MONDRAGON UNIBERTSITATEA

GOI ESKOLA POLITEKNIKOA
FACULTY OF ENGINEERING

# SEMANTIC WEB AND SEMANTIC TECHNOLOGIES TO ENHANCE INNOVATION AND TECHNOLOGY WATCH PROCESSES

*Arrasate-Mondragon, 14 October 2016*


*Alain Perez Riaño*

Supervisor: Dr. Felix Larrinaga Barrenechea
Co-Supervisor: Dr. Juan Ignacion Igartua Lopez

Electronics and Computing Department
Faculty of Engineering – Mondragon University

# Table of contents

# INTRODUCTION

# Contribution area and scope

## 1. Innovation process

- Idea life-cycle, from its **definition** to its **implementation**.
- Idea management Systems (IMSs).

## 2. Technology Watch (TW) process

- Capturing **information relevant to a company's business** from outside and inside the organization.
- Information refers to **competitors, technologies, news, patents**...
- **Software platforms** can help storing and managing that information.

# Problem and Motivation

## 1. Data-overflow

- **Noisy** data.
- **Duplicate** or already processes content.
- Hard to manage large amount of information.

## 2. Successful idea contest identification and replication.

- Hard to **identify why** some contest are successful and other not.
- Most IMSs are **idea centered** (do not gather background information).

# Problem and Motivation

## 3. Interoperability

- Lack of **interoperability** among platforms.
- Data re-usability.

## 4. Non productive tasks

- **Waste of time** of experts in non-productive tasks (reading and classifying).

# Objectives

1. Propose a conceptual model for the **identification** of **successful idea contests** and their **replication**.
2. Identify semantic web methods that provide **additional content** to IMSs from repositories inside and outside the company.
3. Propose a concept model to enable the **interoperability among platforms** linking content.
4. **Reduce the workload of the experts** in the TW process.

# Research Questions

1. Can a conceptual model help on **replicate successful Idea Contests**?
2. Can semantic methods enable **content linking among IMS platforms and semantic repositories** internal and external to the companies?
3. Can **interoperability among content** from Innovation and TW platforms be generalized in a single model?
4. Which are the best semantic technologies that help **reducing the time spent on non-productive tasks** inside the TW process?

# Solution Architecture

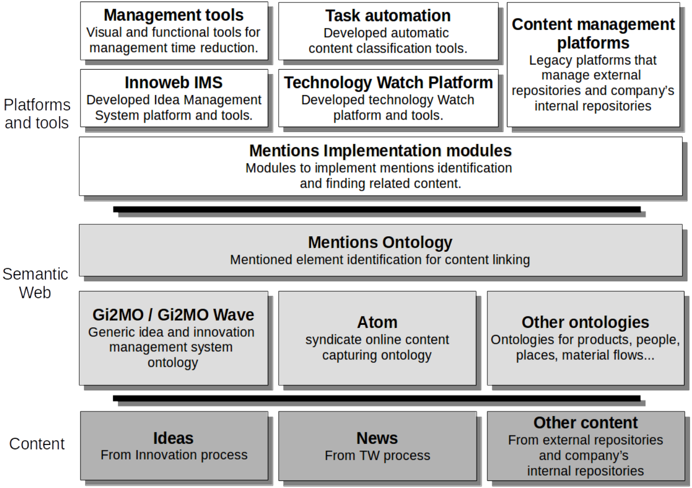



Figure 1: Solution architecture for the thesis.

# FOUNDATIONS

# Innovation Process

- Most of the Innovation process definitions agree on a common baseline[1]:


**Figure 2:** Innovation process baseline stages.

- Thus, management of the ideas is the most critical part.

*[1] N. Errasti, N. Zabaleta, and A. Oyarbide. A review and conceptualisation of innovation models from the past three decades. International Journal of Technology Management, 55(3):190-200, 2011.*

# Innovation Process

- Igartua, J.I.[1] showed relationship between the use of Innovation Management Tools and innovation activity.
- Westerski et al.[2] wrote about innovation, interoperability and linking.
  - *"knowledge management based on interlinking of enterprise systems and web assets to increase information awareness and help innovation assessment".*
- Interoperability issues.
- Relation among Innovation and Technology Watch (TW) processes.

*[1] J. I. Igartua, J. A. Garrigós, and J. L. Hervas-Oliver. How innovation management techniques support an open innovation strategy. Research-Technology Management, 53(3):41-52, 2010.*

*[2] A. Westerski, C. A. Iglesias, and F. T. Rico. A model for integration and interlinking of idea management systems. In 4th Metadata and Semantics Research Conference (MTSR 2010), volume 108 CCIS, pages 183-194, October 2010. Springer Verlag.*

# Technology Watch Process

## TW definition (AENOR's UNE 166006:2011 Standard)

- An organized, selective and permanent process.
- To capture information about science and technology.
- From outside and inside the organization.

## Different domains

- Patents, utility models, industrial designs.
- Legislation, regulations, grants.
- Socio-economic situation of markets.
- News (sector, scientific, doctoral thesis).
- Products.
- ...

# Technology Watch Process

- Society for Competitive Intelligence Professionals (SCIP), identifies five steps in the process:

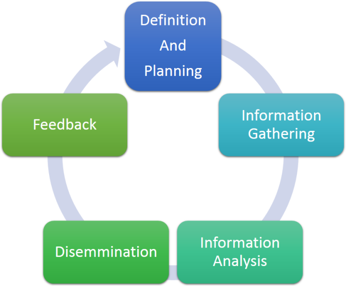



**Figure 3:** Five stages of the Technology Watch process.

- Content gathering may cause a data overflow.

