



UNIONE EUROPEA



UNIVERSITÀ DEGLI STUDI
DI SALERNO



Ministero dell'Istruzione
dell'Università e Ricerca

UNIVERSITA' DEGLI STUDI SALERNO

SCUOLA DI DOTTORATO
Dottorato in Informatica - XII Ciclo

Ph.D. Thesis

**Knowledge management and Discovery for advanced
Enterprise Knowledge Engineering**

Daniele Novi

Coordinator
Prof. Giuseppe Persiano

Supervisor
Prof. Matteo Gaeta

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to thank my wife Maria Rosaria for her personal support and great patience at all times. My children's Luigi e Raffaele, my mom and my dad have given me their unequivocal support throughout, as always, for which my mere expression of thanks likewise does not suffice.

This thesis would not have been possible without the help, support and patience Professors Matteo Gaeta, Vincenzo Loia and Giuseppe Persiano.

I would also like to convey my thanks to the members of CRMPA and CEMSAC.

Abstract

The research work addresses mainly issues related to the adoption of models, methodologies and knowledge management tools that implement a pervasive use of the latest technologies in the area of Semantic Web for the improvement of business processes and Enterprise 2.0 applications.

The first phase of the research has focused on the study and analysis of the state of the art and the problems of Knowledge Discovery database, paying more attention to the data mining systems. The most innovative approaches which were investigated for the "Enterprise Knowledge Engineering" are listed below.

In detail, the problems analyzed are those relating to architectural aspects and the integration of Legacy Systems (or not). The contribution of research that is intended to give, consists in the identification and definition of a uniform and general model, a "Knowledge Enterprise Model", the original model with respect to the canonical approaches of enterprise architecture (for example with respect to the Object Management - OMG - standard).

The introduction of the tools and principles of Enterprise 2.0 in the company have been investigated and, simultaneously, Semantic Enterprise based appropriate solutions have been defined to the problem of fragmentation of information and improvement of the process of knowledge discovery and functional knowledge sharing.

All studies and analysis are finalized and validated by defining a methodology and related software tools to support, for the improvement of processes related to the life cycles of best practices across the enterprise. Collaborative tools, knowledge modeling, algorithms, knowledge discovery and extraction are applied synergistically to support these processes.

Keywords

Knowledge Extraction, Ontology Extraction, Ontology, OWL, Fuzzy Relational Concept Analysis, Fuzzy Formal Concept Analysis, Semantic Web, Best Practice, Information Retrieval, Automotive.

Table of Contents

Acknowledgements	ii
Abstract	iii
List of Figures	vii
List of Tables.....	ix
Introduction.....	1
1.1 Thesis Outline.....	2
Part I: Theoretical Background	4
2. Fuzzy Theory: Formal Concept Analysis & Fuzzy extension FCA	5
2.1 Formal Concept Analysis (FCA)	6
2.2 Fuzzy extension of FCA	8
3. ENTERPRISE 2.0	11
3.1 The roots.....	12
3.2 Models	13
3.3 Tools	15
3.4 Enabling Technologies	17
3.4.1 <i>Solutions to Social Computing</i>	17
3.4.2 <i>Evolution of traditional information systems</i>	21
3.5 Conclusions	22
4. Semantic Technologies.....	24
4.1 Semantic Web	25
4.1.1 <i>XML and RDF levels</i>	27
4.1.2 <i>The level of ontology languages</i>	29
4.1.3 <i>Subsequent levels</i>	30
4.2 Defining Languages for Ontologies	32
4.2.1 <i>RDF and RDFS</i>	32
4.2.2 <i>OWL and OWL 2</i>	35
4.2.3 <i>SKOS</i>	37
4.3 General Purpose Ontologies of Interest	39
4.4 Conclusions	42
5. Enhanced Enterprise 2.0	43
5.1 Domain-Specific Schemes	45
5.1.1 <i>SKOS</i>	45
5.1.2 <i>FOAF</i>	47
5.1.3 <i>ResumeRDF and DOAC</i>	48
5.1.4 <i>SIOC</i>	48
5.1.5 <i>SCOT</i>	50
5.1.6 <i>MOAT</i>	50
5.1.7 <i>Online Presence Ontology</i>	52
5.1.8 <i>GeoNames</i>	52
5.1.9 <i>DOAP (Description of a Project)</i>	54

5.2 Tools of the Enhanced Enterprise 2.0	54
5.2.1 <i>Semantic Search</i>	54
5.2.2 <i>The semantic wiki</i>	56
5.2.3 <i>Semantic Social Tagging</i>	57
5.2.4 <i>Semantic Social Networks</i>	57
5.2.5 <i>Semantic Microblogging</i>	59
5.3 Conclusions	60
Part II: Methodologies & Applications	62
6. Methodology for Knowledge Extraction and Classification	63
6.1 Methodology for Knowledge Classification.....	63
6.2 Methodology for Knowledge Extraction	67
6.2.1 <i>Process Definition</i>	68
6.2.1.1 Natural Language Processing.....	69
6.2.1.2 Vectorization.....	71
6.2.1.3 Concept Data Analysis	71
6.2.1.4 Semantic Technologies Mapping	72
6.2.2 <i>Evaluation Model for Knowledge Extraction Methodology</i>	73
6.2.2.1 Selected Metrics	73
6.2.2.1.1 Micro-Average Precision	73
6.2.2.1.2 Computation of metrics: Classes to Clusters Algorithm	74
6.2.2.1.3 Structural Quality Metric	75
6.2.2.1.4 Example Datasets.....	76
6.2.2.1.5 Experimental Results	78
6.3 Methodology for Ontology and Instance Matching	81
6.3.1 <i>Process Definition</i>	82
6.3.1.1 Configuration Module	83
6.3.1.2 Natural Language Process Task	84
6.3.1.3 Matching Calculation	85
6.3.1.4 Similarity Functions	85
6.3.1.5 Aggregation and Extraction	86
6.3.2 <i>Matching Strategy : Alignment between different SKOS instances</i>	86
6.3.2.1 Configuration Module	86
6.3.2.2 Natural Language Process tasks	86
6.3.2.3 Matching Calculation	87
6.3.2.4 Similarity Function	88
6.3.2.5 Aggregation and Extraction	89
6.3.3 <i>Matching strategy: Alignment between instances in the same schema</i>	89
6.3.3.1 Selected Metrics	90
6.3.3.2 Selected datasets	91
6.3.3.3 Reference Matching	93
6.3.3.4 Experimental Results	96
Part III: Case Studies	97
7. Definition of models of knowledge in the domain Automotive	98
7.1 Modelling	98
7.2 The process applied to the domain Automotive	101
7.2.1 <i>Preliminary Study</i>	101
7.2.1.1 Best Practice Model	101
7.2.2 <i>Formal definition model</i>	102
7.2.2.1 Best Practice Model	102
7.2.2.2 Domain Business Ontologies	103
7.2.3 <i>Model Evaluation</i>	104

8. Enterprise 2.0 Knowledge Management: Best Practices.....	108
8.1 Knowledge management flow	108
8.2 Brenchmarking and Best Practice	111
8.3 Best practice and business competencies	115
8.4 Knowledge Enterprise Model : Best practice.....	116
8.5 Conclusions	118
9. Conclusion and Future work	119
9.1 Summary	119
9.2 Future Work	119
Bibliography	120

