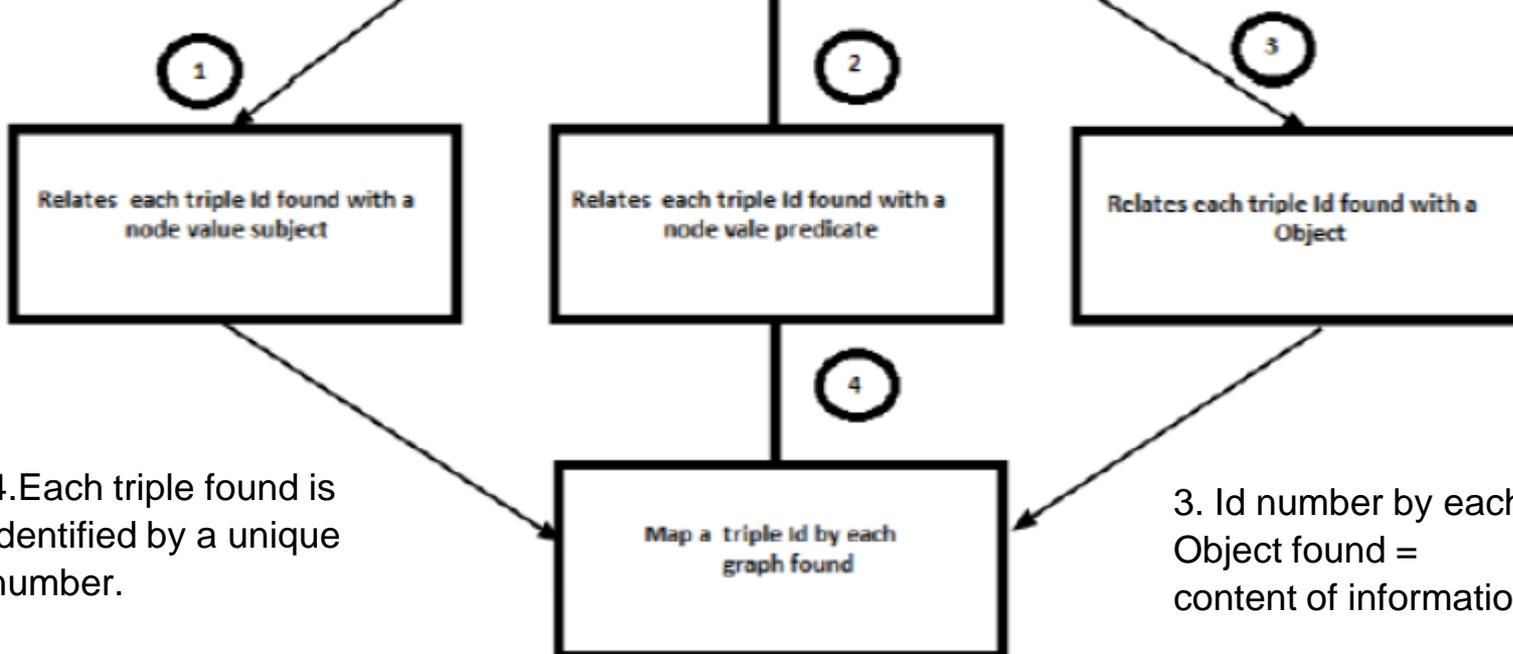
  - ❖ Features.

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

G I 2 S E

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

```
<gi2mo:Comment rdf:about="http://brainstorm.ubuntu.com/idea/21/comment/1">  
<dcterms:created>28 Feb 08 14:49</dcterms:created>  
<gi2mo:content>This will most likely happen with the move to grub2.</gi2mo:content>  
<gi2mo:hasCreator rdf:resource="http://brainstorm.ubuntu.com/contributor/Xhaker/tord"/>  
<foaf:page rdf:resource="http://brainstorm.ubuntu.com/idea/21"/>  
</gi2mo:Comment>
```



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

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

## Third Step - B

- 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 SPARQL 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.

## Third Step – C - II

- 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) }.

## Third Step – C - III

- 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
}
```

- Experiment I

G I 2 S E

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

GI2SE

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

GI2SE

## 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

G I 2 S E

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).



# Gis2e Features – Screenshot I

- 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, in order to import a custom ontology you must perform a query at the endpoint to store the data in cache.

**RDF File URL:**

**Graph Name:**

#### Url Rdfs and Name Graphs Configured

Comments for 1.

Url End Point	Graph Name	Actions - Delete	Actions - Update
http://localhost/rdf/brainstormV2.rdf	http://gi2mo.org	Delete	Update

# Gis2e Features – Screenshot II

## Semantic Search

Add new class

### Output Option:

- Exhibit
- Drupal Native

Choose the option for the Output

Search

### Class:

### Choose the Property:

### Value for the property:

Add new property



## 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

Add new property

# Gis2e Features – Screenshot IV

Resultados Generados de la Búsqueda.

## 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](http://www.gi2mo.org)

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

**GI2MO**  
SEMANTICALLY EMPOWERED IDEA MANAGEMENT

HOME ONTOLOGY TAXONOMY MODEL APPS USE CASES DOCS BLOG ABOUT

**HIGHLIGHTS**

**IDEA BROWSER**

description about primary topic secondary topic ontology diameter fill type

Visualize distributed idea metadata on diagrams

**IDEA BROWSER**

**INNOVATION METRICS**

**OPEN-SOURCE IMS PLATFORM**

**GI2MO in Motion**

**Dashboard Teaser**

Idea Management Systems

**Semantic Web & Linked Data**

**GI2MO Concept in Work**

**GI2MO** is a project which goal is to setup **Semantic Web** technologies in the environment of Idea Management Systems.

**What is Idea Management ?**  
The term of Idea Management relates to systems dedicated to organisation and assessment of large amounts of input from various parties in form of ideas - most often textual content describing innovation related to certain products. For more details see **IMS architecture**.

**What GI2MO delivers ?**  
The primary goal of the project is to construct an **ontology** that will model Idea Management Systems. However, we also devote to study in detail the **appliance** of the data schema onto existing systems and **exploiting it** to achieve real, measurable benefits.

**How can GI2MO help ?**  
GI2MO aims to improve current Idea Management Systems by providing robust data integration capabilities and additional data analysis tools through rich metadata descriptions. Both of those features are derived from the use of Semantic Web technologies to interconnect data inside the Idea Management Systems with assets published by other systems and across the World Wide Web. For details on how it is possible to exploit the benefits of Semantic Web in Idea Management see use cases.



- 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