# Technologies to support Innovation and Technology Watch

## Social Web

- Dynamic bi-directional web pages.
- Hence the concept of Collective Intelligence or Wisdom of crowds[1] (Social Networks).
- Increasing interest in IMS[2].

*[1] J. Surowiecki. The wisdom of crowds . Random House Digital, Inc.,3:22, 2005.*

*[2] A. Conry-Murray. Can enterprise social networking pay off. Informationweek.com, 1224:23-29, 2009.*

# Technologies to support Innovation and Technology Watch

## Natural Language Processing (NLP)

- ”explores how computers can be used to understand and manipulate natural language text or speech”[1].
- Covertino et al. [2] used NLP to identify the ”core” of the ideas.
- Other aim to extract text and map it to an ontology or semantic element [3][4]

*[1] G. G. Chowdhury. Natural language processing. Annual review of information science and technology, 37(1):51-89, 2003.*

*[2] G. Convertino, Á. Sándor, and M. Baez. Idea spotter and comment interpreter: Sensemaking tools for idea management systems. In ACM Communities and Technologies Workshop: Large-Scale Idea Management and Deliberation Systems Workshop, 2013.*

*[3] J. West and S. Gallagher. Challenges of open innovation: the paradox of firm investment in open-source software. R&D Management, 36(3):319-331, 2006.*

*[4] S. Hellmann, J. Lehmann, S. Auer, and M. Brümmer. Integrating nlp using linked data. In 12th International Semantic Web Conference, 21-25 October 2013, Sydney, Australia, 2013.*

# Technologies to support Innovation and Technology Watch

## Artificial Intelligence (AI)

- *"is the study of the design of intelligent agents"*[1]
- Classification => learn from training set and classify new items.
- Clustering => it finds similarities among the objects and groups them in different clusters.

*[1] D. Poole, A. Mackworth, and R. Goebel. Computational Intelligence. Oxford University Press Oxford, 1998.*

# Technologies to support Innovation and Technology Watch

## Semantic Web

- *"The Semantic Web is ... an extension of the current one, ... enabling computers and people to work in cooperation"*[1].
- Main technologies:
  - Resource Description Framework (RDF)
  - Web Ontology Language (OWL)
- Their combination make **explicit descriptions** of content on the web.
- The **content of the web is structured and linked** as in any database, creating the Linked Data (LD)

*[1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, 284(5):28-37, 2001.*

## Conclusions

- Innovation is essential for buisiness survival.
- TW closely related to Innovation.
- With the use of ICT:
  - Collaborative **IMSs** => for idea generation and reaching more collaborators.
  - **Semantic Web** & **Linked Data** => for **interoperability** and adding automatically **relevant data**.
  - **Semantic Technologies** => **task automation** for reducing time spent on non productive tasks.

# WEB BASED PLATFORM FOR INNOVATION PROCESSES

## Main Contributions

- **IMS platform** to enable the identification and replication of successful idea contests.
- **Gi2Mo Wave ontology** to annotate semantically the background of idea contests.

# Introduction

## Probematic

- Most IMSs gather **little context information**.
- Hard to identify **why** an **Idea contest** has been **successful**.
- Interoperability for data re-usability.

## Approach

- Build an interoperable IMS.
- Gather and store context information.

# Existing theories

## Architecture participation analysis[1]

- IMSs → **Idea centered**.
- **Open Source** platform selection reinforced.
- **Social networks** help reaching new collaborators.
- **Drupal** & Liferay → Highest score.

## Semantic Web Ontologies for ideas

- Innovation management[2] & Gi2MO ontology [3].

*[1] N. Errasti. Social software in support of collaborative innovation processes. Projectics/Projectique, (2):81-104, 2010.*

*[2] C. Riedl, N. May, J. Finzen, S. Stathel, V. Kaufman, and H. Krcmar. An idea ontology for innovation management. International Journal on Semantic Web and Information Systems, 5(4):1-18, 10 2009.*

*[3] A. Westerski. Gi2mo: Interoperability, linking and filtering in idea management systems. In Extended Semantic Web Conference 2011. PhD Symposium Poster., Heraklion, Greece, May 2011.*

# Research approach

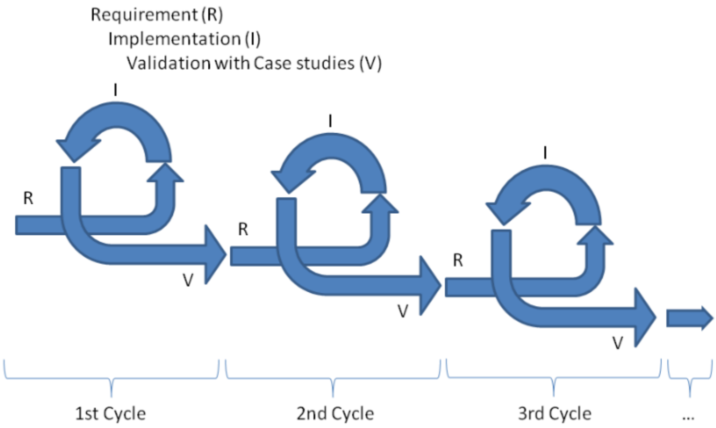



Figure 4: Incremental Development Cycle methodology.

# Research approach

## Platform

- Idea Generation => Blog.
- Idea Analysis => Selection table (with data).
- Idea Enrichment => Wiki.
- Idea Selection => Selection Matrix.