List of Figures

Figure 1 - FCA vs Fuzzy FCA	7
Figure 2 - : The user needs emerging in the Enterprise 2.0 model [25]	14
Figure 3 - Model for the classification of the emerging needs of businesses and 2.0 proposed by the Observatory of the Politecnico di Milano	15
Figure 4 - The model of FLATNESSES Hinchcliffe compared with the model SLATES	17
Figure 5 - Social Network example	20
Figure 6 - Semantic Web Wedding Cake	26
Figure 7 - A simple document RDF / XML	33
Figure 8 - Example of the relationship between level information RDF and those at RDFS	35
Figure 9 - Comparison between OWL and OWL2	37
Figure 10 - Example of serialization concept of love	38
Figure 11 - Serialization concept of love	38
Figure 12 - Linked Data Cloud diagram of the project	42
Figure 13 - From [35], the difference between instruments and SLATES SemSLATES	44
Figure 14 - Example of an RDF graph created using the SKOS	46
Figure 15 - FOAF ontology classes	47
Figure 16 - SIOC ontology classes and properties	49
Figure 17 - Ontology connection between SIOC, SKOS and Foaf	49
Figure 18 - Skos Classes and Properties	50
Figure 19 - Processes interaction KE, IOM and KC	64
Figure 20 - Knowledge Classification Pipe Phase	65
Figure 21 - Rule Based Classifier Phase	66
Figure 22 - Aggregation Classification Result Module	67
Figure 23 - Knowledge Extraction Process	69
Figure 24 - Knowledge Extraction: NLP Pipe	70
Figure 25 - DMOZ homepage	76
Figure 26 - Top category and sub-categories inside DMOZ directory	77
Figure 27 - RDF dump structure	77
Figure 28 - Micro-Average Precision Recall	80
Figure 29: Precision/Recall for specific category (i.e., Consulting)	80
Figure 30 - Ontology and Instance Matching Process	83
Figure 31 - LCS between two concepts in a Taxonomy	88
Figure 32: MSLCS between two concepts in two SKOS Taxonomies	89
Figure 33 - Evaluation model for a matching process	90
Figure 34 - Dmoz_Subset1	92
Figure 35 - Dmoz_Subset2	93
Figure 36 - Semi automatic process to generate a Reference Matching	93
Figure 37 - Results of Micro Average Precision and Recall of Matching Strategy on test datasets	96
Figure 38 - Knowledge Meta Process	98
Figure 39 - Best Practice - Preliminary study	102
Figure 40 - Best Practice Model	103
Figure 41 - Creating classes in Protégé	104
Figure 42 - Relations in Protégé	105
Figure 43 - Graphical representation of the ontology created in Protégé	105
Figure 44 - Assessment of the consistency of the model made by Pellets 2.3.0	106
Figure 45 - Evaluation of the model achieved by Pellet 2.3.0	106
Figure 46 - Spiral of organizational knowledge creation	109
Figure 47 - Seci Model	109
Figure 48 - Four categories of ba	110

<i>Figure 49 - Novice-to-Expert scale</i>	115
<i>Figure 50 - Best Practice Life Cycle</i>	117

List of Tables

<i>Table 1 - From Enterprise 1.0 to Enterprise 2.0</i>	12
<i>Table 2 - The three major categories of Social Networks</i>	20
<i>Table 3 -- Example of Ambiguity representation</i>	28
<i>Table 4 - Input/Output of Knowledge Classification methodology.....</i>	65
<i>Table 5 - Input/Output of Knowledge Extraction process.....</i>	68
<i>Table 6 - Input/Output of NLP Pipe phase.....</i>	70
<i>Table 7 - Input/Output of Vectorization phase.....</i>	71
<i>Table 8 - Input/Output of Concept data Analysis micro-phase.....</i>	71
<i>Table 9 - Main features of Data Analysis techniques</i>	72
<i>Table 10 -Input/Output of Semantic Technologies Mapping micro-phase</i>	73
<i>Table 11 - Example of classes to clusters algorithms</i>	75
<i>Table 12 - Dataset used to evaluate methodology of Knowledge Extraction</i>	79
<i>Table 13 - Structure Quality Metric: Precision and Recall (see Section 4.2.2.1.3 for definitions).....</i>	80
<i>Table 14 - illustrates the overall of Ontology and Instance Matching process</i>	83
<i>Table 15 - Input/Output of the Configuration Module</i>	84
<i>Table 16 - Input/Output of NLP Tasks</i>	85
<i>Table 17 - Input /Output of Matching Calculation</i>	85
<i>Table 18 - Input/Output of Similarity Functions</i>	85
<i>Table 19 - Reference Matching on Dmoz subsets</i>	95
<i>Table 20 - Dreyfus Model and Best Practice</i>	116

Introduction

This section describes the purpose of the doctoral thesis, which consists in the realization of a project aimed at improving the processes related to the life cycles of best practices across the enterprise.

The paradigm Enterprise 2.0 can introduce an innovation process and technology. Currently, however, this type of innovation is technology- driven, that is driven mainly by the use of social networking tools (wiki, blog, forum) collaboration and communication within the company. To achieve the above-mentioned organizational renewal, it is necessary to pursue modeling and methodological innovation that allows companies to capitalize, in terms of reduction of time and costs associated with the business processes and the intangible value (defined as the ideas, experiences, knowledge, attitudes, ability) that the use of Enterprise 2.0 tools escapes more and more to traditional management systems being embedded in social networks business. For this purpose, the research project will analyze issues related to the architectural design and integration of legacy systems (and not). The contribution of the research consists, in particular, in identifying and defining a uniform and general model, a " Knowledge Enterprise Model", which is an original goal in respect with the canonical approaches to enterprise architecture (for example with respect to the Object Management - OMG - standard), in particular for the explanation and indexing of tacit knowledge as best practices.

The survey investigates also approaches to machine learning and text mining to extract concepts and relations from informal textual content resulting from activities related to business processes; examples are personal blogs , chat , forums, and information content deriving from the extraction of the most relevant terms to identify ontological classes which will then be organized into hierarchical structures (taxonomies) or paragraphs (applying different types of semantic relationships) supporting the formalization of the best practices.

From a technological point of view, the goal can be summarized in the definition, design and prototype implementation of a demonstrator aimed at validating the research results, which will be validated in Automotive scenarios, based on the actual industrial needs and requirements identified.

1.1 Thesis Outline

The thesis work is described according to the following structure.

Part I: Theoretical Background

Chapter 2 “*Fuzzy Theory: Fuzzy Formal Concept Analysis & Fuzzy Relational Concept Analysis*” - introduces the mathematical model of Formal Concept Analysis. Furthermore, this chapter defines fuzzy extension of Formal Concept Analysis and Relational Concept Analysis that have been applied in this thesis work in order to manage data uncertainty and conceptualization.

Chapter 3 “*Enterprise 2.0*” - defines a system of web-based technologies that provide rapid and agile collaboration, information sharing, emergence and integration capabilities in the extended enterprise.

Chapter 4 “*Semantic Technologies*” - describes the basics of the semantic technologies and the semantic web standard exploited in this research work.

Chapter 5 “*Enhanced Enterprise 2.0*” – defined collaborative system of Knowledge Management that makes it easier to generate and share information content.

Part II: Methodologies & Applications.

Chapter 6 “*Methodology for Knowledge Extraction and Classification*” - defines the general framework for knowledge extraction. Furthermore, the general framework has been extended and applied in order to support specific research objectives.

Part III: Case Studies.

Chapter 7 “*Definition of models of knowledge in the domain Automotive*” - defined an overall picture of the knowledge models for general purpose representation of resources and ontological models in the domain that will be used for the thesis work

Chapter 8 “*Enterprise 2.0 Knowledge Management: Best Practices*” - defined the art of externalizing enterprise knowledge, i.e., representing the core knowledge of the enterprise.

Part I: Theoretical Background

2. Fuzzy Theory: Formal Concept Analysis & Fuzzy extension FCA

The theory for Knowledge Extraction model based on existing experimental techniques. In Automotive design there is a need to transform natural language text into a format readable by a computer: they must be identified significant portions of the text with the ability to disambiguate polysemous words. Disambiguation techniques used are based on work previously carried out by Milne et al. [1]. The authors describe Wikipedia Miner, an open source toolkit that provides:

- object-oriented structure and content of Wikipedia;
- comparison between semantic concepts and terms;
- detection of topics (items) of Wikipedia, when mentioned in the documents.

Milne explains how machine learning can be used to identify the meaning of terms within unstructured text, enriching it with links to the appropriate Wikipedia articles. The result of the identification and disambiguation is commendable and has a precision of nearly 75%. This performance is constant whether the system is evaluated on Wikipedia articles or documents related to the "real world."

In the research work the texts have to be synthesized into concepts hierarchically organized. The Formal Concept Analysis is widely used in Knowledge Discovery for the automatic acquisition of taxonomies or concept hierarchies from a text corpus. Other authors [2] use an approach that includes Natural Language Processing for:

- text entry;
- construction of a lattice ;
- the conversion of the lattice in a taxonomy .

In addition, the same authors [2] have also compared the performance of the ontologies generated with those created using hierarchical clustering.

The performance of the FCA are best compared to the clustering algorithms because they determine a value of "*recall*" much higher ($> 60\%$), while that for "*precision*" is more or less similar. This is due to the fact that the FCA generates a high number of concepts compared to other clustering algorithms, thus increasing the value of "*recall*".