- Wave for multiple idea contests.
- Real Time Web (Twitter-email).
- Events.
- Semantic Web (Gi2Mo Wave).
- Visualization tools.

- Community modules (RDFme, workflow, voting..)

# Research approach

## Ekiten case-study

- Driven by the Engineering, Business and Humanities faculties of MU.
- To promote entrepreneurship among students.
- Each year => duration: 4 months.
- 3 waves each year

- Innoweb => since 2010
- Wave module => since 2011

# Results

| Context parameter | 2010-11 | 2011-12 | 2012-13 | 2013-14 | 2014-15 |
|---|---|---|---|---|---|
| *Groups* | 10 | 27 | 46 | 98 | 78 |
| *Participants* | 40 | 92 | 155 | 240 | 276 |
| *Experts* | 10 | 10 | 10 | 14 | 14 |
| *Evaluators* | 9 | 9 | 9 | 10 | 10 |
| *Events* | 11 | 16 | 24 | 24 | 15 |
| *Activities integrated in regular courses* | | | | 5 | 13 |

**Table 1:** Inputs from Ekiten case study.

| Outcome | 2010-11 | 2011-12 | 2012-13 | 2013-14 | 2014-15 |
|---|---|---|---|---|---|
| *Ideas* | 10 | 27 | 49 | 105 | 84 |
| *Promoted ideas* | 3 | 2 | 3 | 6 | 8 |
| *Spin-offs* | 0 | 1 | 1 | 1 | 1 |
| *Average rating of ideas* | - | 3.76 | 4.77 | 5.9 | 5.76 |

**Table 2:** Outcomes from Ekiten case study.

## Conclusions

- **1st RQ**: *Can a conceptual model help on replicate successful Idea Contest?*
- Offers data in **semantic format** (enabling interoperability).
- **Visual tools**, ease data analysis for managers.
- Enables **management decisions** impact analysis.
- Better **campaign control/design**.
- Co-creation traceability.

# SEMANTIC WEB TO LINK THE INNOVATION PROCESS WITH INTERNAL AND EXTERNAL REPOSITORIES

## Main Contribution

- Proposal of a method to enable **interoperability** of IMS platforms with semantic **repositories internal and external** to the company.

# Introduction

## Probematic

- Lack of interoperable platforms => **Data silos.**
- Sustainability repository.
  - Energy consumption.
  - GHG emissions.
  - ...

## Approach

- Analyse semantic web methods to link IMSs with internal & external repositories.
- Tools to add information to IMSs from repositories.
- Create richer ideas.

## Use cases

### 1. Energy Reduction

- *I would change the incandescent bulb in desk number 333 for a LED bulb in order to save energy.*
- Identify: *desk number 333*, *incandescent bulb*, *LED bulb.*
- Link to the data in the repository.
- Show location of the desk, and power consumption difference.


**Figure 5:** Energy Reduction and LCA ideas

# Use cases

## 2. Life-cycle Assessment (LCA)

- *If we buy 13-Inch MackBook Air laptops instead of 13-Inch MacBook Pro, we can reduce the amount of GHG emissions in the manufacture of our devices.*
- Identify: *MackBookAir*, *MacBook Pro* and *GHG*.
- Show GHG emissions for both laptops.

## 3. Similar ideas recognition

- Compare idea *mentions*.
- Show the ideas with the largest amount of matchings.
- Why similar ideas have been rejected.

# Platform

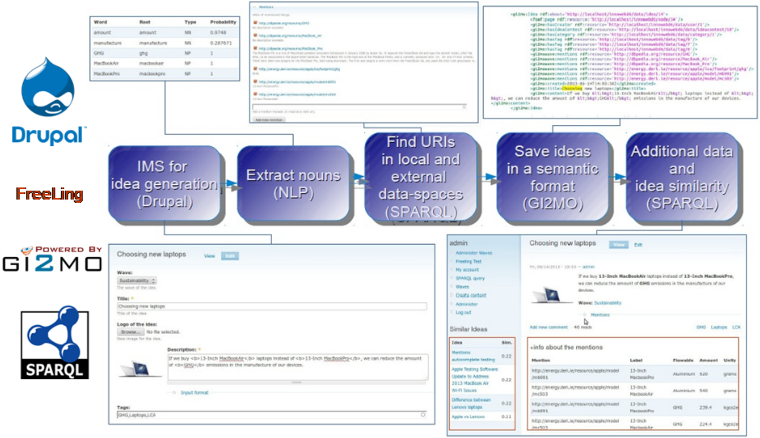



Figure 6: Process to find related elements in Internal and External repositories.

# Platform

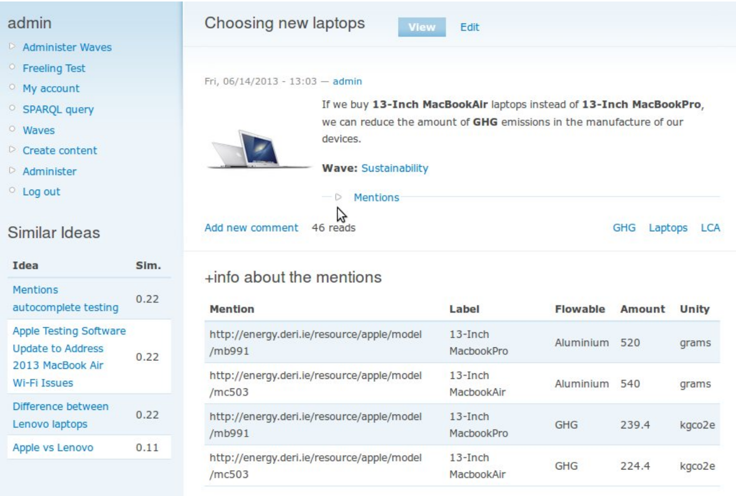



Figure 7: Example of automatically added information.