The Fuzzy Formal Concept Analysis has been the subject of study in recent years. Zhou et al. [3] use Fuzzy FCA for the construction of a model of user behavior, using the log to identify what resources they are more interested in a given period. The Conceptual Data Analysis is strongly linked to the work in [4] [5] [6], in which the extraction of knowledge is performed through Natural Language Processing task on the input text , and through the creation of a Fuzzy Concept Lattice. The use of FCA instead of FFCA allows to take into account the strength of relationships between objects (documents) and attributes (features of text). The ends of knowledge extraction process with the step of Ontology Building, in which the concepts of the lattice are converted into fuzzy classes of an ontology OWL.

The hierarchical conceptualization can also improve the way in which workers have access to the resources in the set of available documents. In [4] and [6] the authors describe the process of building ontology using Facet for viewing. Some initial experiments were conducted to validate the process of construction of ontology and to classify the text. The test was applied on a sample of 443 objects, which represent Web pages or only portions of them. 87% of the entire collection of Web resources has been classified in a consistent manner, while the remaining 13% wrong. From an analysis, there was evidence that some resources reveal the ambiguity in the content because they contain more than one argument.

When techniques are used as FCA and FFCA, the computational complexity becomes an important issue, because the cost of construction of a lattice of concepts is super - linear function to the size of the corresponding context [7] [8]. Thus, the goal is to minimize the input data before constructing the lattice of concepts. The decrease of input data reduces the calculation of the grating by creating a graphical representation smaller.

Cheung and Vogel [9] have proposed the application of a singular value decomposition (SVD) to reduce the size of the matrix. Although SVD provides a good approximation of the matrix, its computational complexity makes it impractical to apply the previous with large arrays [7]. The cluster analysis can thus be used as a method for data reduction.

The Fuzzy C -Means (FCM) having a low computational complexity can be considered as an alternative to SVD to reduce the size of the term - document matrix in applications of Information Retrieval [10]. A recent survey [11] has also demonstrated the superior performance of FCM on other techniques such as hierarchical clustering, partitioning and iterative two-step cluster analysis. For all these reasons Fuzzy C -Means was one of the chosen methods to improve the performance of FCA and FFCA.

Aswani et al. [7] have shown that the reduction of the formal context with FCM is computationally more advantageous compared to the technique of decomposition of the matrix used by SVD. The pattern obtained is reduced also in agreement with the lattice quotient obtained by Cheung and Vogel [9] with SVD.

The Fuzzy C –means [12] is used instead in combination with FFCA. In research carried out the performance of clustering before the construction of the lattice can replace the original files with the centroids of the clusters, greatly reducing the size of the array.

2.1 Formal Concept Analysis (FCA)

Formal Concept Analysis defines a criterion for automatically derive an ontology from a set of objects and properties that bind them. FCA provides a conceptual basis for data analysis and knowledge processing. It allows the representation of relationships between objects and attributes of a given domain. The FCA provides a graphical alternative to the tabular data more natural to navigate and use [13]. The formal concepts can be interpreted by the lattice of concepts latex [14] using the FCA.

Fuzzy Formal Concept Analysis is an extension of FCA with the use Fuzzy Formal Context.

Figure 1 shows a comparison between FCA and Fuzzy FCA .

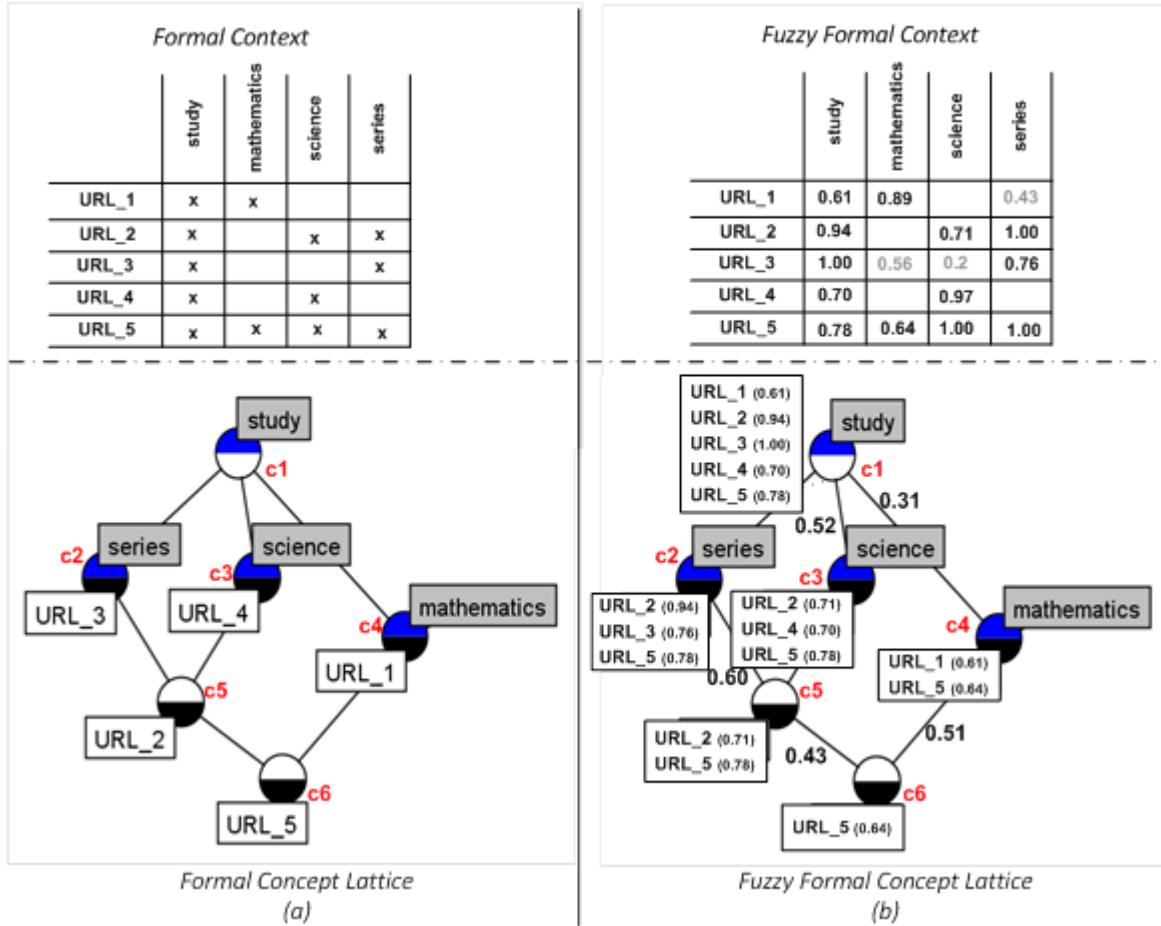


Figure 1 - FCA vs Fuzzy FCA

The following sections describe the mathematical model concealed from both theories. The FCA takes as input a formal context, which can be seen as a binary relation between the set of objects and the set of attributes. The formal context is defined as a triple $K = (G, M, I)$, where G is a set of objects , M is a set of attributes, and I is a binary relation , s. t. $I \subseteq G \times M$. $(g, m) \in I$ read as "the object g has an attribute m". The use of "*object*" and "*attribute*" is indicative, because in many applications it may be useful to choose items such as formal objects and their features as formal attributes. For example, in the methodology defined, the incoming digital resources could be considered formal objects and terms of the Feature Set Cover could be considered as attributes. The context is often represented by a "cross table" in which the rows represent the formal objects and the columns represent the formal attributes , the "x" represent the existence of relations between objects and attributes.

“Definition 1. Formal Concept. Given a context (G, M, I) , per $A \subseteq G$, applying a derivation operator, $A' = \{m \in M \mid \forall g \in A : (g, m) \in I\}$ and for $B \subseteq M$, $B' = \{g \in G \mid \forall m \in B : (g, m) \in I\}$. A formal concept is identified with a pair (A, B) , where $A \subseteq G$, $B \subseteq M$, such that $A' = B$ e $B' = A$. A called the extent and B is called the intent of the concept (A, B) . [15]”

The relation subconcept-superconcept is formalized as follows:

“Definition 2. Given two concepts $c1 = (A1, B1)$ e $c2 = (A2, B2)$, $c1$ è un subconcept di $c2$ (equivalently, $c2$ is a superconcept of $c1$) se $(A1, B1) \leq (A2, B2) \Leftrightarrow A1 \subseteq A2 (\Leftrightarrow B2 \subseteq B1)$. The set of all concept of a particular context, ordered in this way forms a complete lattice [15]”

Figure 1 (a) shows a so-called line diagram of a lattice of concepts corresponding to the formal context in the upper part of the figure. A network of concepts is the set of concepts of a formal context and the subconcept-superconcept relationships between concepts. The nodes represent formal concepts. The formal objects are labeled below and above the formal attributes of the nodes that label (see below). Note that the names of $c1$, $c2$, etc. represent identifiers concept. These identifiers have been introduced to reference easily concepts, but are not part of the representation of the lattice. For example, under this name, the node labeled with the 'formal attribute "series" and a formal object "*URL_3*" is referred to as $c2$.

To identify the extent of a formal concept one needs to trace all the paths that lead down from the node to collect the formal objects. In Figure 1(a), the formal objects of $c2$ are *URL_3*, *URL_2* and *URL_5*. To identify the intension of a formal concept, however, one needs to trace all paths in order to collect all the formal attributes. In the example, there is a node above $c2$ "study" as formal attributes attached. Thus $c2$ is the formal concept with the extension "*URL_3*, *URL_2*, *URL_5*" and the intension "study, series". Finally, $c2$ is a sub -concept of $c1$. The report subconcept-superconcept is transitive: a concept is subconcept of all concepts that can be achieved through the lattice from the top right down to it. If a formal concept has a formal attribute then its attributes are inherited by all its subconcept. This corresponds to the notion of "*inheritance*". In fact, the network can also support multiple inheritance relationships.

2.2 Fuzzy extension of FCA

Recently, the FCA has been used in many applications whose domain includes representation of information is uncertain and vague [6]. Pioneering studies exploit the "fuzziness" in FCA generalizing the model of Wille [16] for formal contexts of FCA or, more importantly, take advantage of the extension of the original analysis of the formal concept by setting the degree of truth for the propositions "the object x has an attribute y" contexts" fuzzy formal "using a" residuated lattice¹ [17], [18]. The degrees are taken from a scale L of truth degrees. Usually, L is evaluated with real values

¹ In abstract algebra a "residuated lattice" is an algebraic structure which is both a lattice $x \leq y$ and $x \bullet y$ a monoid which admits the operations $x \setminus z$ and z / y is roughly analogous to the division or implicitly when $x \bullet y$ is seen respectively as multiplication or conjunction.

in $[0, 1]$. So the rumors of a table that describes the objects and attributes become degrees of L instead of values $\{0, 1\}$ as in the case of the basic settings of the FCA. This extension is known as Fuzzy Formal Concept Analysis (FFCA).

Definition 3. A **Fuzzy Formal Context** is a triple $K = (G, M, I = \varphi(G \times M))$, where G is a set of objects, M is a set of attributes, and I is a fuzzy set on domain $G \times M$. Each pair $(g, m) \in I$ has a membership value $\mu_I(g, m)$ in $[0, 1]$.

The set $I = \varphi(G \times M) = \{(g, m), \mu_I(g, m) | \forall g \in G, m \in M \mu_I: G \times M \rightarrow [0, 1]\}$ is a fuzzy relation $G \times M$.

Definition 4. Fuzzy Representation of Object. Each object g in a fuzzy formal context K can be represented by a fuzzy set $\Phi(g)$ as $\Phi(g) = \{(m_1, \mu_I(m_1)), (m_2, \mu_I(m_2)), \dots, (m_m, \mu_I(m_m))\}$ where $\{m_1, m_2, \dots, m_m\}$ is a set of attributes in K and $\mu_I(m_i)$ is the membership to attribute m_i . $\Phi(g)$ is called the fuzzy representation of g .

Figure 1 (b) is shown a fuzzy version of the formal context through a "cross table". According to the fuzzy theory, the definition Fuzzy Formal Concept is the following.

Definition 5. Fuzzy Formal Concept. Given a fuzzy formal context $K = (G, M, I)$ and a confidence threshold T , we define $A^* = \{m \in M | \forall g \in A: \mu_I(g, m) \geq T\}$ for $A \subseteq G$ e $B^* = \{g \in G | \forall m \in B: \mu_I(g, m) \geq T\}$ for $B \subseteq M$. A fuzzy formal concept (or fuzzy concept) A_f , of fuzzy concept K with confidence threshold T , is a pair $(\varphi(A), B)$, where $A \subseteq G$, $\varphi(A) = \{g, \mu_{\varphi(A)}(g) | \forall g \in A\}$, $B \subseteq M$, $A^*=B$ and $B^*=A$. Each object g has a membership $\mu_{\varphi(A)}(g)$ defined:

$$\mu_{\varphi(A)}(g) = \min_{m \in B} \mu_I(g, m)$$

where $\mu_I(g, m)$ is the membership value between object g and attribute m , which is defined in I ; Note that if $B=\{\}$ then $\mu_g = 1$ for every g . A and B are the extent and intent of a formal concept $(\varphi(A), B)$ respectively.

Figure 1 (b) the formal context fuzzy has a confidence threshold $T = 0.6$ (as said, all the relationships between objects and attributes with membership values less than 0.6 are not shown).

Definition 6. Let $(\varphi(A_1), B_1)$ and $(\varphi(A_2), B_2)$ be two fuzzy concepts of a fuzzy formal context (G, M, I) . $(\varphi(A_1), B_1)$ is the sub-concept of $(\varphi(A_2), B_2)$, denoted as $(\varphi(A_1), B_1) \leq (\varphi(A_2), B_2)$, if and only if $\varphi(A_1) \subseteq \varphi(A_2)$ ($\Leftrightarrow B_2 \subseteq B_1$). Equivalently, $(\varphi(A_2), B_2)$ is the Super-concept of $(\varphi(A_1), B_1)$.

For example, observe that in (b) the concept C_5 is a sub-concept of the concepts c_2 and c_3 . Equivalently concepts C_2 and C_3 are super-concept of the concept c_5 .

Definition 7. A **Fuzzy Concept Lattice** of a fuzzy formal context K with confidence threshold T is a set $F(K)$ of all concepts of K with the partial order \leq with confidence threshold T .

The theory of the FCA proposes a hierarchical model in which the concepts (objects and their attributes) are arranged in a subsumption relation (known as the "hyponym-hypernym" or "is-a"). The grid shows the formal fuzzy membership associated with the objects and the class-subclass relationship. More formally:

Definizione 8. *The Fuzzy Formal Concept Similarity between concept $K_1=(\varphi(A_1), B_1)$ and its subconcept $K_2=(\varphi(A_2), B_2)$ is defined as:*

$$E(K_1, K_2) = \frac{|\varphi(A_1) \cap \varphi(A_2)|}{|\varphi(A_1) \cup \varphi(A_2)|}$$

where \cap and \cup refer intersection and union operator²; on fuzzy sets respectively.

As an example, the Fuzzy Formal Concept Similarity computed between the concept $c_2 = \{(URL_2, URL_3, URL_5), (series, study)\}$, and the concept $c_5 = \{(URL_2, URL_5), (science, study, series)\}$, Figure 1 (b) shown is:

$$E(c_2, c_5) = \frac{|(\min\{0.71, 0.94\}) + (\min\{0.78, 0.78\})|}{|(\max\{0.71, 0.94\}) + (\max\{0.78, 0.78\}) + (\max\{0.76\})|} = 0.60$$

Comparing FCA with the FFCA highlighting the differences in the modeling of the two methods on the same sample of objects (e.g., web resources). In the classical FCA, the matrix that represents the formal context contains binary values that indicate the existence of relations between objects and attributes. In the table "fuzzy" corresponding each cell contains a value in the range [0, 1] through which provides an estimate of the strength of the bond expressed by the relation between the object and attribute. The lattice of concepts provides a mathematical modeling of knowledge that contains more information than the traditional tree-like conceptual structure [19]. Compared to the formal lattice, the lattice fuzzy introduces more information on how you structure the knowledge and existing relationships, such as the "fuzziness" is included with each item/resource and the similarities between the concepts fuzzy formal concept

² The intersection and union fuzzy, respectively are calculated using the t-norm and the t-conorm. The t-norm most commonly used is the "minimum", while the t-conorm most common is the "maximum". That is, given two fuzzy set A and B with membership functions $\mu_A(x)$ and $\mu_B(x)$ $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$ and $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$.