## Conclusions

- **2nd RQ**: *Can semantic methods enable content linking among IMS platforms and semantic repositories internal and external to the companies?*
- Apply **semantic web** and **LOD technologies** to link sustainability repositories with IMSs.
- Identify elements **mentioned** in the ideas.
- Tools tested with **real sustainability repository**.

# SEMANTIC WEB TO ENHANCE INNOVATION AND TECH. WATCH LINKING

## Main Contributions

- Architecture to enable **interoperability** among Innovation and TW platforms.
- **Mentions Ontology (MO)** to annotate semantically mentioned elements inside text based content.

# Introduction

## Probematic

- Interoperability
  - Linking content in the same platform (idea-idea).
  - **Interoperability** among platforms needed for data re-usability. (e.g. Innovation-TW or Idea-news item)

## Approach

- Propose a **conceptual model** for linking content from IMS and TW platforms.
- Deliver **definitions** from external repositories.
- Show **content similarities**.

# Theory

- A need to enable interoperability among **platforms**[1].
- **Duplicates** make idea assessment time consuming[2].
- Innovators **do not find worthwhile** to search in a huge database before posting their own ideas (proper tools)[3].

*[1] A. Westerski, C. A. Iglesias, and F. T. Rico. A model for integration and interlinking of idea management systems. In 4th Metadata and Semantics Research Conference, volume 108 CCIS, pages 183-194, 2010*

*[2] C. Geffen and K. Judd. Innovation through initiatives: a framework for building new capabilities in public sector research organizations. Journal of Engineering and Technology Management , 21(4):281-306, 2004.*

*[3] E. Ford and S. Mohapatra. Idea de-duplication in an innovation community. Urbana , 51:61801-2302, 2011.*

# Theory

- TW as **starting point** of innovation process[1].
- **News domain** a "clear area where it is important for companies to keep a close eye on technologieal developments in their field"[2].

*[1] B. Fernández Fuentes, S. Pérez Álvarez, F.D. Valle Gastaminza. Metodología para la implantación de sistemas de vigilancia tecnológica y documental: El caso del proyecto INREDIS. Investigación bibliotecnológica, 23(49), 149-177, 2009*

*[2] D. Maynard, M. Yankova, A. Kourakis, and A. Kokossis. Ontology-based information extraction for market monitoring and technology watch. In ESWC Workshop "End User Apects of the Semantic Web", Heraklion, Crete , 2005.*

# Material and Methods

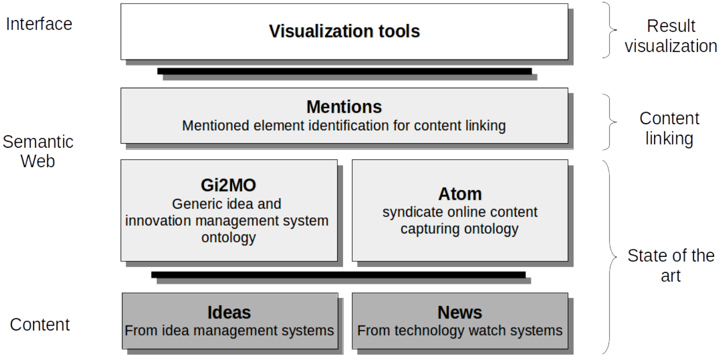



Figure 8: Solution architecture

# Material and Methods

- Based on Modular Uniffied Tagging Ontology (MUTO) [1].
- Methodology for ontology modelling.[1][2]

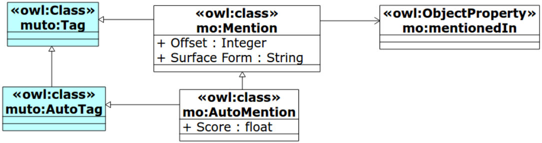



**Figure 9:** Mentions Ontology's[2] Class Diagram.

- DBPedia Spotlight.

*[1] A. Gómez-Pérez. Ontological engineering: A state of the art. Expert Update: Knowledge Based Systems and Applied Artificial Intelligence , 2(3):33-43, 1999.*

*[2] M. Uschold and M. King. Towards a methodology for building ontologies . Citeseer, 1995.*

[1] `http://purl.org/muto/core#`

[2] `https://w3id.org/mo`

# Material and Methods

## Ubuntu Brainstorm (27000)

- 10 most commented.
- 10 not commented.
- 10 highest rated.
- 10 not rated.
- 40 implemented.
- 120 random.

## News feeds (550)

- *OMG! Ubuntu.*
- *Ubuntu Security Notices.*
- *I heart Ubuntu.*
- *Web Upd8.*
- *Ubuntu Manual.*
- 200 random.

### Experiments

1. Concept definition experiment.
2. Idea similarity experiment.
3. News similarity experiment.

# Material and Methods

## Evaluation - Statistical inference

- Size of the sample (200 ideas * 10 links/idea = 2000 links).
- Try to reject a Null hypothesis.

| | |
|---|---|
| **H0 (null hypothesis)** | The system finds low quality relationships |
| **H1 (alternative hypothesis)** | The system does not find low quality relationships |

Table 3: Null and alternative hypothesis for similarity experiments.