3. ENTERPRISE 2.0

The Enterprise 2.0 is a tool for business innovation, which aims to make organizations more open, flexible and dynamic, which makes creativity and collaboration between people a key factor in the achievement and the improvement of the business objectives of companies. The collaboration of individuals, businesses and things is the new foundation of competitiveness, since it can lead to new ideas, creativity and, consequently, innovation, universally regarded as the starting point for improving the competitiveness, helping to improve the performance of the processes that characterize the corporate structure .The optical Enterprise 2.0 moves his feet from the consideration that companies do not possess internally all the intellectual capital necessary to innovate and so, you can benefit from the knowledge and know-how, present in markets and networks, to improve and evolve their innovation policies. The Enterprise 2.0 aims to revolutionize the traditional way of thinking about the organization passing from:

- a hierarchical model, typically a top-down to another based on cooperation, typically bottom-up
- a technology-driven innovation to a user-driven innovation,
- the formation of a centralized team to that of distributed teams, breaking, in this sense, real geographical barriers,
- an organization, content and business documents, based on taxonomies (classification created by experts) to one based on folksonomies (classification created by users)
- an opening of the boundaries of organizations to the outside,
- an open knowledge management ,
- a greater flexibility of business models a reduction of time to market, or the time it takes to turn the company a new opportunity of market to product,
- a corporate knowledge management ,enriched by the collaboration and the exchange of information created by users and external to an organization.

The steps indicated express the evolution of Enterprise 1.0, associated with the traditional business models of enterprises, Enterprise 2.0, associated with new business models arising from the collaboration of activities and the exchange of users knowledge (Table 1).

Enterprise 1.0	Enterprise 2.0
Hierarchy	Flat Organization
Friction	Ease of Organization Flow
Bureaucracy	Agility
Inflexibility	Flexibility
IT-driven technology / Lack of user control	User-driven technology
Top down	Bottom up
Centralized	Distributed
Teams are in one building / one time zone	Teams are global
Silos and boundaries	Fuzzy boundaries, open borders
Need to know	Transparency
Information systems are structured and dictated	Information systems are emergent
Taxonomies	Folksonomies
Overly complex	Simple
Closed/ proprietary standards	Open
Scheduled	On Demand
Long time-to-market cycles	Short time-to-market cycles

Table 1 - From Enterprise 1.0 to Enterprise 2.0

Will be defined a methodology for Knowledge Classification able to associate the resources available to known categories, belonging to a classification created and updated based on the contents of the input resources. The latter can be characterized by unstructured data (documents, presentations, etc. available within the organization), but also by instances of the models defined on the domain of interest.

3.1 The roots

The social and the collaborative aspect among the employees of an organization is the fulcrum around which they are born and evolved new models of business. The user goes from being a mere user to a creator of business knowledge.

The emphasis on individuals and their ability to interact, collaborate and share information is certainly the aspect that distinguishes the new form of social interaction enabled by Web 2.0 technologies. It is precisely from the approaches and tools of Web 2.0 that Enterprise 2.0 reuses collaborative technologies and derives new organizational models. Web 2.0 has given rise to a new way of conceiving the network, triggering a mechanism aimed at change also the business and the consumption models, with important consequences for both consumers and businesses. Thanks to social platforms, software, allowing communication between users and the creation of information from it, it has established a new mode of knowledge construction, which is based primarily on the interaction. Blogs, wikis, feed RSS and folksonomies are classic examples of that set of tools from Web 2.0, responsible for the reconfiguration of social relations and the way people communicate. A simplistic view frames the Enterprise 2.0 as a corporate application of technologies and services that characterize Web 2.0. A more critical view, however, emphasizes the Enterprise 2.0 as a fundamental shift in the way the organizations operate, that is, to their way of doing business. [20]. Enterprise 2.0 is a social phenomenon driven, a profound change in the relationship between the employees of an organization and between these people and the tools that they usually seek in their

work environment .This brief introduction allows us to understand the meaning of the term coined Enterprise 2.0 and explained by Professor Andrew McAfee: "use of social software platforms within companies in emerging way or between companies and their partners and customers," [21].

Based on the definition of McAfee is possible to identify the key elements of this new organizational and business paradigm:

- **The platforms:** digital environments in which the interaction between users within the company and their contributions are visible to all participants in a persistent manner over time.
- **Social software:** set of tools that allow people who are part of a company to get in touch and work together, activating, using computer technology, real communities internal to the organizational context.
- **The emergency:** the ability to bring out information from the interaction between people. [22]

Using again the words of McAfee, "Enterprise 2.0 technologies make the intranet more similar to what the web is already: an online platform, with an evolving structure, determined in a distributed and independent way from the users actions". [23]

3.2 Models

The user and his emerging needs are the focus of Enterprise 2.0 models. According to a study by the Observatory of the Politecnico di Milano the emerging needs that the authors have identified can be classified into six types [24]:

- **Membership open:** the need to open the borders to the outside of your organization, involving subjects with which we relate on a daily basis, such as suppliers, partners and customers.
- **Social networking:** the need to create professional relationships internally and externally, through evolved profiles through which trace people.
- **Emerging collaboration:** the need for cooperation, which is essential in any work activity. Ability to reconfigure adaptive: the ability to reconfigure their processes in order to keep them in line with the strategic objectives in rapid and continuous change.
- **Knowledge on the Web:** the need to be able to get knowledge directly from the network is adequately managing the tacit and explicit.
- **Global mobility:** the need for connectivity to their work environment, anywhere and at any time, independent of the user's location in order to share and access to knowledge of the company.

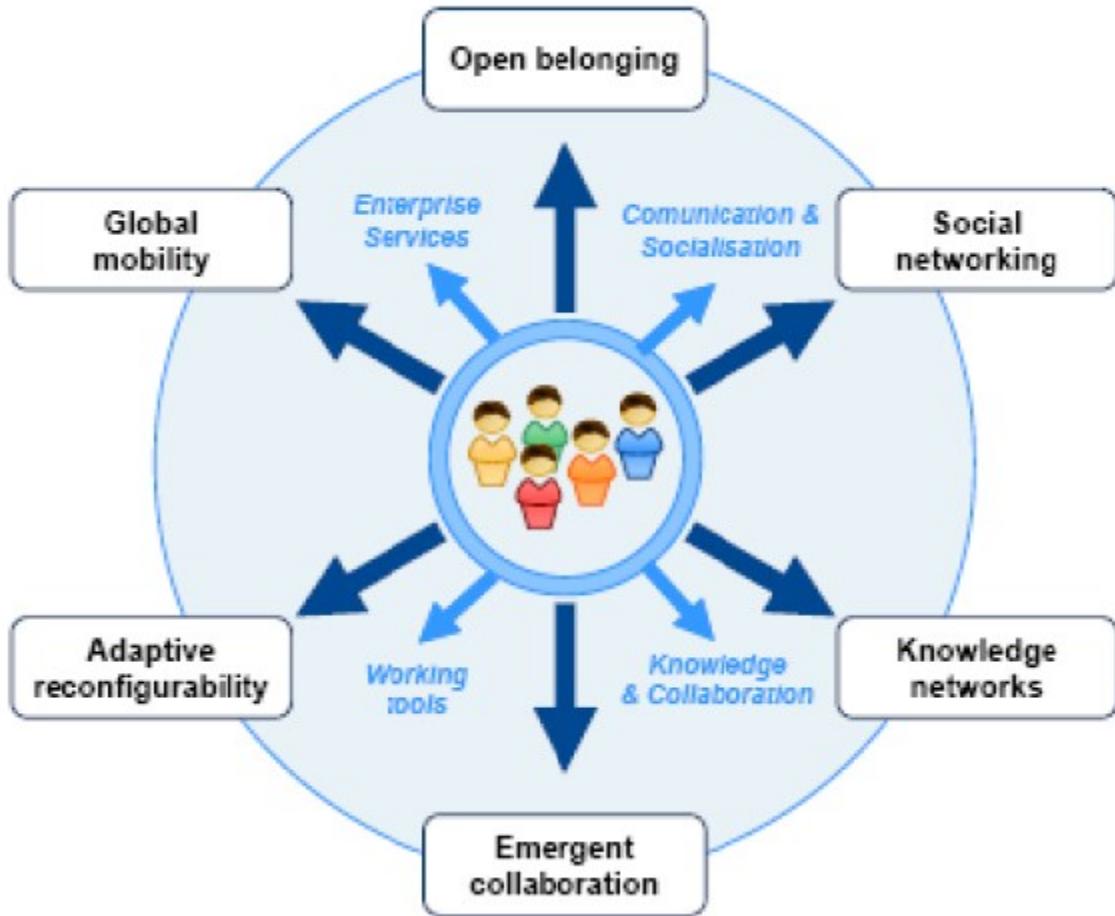


Figure 2 - : The user needs emerging in the Enterprise 2.0 model [25]

These needs (see Figure 2) are already present or they are springing up in most organizations, although to different degrees. In light of these six dimensions that characterize the needs of users, the authors of the study conducted by the Observatory of the Politecnico di Milano 2.0, have identified three business models [24]:

- **Social Enterprise:** business profile that points to the creation of new patterns of collaboration, sharing of knowledge and relationship management aimed at overcoming the barriers of hierarchical organizations.
- **Adaptive Enterprise:** business profile that aims to support, by increasing flexibility, in an adaptive business processes.
- **Open Enterprise:** business profile, where there is a constant exchange of content and information with outside. For this profile it becomes imperative to orchestrate an IT infrastructure that interfaces with customers, suppliers, partners, and consultants from which they can often get real process innovations or product. This model often meets the needs of mobility, connecting people scattered on the territory to their company network.

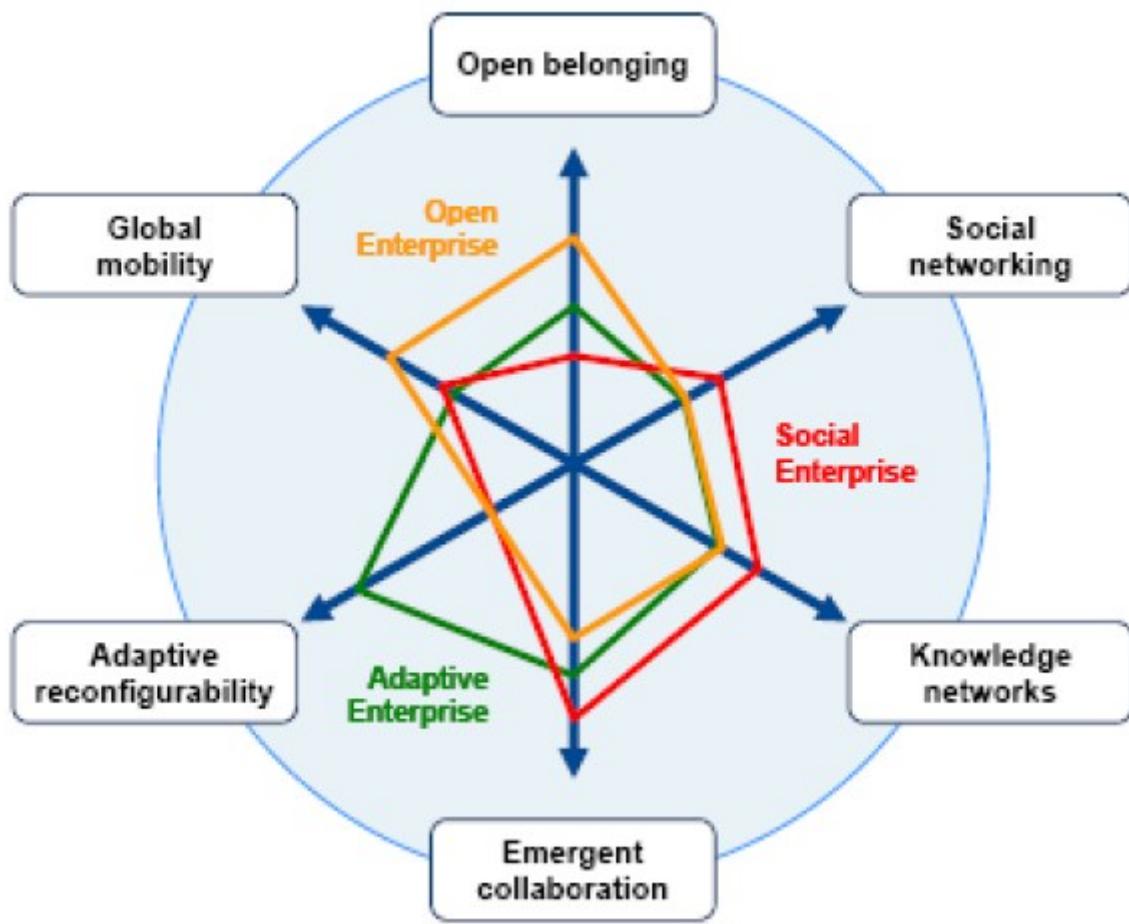


Figure 3 - Model for the classification of the emerging needs of businesses and 2.0 proposed by the Observatory of the Politecnico di Milano

Figure 3 shows the model proposed by the Observatory of the Politecnico di Milano 2.0, which shows the mapping between the three models of Enterprise 2.0 and the needs of users within an organization.

3.3 Tools

The definition of the term Enterprise 2.0 that was coined by Professor McAfee, is characterized by the use of social software platforms in order to provide a digital environment in which contributions and interactions are made available over time. In this perspective, the tools and technologies that characterize a social software platform in the enterprise are identified by the acronym **SLATES** that defines the core elements [21] [26]:

- **Search:** an internal search engine in order to introduce in the system the same way for business research typically used on the Internet. In particular, it must be implemented a search system more based on keywords than on the browsing of documents.
- **Links :** ability to add internal or external links . The platform should give the possibility to link the content between them, in order to exploit the information content inherent in the same link. The value of content is also determined by the number of links to it, similar to what

happens with the PageRank algorithm of Google. For this purpose it is important that all users, not just administrators, could create links between resources.

- **Authorship:** ability to create and edit documents and content. The massive use of blogs and wikis on the Internet lets you easily reflect the intention that many people have in expressing their ideas through appropriate communication tools. The set of posts, comments and contributions are an enormous knowledge base that, without the introduction of these tool could never be. With the excellent results obtained by these tools on the Internet, just simply think of Wikipedia, it is clear that these must also be introduced in the company, in order to improve the content in a collaborative way.
- **Tags:** ability to categorize documents and content, through the affixing, by user, tags, or labels. In this way it is possible to structure the contents implicitly, by using directly the users' actions (tagging users). This technique produces categorization from the bottom of folksonomies, or a categorization of information generated by users through the use of keywords (tags), freely chosen, which are opposed to the traditional taxonomies, which are a categorization of information developed by experienced users. The advantage of folksonomies is to reflect the information structures actually adopted by users, rather than those designed by the creators of the content.
- **Extensions:** ability to add features that enable it to identify and suggest intelligently patterns of use. Using algorithms of affinity between the content, the company's platform should be able to recommend to a user of the contents similar to those already enjoyed.
- **Signals:** ability to notify users e-mail or RSS feed, all the changes of interest.

The elements highlighted by the acronym SLATES have a concept of Enterprise 2.0 exclusively linked to the technological aspects of Web 2.0. Previously, it has already been pointed out that Enterprise 2.0 does not want to bring only the tools and technologies of Web 2.0 in the enterprise, especially as changing of the traditional business logic. In order to identify also the methods and the strategies that companies must follow and take to reconfigure its organizational logic with an optical Enterprise 2.0, the scholar Dion Hinchcliffe has integrated into the model SLATES a more complex model which is called FLATNESSES [27]. This model adds to the model SLATES aspects that characterize the size and the organizational culture of an organization:

- **Freeform:** no barriers to authorship, the creation of documents and content. The format of the content should be as free as possible, avoiding overly constrain users.
- **Network-oriented:** orientation to the collaborative network and sharing between users of this organizational and cultural model in the company. Not only the application of the platform must obviously be network-oriented, but these must be also the documents, in the sense that they must be designed and structured for the network.
- **Social:** emphasizing of relationships, collaboration and sharing between users. All content and all contributions must be associated with a user with a public profile, so you can make

the reputation management. Must be permitted communication between users, whether public or private.

- **Emergence:** spontaneous dynamics and bottom-up without structures imposed from above. A system that is based on the Enterprise 2.0 must unleash the business knowledge in a structured way, starting from the interaction and the communication of the users.