# Results

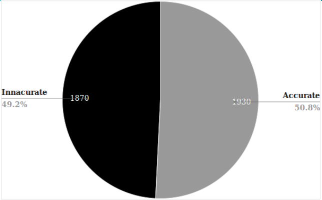



Figure 10: Accuracy of identified concepts from the body.

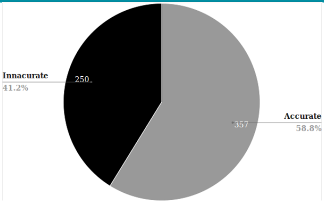



Figure 11: Accuracy of identified concepts from the title.

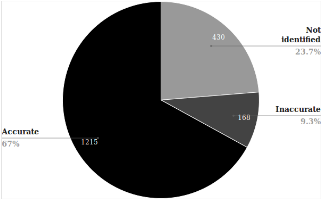



Figure 12: Accuracy of critical concepts from the body.

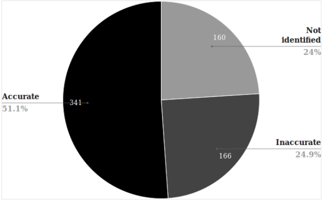
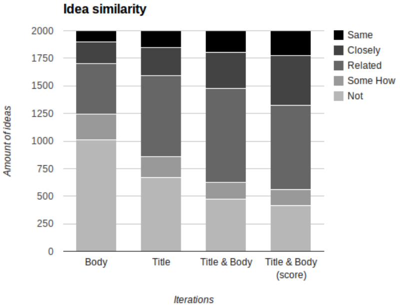



Figure 13: Accuracy of critical concepts from the title.

# Results

| | Body | Title | T&B | T&B+Score |
|---|---|---|---|---|
| Same | 102 | 153 (+51) | 197 (+44) | 228 (+31) |
| Closely related | 194 | 255 (+61) | 330 (+75) | 448 (+118) |
| Related | 463 | 733 (+270) | 851 (+118) | 767 (-84) |
| Some how related | 233 | 193 (-40) | 147 (-46) | 143(-4) |
| Not related | 1008 | 666 (-342) | 475 (-191) | 414 (-61) |

Table 4: Idea relationships in each iteration.



# Results

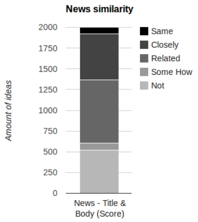



**Figure 14:** News items and ideas relationships.

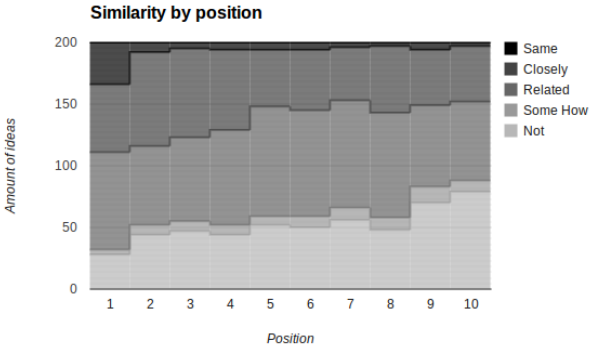



**Figure 15:** News items and ideas relationships by position.

# Results

LOW->0 MEDIUM->0.5 HIGH->1

| | 10 relationships | 5 relationships | 3 relationships |
|---|---|---|---|
| Iu1 | 0.529 | 0.582 | 0.608 |
| Iu2 | 0.530 | 0.584 | 0.612 |

**Table 5:** Statistical inference results by amount of *idea-idea* relationships.

| | 10 relationships | 5 relationships | 3 relationships |
|---|---|---|---|
| Iu1 | 0.506 | 0.560 | 0.591 |
| Iu2 | 0.507 | 0.563 | 0.594 |

**Table 6:** Statistical inference results by amount of *idea-news item* relationships.

# Conclusions

- **3rd RQ**: *Can interoperability among content from Innovation and TW platforms be generalized in a single model?*
- Statistical inference suggest that the architecture is feasible for:
  - Link Ideas among each other.
  - Link Ideas-News items.
- Higher in the list -> More related.
- Best results using *Title&Body+Score*.
- Mentions ontology.

# SEMANTIC TECHNOLOGIES TO ENHANCE TECH. WATCH PRODUCTIVITY

## Main Contributions

- **Reduction of human workload** on document classification tasks in the TW process.
- Identification of best semantic technologies for **automatic document classification**.

# Introduction

## Probematic

- Real TW case study.
- Too much time wasted in document classification.

## Approach

- Semantic Technologies (AI) for document classification.
- Different approaches.
  - Input text (RAW & Semantic annotations).
  - AI algorithms.

## Theory

- NLP tasks over electronic text to treat them[1]:
  - Tokenization.
  - Function words elimination.
  - Stemming.
  - Word frequency.
- AI Algorithms for text mining[2]. 3 selected belonging to different approaches.
  - SVM.
  - Decision Tree (J48).
  - Naive Bayes.

*[1] R. Mitkov. The Oxford handbook of computational linguistics. Oxford University Press, 531:532, 2005.*

*[2] B. Baharudin, L. H. Lee, and K. Khan. A review of machine learning algorithms for text-documents classiffication. Journal of advances in information technology, 1(1):4-20, 2010.*

# Material and Methods

## Dataset 1. Previously categorized documents.

- 7379 documents.
- 14 classes.
- The text of these documents → *RAW Text*.

## Dataset 2. Additional content added by the experts.

- 6968 documents.
- 14 classes.
- Additional content of these documents → *Experts information*.

## Semantic Annotations.

- URIs for Mentioned Elements.
- For both.

# Material and Methods

## Tools

- DBPedia Spotlight → semantic annotation.
- WEKA → text mining (10 fold cross validation).

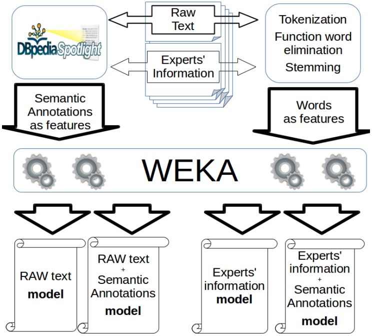



Figure 16: Model generation.

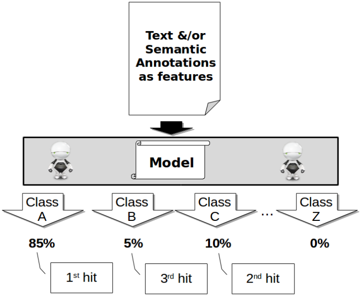



Figure 17: Single document's hits.

## Material and Methods

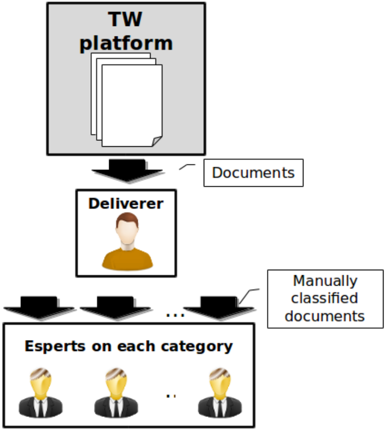



Figure 18: Information flow with no interactions.

# Material and Methods

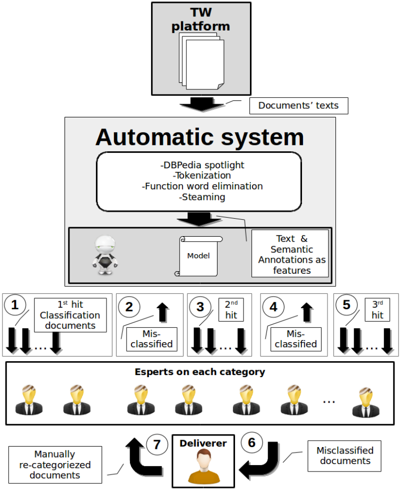



Figure 19: Information flow using automatic system.

# Results

| 7379 Instances | | Raw text (36635 attributes) | | | Raw text + Semantic annotations (46336 attributes) | | | No intervention |
|---|---|---|---|---|---|---|---|---|
| | | J48 | SVM | Naive Bayes | J48 | SVM | Naive Bayes | |
| Accuracy | 1st Hit | **0,978** | 0,947 | 0,908 | 0,977 | 0,946 | 0,908 | 1 |
| | 2nd Hit | 0,969 | 0,900 | 0,908 | 0,968 | 0,898 | 0,908 | |
| | 3rd Hit | 0,963 | 0,837 | 0,908 | 0,964 | 0,836 | 0,908 | |
| Precision | 1st Hit | **0,846** | 0,632 | 0,359 | 0,840 | 0,622 | 0,359 | 1 |
| | 2nd Hit | 0,731 | 0,401 | 0,359 | 0,732 | 0,392 | 0,359 | |
| | 3rd Hit | 0,687 | 0,287 | 0,359 | 0,691 | 0,283 | 0,359 | |
| Recall | 1st Hit | 0,846 | 0,632 | 0,359 | 0,840 | 0,622 | 0,359 | 1 |
| | 2nd Hit | 0,886 | 0,801 | 0,359 | 0,882 | 0,784 | 0,359 | |
| | 3rd Hit | **0,893** | 0,860 | 0,359 | 0,888 | 0,849 | 0,359 | |
| FScore | 1st Hit | **0,846** | 0,632 | 0,359 | 0,840 | 0,622 | 0,359 | 1 |
| | 2nd Hit | 0,801 | 0,534 | 0,359 | 0,800 | 0,523 | 0,359 | |
| | 3rd Hit | 0,776 | 0,430 | 0,359 | 0,777 | 0,425 | 0,359 | |
| Amount of readings | 1st Hit | **9659** | 12807 | 16843 | 9736 | 12956 | 16838 | 14758 |
| | 2nd Hit | 10201 | 13027 | 21565 | 10305 | 13347 | 21568 | |
| | 3rd Hit | 10938 | 13626 | 26292 | 11085 | 13981 | 26298 | |