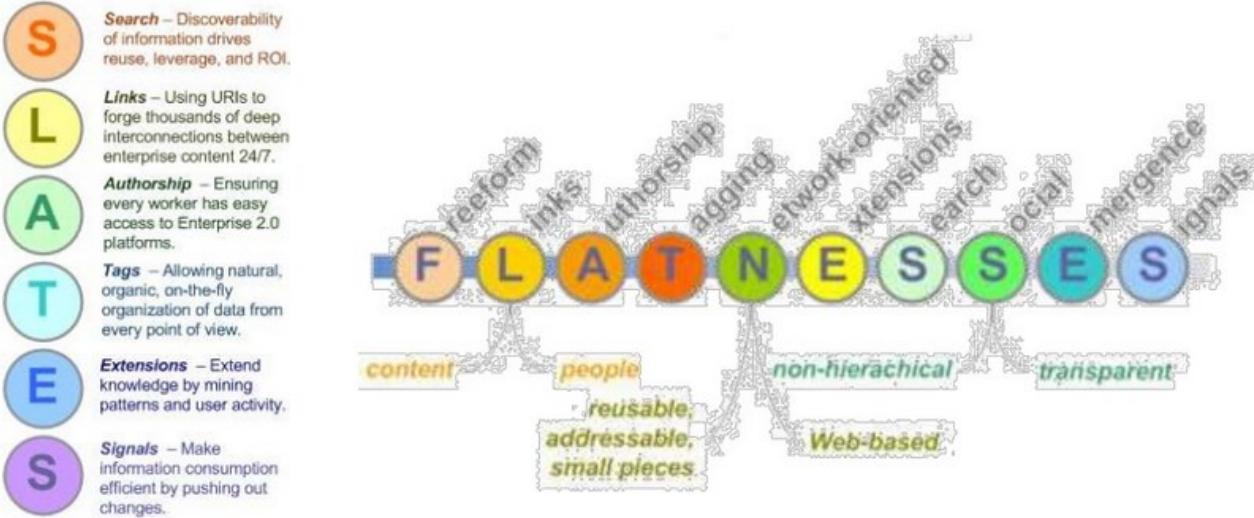


Figure 4 - The model of FLATNESSES Hinchcliffe compared with the model SLATES

3.4 Enabling Technologies

According to the Observatory Enterprise 2.0 [28], the answer involves both technological solutions for **social computing** (wikis, blogs, microblogging, social networking, social tagging, RSS feeds, podcasting, video sharing, instant messaging, etc.). Solutions aimed at 'evolution of traditional informative systems (SOA - Service Oriented Architectures and BPM - Business Process Management), which found new life thanks to Mashup and the delivery of services in SaaS - Software as a Service [20].

3.4.1 Solutions to Social Computing

To Instruments of Social Computing are a significant support for companies in order to simplify, improve and optimize the management and the organization of strategic information for the organization. These are tools that help companies to approach an Enterprise 2.0 through a review of its organizational model geared to participation and sharing, at the opening of business boundaries, the establishment of the dynamics of collaboration between the company, people who are part and external communities.

Below is an overview of the tools of social computing, all from the world of Web 2.0, reusable and configurable within companies in an optical Enterprise 2.0:

- **Collaborative Tagging, Folksonomies, RSS Feeds and Social Bookmarking** [29] [30]: the collaborative tagging is characterized by the possibility that users have to label (or

associate tags) documents and content. This work paves the way for a new model of classification and organization of information within a company's information system, the folksonomies. This technique of categorizing the content, structures the most strategic and relevant information within the organization, using directly the actions of users, the tagging of users. It is a mechanism that is opposed to the traditional categorization policy expressed by taxonomies, based on an organization of documents and hierarchical content. Therefore, the folksonomies expresses a bottom-up approach to nature , since it allows the same individuals within the company to classify content according to label it thinks fit ; taxonomies, however, express a nature approach to top down , since they only allow experienced users to classify content . The advantage of the collaborative tagging and classifying information through folksonomies is expressed in the facility of finding information. In a scenery designed to speed up and improve the quality of information obtained involved the RSS (Really Simple Syndication). These are instant notifications on creation of new content on the platforms that it was decided to monitor (New Story of an online journal, blog post, etc.). Which allow the following link to information content provider. RSS does not feed only to speed up the research and the information retrieval but also make faster and more productive the acquisition and the use of relevant content in the company. The management of information through the system of tagging and the opportunities related to the classification through the folksonomies are frequently exploited by individuals through these social bookmarking platforms. These allow people to store, manage and share their bookmarks (bookmark) on the web. Bookmarks are associated with links or web pages with tags that characterize the content. Through social bookmarking each user has a personal library of bookmarks, indexed by tag, share with other users, facilitating the search for some content. The sharing of bookmarks is also possible through the use of RSS feeds.

- **Corporate Blog and Blog** [26]: have become, within a few years, one of the most popular tools of the Network, simplifying the process of content production from the users and making more flexible the use of the information. As part of Enterprise 2.0, the Blog is able to offer renewed opportunities for companies both as an innovative channel being configured to create an actual connection, direct and transparent link between the people who are part of the company and external stakeholders (the so-called Corporate Blog) , and as a tool for collaboration and sharing of design information, experiences and knowledge between business departments , workgroups and individual employees , and employees. All the most famous organizations (like Google) have begun to use blogs for the global spread of news instead of organizing press conferences. Blogs allow immediate use, can potentially reach a large number of users, are low-cost and offer the opportunity for a two-way transfer of information. blog , therefore , are considered tools of communication and interaction that can enable companies to get closer to the organizational model of Enterprise 2.0 , both from the point of view of opening up to the outside corporate boundaries , both as part of ' internal

organization, where they represent and considering that they represent media ,and applications that stimulate and facilitate the processes of communication, relationship, cooperation and sharing of knowledge among the people who are part of the company.

- **Microblogging and Activity Stream:** Microblogging gives you the ability to post brief information on their status, or on its activities and subscribe to notifications of updates of other users of interest. The microblogging, made famous by Twitter, can increase its utility by the semantic enrichment of distributed information. Microblogging is a tool similar to the Activity Stream, which is a notification system that provides a user updates on the activities of people of interest. The participants in the community can choose which part of their activity streams visible to particular users. The majority of Enterprise Social Software supports notifications through Really Simple Syndication (RSS).
- **Wiki** [29] [30]: are systems for content management products by many people , which are used within the company , may allow employees and contractors to integrate , monitor , edit or delete information, experiences and knowledge, in a ' perspective of sharing and collaboration continues. In this direction, it is possible to imagine a corporate Wiki as a bulletin board type of participatory that is updated in real time by all members within the company. Businesses can use the flexibility and the versatility of participatory platforms such as Wikis to tap into and leverage the collective intelligence by people living outside the organization, such as that of partner and professional communities. This process is known in the jargon as a process of crowd sourcing. It is increasingly common, in fact, the case in which the company relies on community ad hoc management of business processes to achieve successful results in terms of quality, speed and costs. A wicked -known wiki is Wikipedia.
- **Social network** [29] [30]: it consists in an online space where each user creates his own profile and connect to one other in order to share information and content with their network of acquaintances. The outstanding advantage of social networks is the ability to relate people with the same interests, passions, attitudes, creating bonds that can have feedback working. The Social Network Social Computing incorporate other instruments such as tags, bookmarking, RSS feeds, etc.. Figure 5 are shown examples of well-known social networks such as Facebook, mySpace, Flickr e Linkedin



Figure 5 - Social Network example

Social Network can be divided into three broad categories

AGGREGATORS MEDIA	SOCIAL AGGREGATORS	PROFESSIONAL AGGREGATORS
The first of which is that of the media aggregators that allow to seek and download material :photos(flickr) ,music(last.fm) ,video (you tube) ,news and knowledge (wikipedia) or simple links.	The second category is that of social aggregators that is used to meet people and interact with them(facebook,myspace;Friendster;second life)	The third relates to professional aggregators born with the aim to expand their personal network to provide meeting and gathering places as forum ,events, communities,formation ,observatory and so on (eccellere business community,manager zen ,customer, think ,mymarketing.net)

Table 2 - The three major categories of Social Networks

The characteristics of social networks allow companies to employ them at different levels, both within the organization and outside the contact and relationship with stakeholders. The use of social networks in the enterprise improves the system of relationship between:

- employees

- business units or departments
 - workers and professional communities to which they belong
 - company and supply chain partners
 - public and company
 - demand and supply of labor.
- **Podcasting** [29] is the ability to access to the audio files available for download both streaming and local download .This tool abilities to the opportunity of learning content through audio and visual media.
 - **videosharing** [29] is the ability to access video streaming that is available in local download. This tool abilities the opportunity of learning content through audio and visual media. . A well-known example is given by video-sharing YouTube.
 - **Document sharing** [29]: There are many platforms to work shared a single version of a document. An example is the Google Documents application that allows you to work remotely on a document that can be shared with a group of people. This allows you to have a single version of the documents in such a way that everyone can always work on the latest version. The comments and changes are tracked so that they may or may not be accepted by the other participants in the working group.
 - **Chat and instant messaging and conferencing systems** [29] : services that enable a dialogue between users in real time. The web conferencing software include slide presentations, messages and real-time chat, VoIP audio, recording, etc.. An important example that can resize telecommunications traditional systems Skype, is used in both private and corporate.