**Table 7:** Results using “Raw text”, without and with “Semantic Annotations” (1st experiment).

# Results

| 6968 Instances | | Experts' information (6804 attributes) | | | Experts' information + Semantic annotations (8483 attributes) | | | No intervention |
|---|---|---|---|---|---|---|---|---|
| | | J48 | SVM | Naive Bayes | J48 | SVM | Naive Bayes | |
| Accuracy | 1st Hit | **0,975** | 0,958 | 0,935 | 0,974 | 0,962 | 0,931 | 1 |
| | 2nd Hit | 0,958 | 0,912 | 0,934 | 0,960 | 0,914 | 0,931 | |
| | 3rd Hit | 0,948 | 0,849 | 0,934 | 0,950 | 0,850 | 0,931 | |
| Precision | 1st Hit | **0,824** | 0,709 | 0,542 | 0,819 | 0,732 | 0,519 | 1 |
| | 2nd Hit | 0,655 | 0,442 | 0,536 | 0,667 | 0,450 | 0,515 | |
| | 3rd Hit | 0,590 | 0,313 | 0,536 | 0,604 | 0,316 | 0,515 | |
| Recall | 1st Hit | 0,824 | 0,709 | 0,542 | 0,819 | 0,732 | 0,519 | 1 |
| | 2nd Hit | 0,873 | 0,884 | 0,548 | 0,868 | 0,900 | 0,524 | |
| | 3rd Hit | **0,887** | 0,940 | 0,548 | 0,879 | 0,949 | 0,524 | |
| FScore | 1st Hit | **0,824** | 0,709 | 0,542 | 0,819 | 0,732 | 0,519 | 1 |
| | 2nd Hit | 0,748 | 0,589 | 0,542 | 0,754 | 0,600 | 0,519 | |
| | 3rd Hit | 0,709 | 0,470 | 0,542 | 0,716 | 0,475 | 0,519 | |
| Amount of readings | 1st Hit | **9421** | 11019 | 13355 | 9492 | 10708 | 13676 | 13934 |
| | 2nd Hit | 9962 | 10615 | 16459 | 10072 | 10232 | 16962 | |
| | 3rd Hit | 10646 | 10636 | 19606 | 10841 | 10245 | 20282 | |

**Table 8:** Results using "Experts' information" , without and with "Semantic Annotations" (2nd experiment).

# Results

| classified as -> | a | b | c | d | e | f | g | h | i | j | k | l | m | n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 352 | 6 | 7 | 3 | 8 | 13 | 3 | 2 | 8 | 1 | 3 | 1 | 3 | 10 |
| b | 5 | 412 | 26 | 3 | 5 | 5 | 1 | 2 | 4 | 1 | 8 | 2 | 17 | 1 |
| c | 12 | 18 | 766 | 2 | 4 | 6 | 5 | 1 | 5 | 1 | 3 | 0 | 7 | 6 |
| d | 2 | 0 | 2 | 313 | 2 | 1 | 0 | 19 | 0 | 5 | 2 | 1 | 1 | 1 |
| e | 2 | 7 | 4 | 2 | 562 | 16 | 7 | 4 | 126 | 1 | 9 | 8 | 7 | 2 |
| f | 20 | 10 | 3 | 0 | 23 | 737 | 21 | 2 | 7 | 1 | 2 | 3 | 3 | 3 |
| g | 5 | 3 | 3 | 0 | 5 | 21 | 388 | 3 | 11 | 1 | 7 | 7 | 4 | 2 |
| h | 1 | 0 | 2 | 16 | 4 | 0 | 4 | 391 | 12 | 5 | 2 | 1 | 1 | 2 |
| i | 4 | 8 | 4 | 2 | 93 | 5 | 3 | 5 | 377 | 3 | 4 | 2 | 7 | 2 |
| j | 0 | 2 | 0 | 9 | 2 | 4 | 1 | 11 | 0 | 273 | 3 | 0 | 1 | 0 |
| k | 4 | 4 | 7 | 1 | 6 | 5 | 2 | 3 | 15 | 0 | 633 | 3 | 42 | 0 |
| l | 3 | 3 | 2 | 0 | 4 | 3 | 10 | 1 | 5 | 0 | 7 | 449 | 13 | 2 |
| m | 3 | 22 | 11 | 0 | 11 | 4 | 6 | 3 | 5 | 3 | 41 | 5 | 284 | 1 |
| n | 10 | 2 | 7 | 0 | 4 | 6 | 2 | 2 | 0 | 1 | 0 | 1 | 1 | 302 |

**Table 9:** Confusion matrix of the model with the best results (J48 algorithm - 1st hit - RAW text) (Market | Competitors | Technologies | Subsidies).

# Conclusions

- **4th RQ**: *Which are the best semantic technologies that help reducing the time spent on non-productive tasks inside the TW process?*
- Semantic Annotations did not improve results in this case.
- RAW text does not need a treatment.
- Hits only improved using SVM and Experts information.
- J48, 1st hit & "RAW text" gave best results (34.55%).
- Taking into account super-classes, better results.

# CONCLUSIONS

# Conclusions

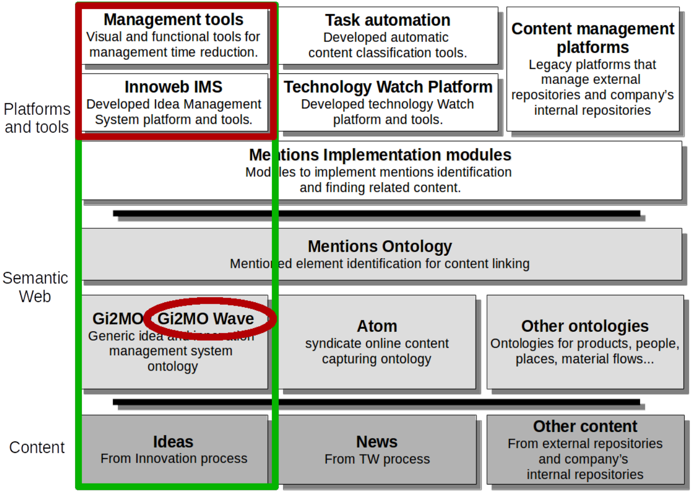



**Figure 20:** Solution architecture implemented in the 1st part of the thesis.

# Conclusions

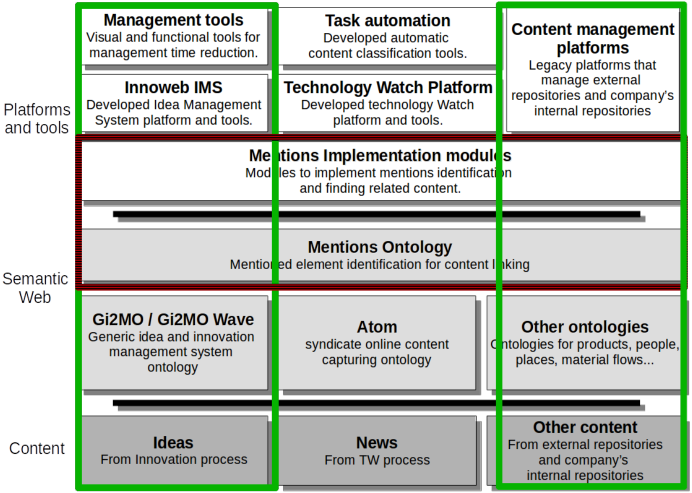



Figure 21: Solution architecture implemented in the 2nd part of the thesis.

# Conclusions

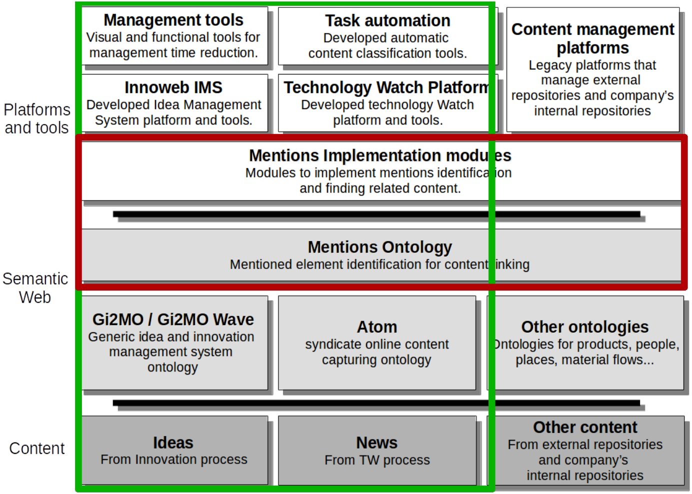



Figure 22: Solution architecture implemented in the 3rd part of the thesis.

# Conclusions

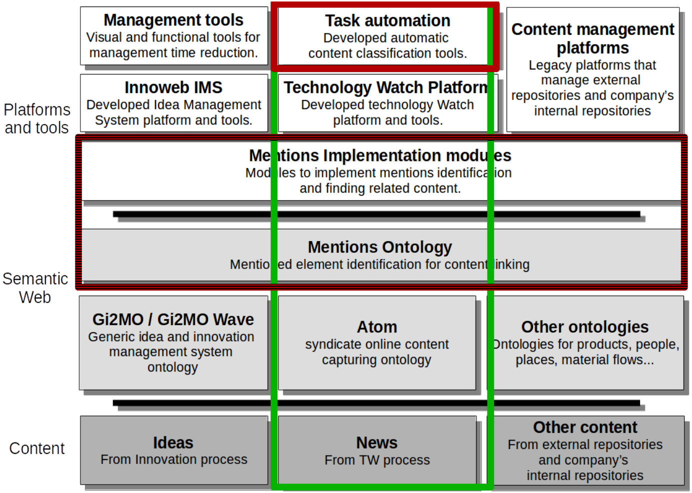



Figure 23: Solution architecture implemented in the 4th part of the thesis.

# Future Work

- Test linking functionalities in other scenarios and content types.
- Case study that implement the whole architecture at once.
- Testing alternatives to DBpedia Spotlight API.
- Clustering for enhance document classification.
- More visual tools development.

# CONTRIBUTIONS

## Cross university and company collaboration

- Digital Enterprise Research Institute (DERI) of the National University of Ireland (NUI)
- Universidad Politécnica de Madrid (UPM)
- ISEA
- Koniker
- Mondragon Unibertsitatea (MU)

# Projects

- SELENE: *Support System for Strategic Decision and Patentability in the Field of Home Energy*
- COLABORANOVA: *Collaborative Innovation System Based on Participation*
- ELKARWEB: *Social and Semantic Web Platform to Support Collaborative Innovation*
- PATENTAWARE: *Support System for Strategic Decision and Patentability in the Field of Home Energy*
- INNES: *Smart Platform for Strategic Innovation in the Field of Health* (Strategin Innovatioin) (MU)
- K-INNES: *Smart Platform for Strategic Innovation in the Field of Health , Servicing for KONIKER*
- ACCELERATE: *A Platform for the Acceleration of GO-TO Market in the ICT Industry*

# Scientific Publications

- *A case study on the use of community platforms for inter-enterprise innovation*, 2011, 17th International Conference on Concurrent Enterprising (ICE 2011)
- *INNOWEB: Gathering the context information of innovation processes with a collaborative social network platform*, 2013, 19th International Conference on Concurrent Enterprising (ICE 2013)
- *The Role of Linked Data and Semantic-Technologies for Sustainability Idea Management*, 2013, 2nd International Symposium on Modelling and Knowledge Management for Sustainable Development (MoKMaSD 2013)
- *Semantic Annotations to enhance Innovation and Technology Watch Interoperability*, pending at International Journal on Semantic Web and Information Systems (IJSWIS)
- *A case study on the use of Machine Learning techniques for supporting Technology Watch*, pending at Data and Knowledge Engineering (DKE) journal

ESKERRIK ASKO
THANK YOU
GRACIAS

aperez@mondragon.edu