3.4.2 Evolution of traditional information systems

The use of enabling technologies, Enterprise 2.0 within organizations, requires modification of interventions and also update the information system of the organization. The goal is to make enterprise information systems capable of responding quickly to the needs of the users. With this approach, if the past were organizations to converge towards Web technologies, allowing access only through the browser, today is the web that flows to the enterprise, allowing access to the information system through a wider range of channels and media [30]. Business users access the system through new applications and interaction models, are an example of the RWA (Rich Web Application), which make accessible from the browser functionality and capabilities of typical desktop applications and RIAs (Rich Internet Application), which make usable web applications outside of the browser.

The ' Enterprise 2.0 also opens the way for the adoption of new approaches and infrastructure applications such as SOA (Service Oriented Architecture) , as well as new models offered as SaaS (Software- as-a- Service). The first introduces an architectural evolution in informative systems

aimed at defining a strategy for the development of applications such as service composition with quite specific characteristics oriented to the reuse and the integration. The second introduces a software application delivery model where a software vendor develops, operates and manages a web application that makes available to its customers the Internet.

3.5 Conclusions

Companies have always invested on both channels and platforms, but it has become more and more a growing awareness of the enormous potential arising from the use of the collaborative platforms of Web 2.0.

The adoption of new technologies inevitably involves risks and opportunities. While the use of collaborative tools within an enterprise promotes and enhances knowledge sharing, on the other hand it increases the concern of business managers about the loss of control over sensitive information to the company, caused by 'open network outside the organizations'. To face with these risks is a clear need to adopt appropriate safety systems concurrently with the adoption of the tools that characterize the Enterprise 2.0.

It is evident that the Enterprise 2.0 identifies the centrality of the role played by the users of an organization and creativity derived from their collaboration opportunities for innovation and competitiveness of an enterprise. The reasons of this approach come from a variety of social factors introduced by social technologies, but above all by the emergence of the knowledge economy (economics knowledge) and the knowledge-based economy (knowledge economy). The first identifies a new discipline of economic theory, which deals with the knowledge as an economic good and its effects on both the individual and well-being of the collective, the second identifies the new historical period pervaded by the use of social platforms. In this context, the ease with which you can collect information and compare the multitude of products and services available on the market allows the customer to make his own choices in a more conscious, quick and focused, leading businesses in a state of hyper competition, characterized by a high rate of innovation and a very short life cycle of the product. The knowledge economy and knowledge-based economy will have encouraged the use of the typical tools of Web 2.0 but also highlighted the need to manage different kinds of knowledge: tacit, explicit, individual or collective. This need expresses the impact that Enterprise 2.0 has been with part of the Knowledge Management (KM). It is only through interaction among the users that you can create knowledge continuously and profitable, in order to create the so-called collective intelligence [31].

It is necessary to permit such sharing through a promotion of all the practices and tools that can lead to an acceleration of the distribution of information, creating communities where information sharing is consistently implemented.

The limitations of Enterprise 2.0 refer to those that are the inherent limitations of Web 2.0. In particular, it collides with the consideration, apparently trivial, for what a system which understands and is able to do is much lower than the user's knowledge. The limit system permeated by Web 2.0

features is due to the difficulty in understanding natural language and to efficiently manage multimedia content. Although the information can be trivially interpreted by a human interpreter, very often a classical information system contains a textual search based on the comparison between strings. Even though the contents were described with the help of tags, the system would simply compare them based on their textual form, without having a clue of their real meaning. This is a trivial problem that affects , for example, many tools of documentary research , for which there is no intrinsic difference between the word 'Red' understood as a surname and the same word understood as a plural adjective . If done with the keyword "Rossi" in a search will return all documents related to any Mr. Smith and all those that contain the description of objects with that color. It is evident that a user can scroll through the results and understand trivially different contexts, but it is clear how difficult it is in parallel to an information system of the Enterprise 2.0 get to that distinction. If you think to the heterogeneity and the abundance of data produced by various collaborative tools, the situation is clearly unmanageable without the technologies that attempt to organize the information so as to allow the various systems to create and share knowledge so as much as possible autonomous and relevant reasoning implemented by a user.

4. Semantic Technologies

Given the need to extract information from the content written by the workers through collaborative tools of the social network business, there are striking structural limitations related to how the data are managed by these tools. The limits of Web 2.0 lies, in fact, in its nature "text" that prevents to implement a form of automated reasoning on the basis of the content expressed. On this basis, introduces the Semantic Web formalisms for representing knowledge in such a way that the information is processed by machines. In particular facts revealed some new requirements to be fulfilled:

- **Identifiability:** each entity in the Web must be uniquely identifiable. This is certainly the basis from which to make the information understandable to machines. To understand how crucial is this feature, just think of what is happening in any information system with the uniqueness constraint of the primary keys.
- **Relations:** reporting an entity to another produces a substantial amount of knowledge and allows to use a variety of known algorithms on data structures arising from these connections.
- **Extensibility:** Due to the enormous variety of related entities on the Web, there is a need for a representative scheme that is easily extensible and adaptable.
- **Definition:** Each entity must be defined on the basis of a shared reference vocabulary (ontology).

These requirements are addressed primarily with the goal of dating resources and then with the combination of ontologies. For this reason, in the context of the Semantic Web, arose numerous initiatives aimed at standardization of languages and models for ontologies. With ontology refers to the collection and sharing of unique terms which, in a particular application domain, a particular coding knowledge. Ontologies, describing the knowledge in one or more domains are used as instruments of knowledge and reuse of integration between the different sources, strictly in accordance with the instance of unity that they bring with them (the computer and the philosophical meaning of the term ontology are not in this respect, far away).

The definition that we believe more technical and concise about what exactly is meant by "ontology" in information given to us by Gruber in 1993: "An ontology is a formal explicit specification of a shared conceptualization" where we find the essentiality of the following terms:

- "**explicit**" by which is meant that the concepts and constraints defined in an ontology should be explicitly defined ;
- "**formal**" which refers to the fact that the ontology is written in a formal manner to enable the understanding of software agents ;
- "**shared**" that reflects the notion that an ontology arises from the common consensus of a group and not an individual ;
- "**conceptualization**" to conceive an ontology as an important linguistic abstraction fundamental to the understanding of the universal language.

The ontology thus understood as a shared conceptualization of a domain , which contains the set of concepts (entities, attributes , processes), the definitions and relationships between concepts can be viewed formally as a pair of sets $\langle V, A \rangle$, where :

- V is the set of terms , and vocabulary ;
- A is the set of constraints and relationships on terms defined in V and enables inferences.

The term ontology has been used to describe artifacts with different degrees of structuring, ranging from simple taxonomies (such as the hierarchy of Yahoo!) to metadata schemas (such as the Public Core), up to logical theories. The Semantic Web, in particular, makes use of ontologies with a significant degree of structuring. The main issues that are called upon to solve are the description of the following concepts:

- Classes (general concepts) in the various domains of interest;

- The relationship that exist between the classes ;
- The properties (or attributes) that classes can have.

Generally, ontologies are expressed in a logical language, so you can make distinctions detailed, accurate, consistent, clear and meaningful between classes, properties and relationships. The creation of an ontology is a very demanding process that requires a thorough knowledge of the context information to describe and mainly concerns the area of Knowledge Engineering. Currently many tools attempt to facilitate the task of creating and some modeling tools have been introduced to improve the design phase and to facilitate the reuse of information.

In addition, many experiments are turning to the activity of Ontology Discovery, which attempt to automate the process of creating ontologies taking the road of clustering on the information retrieved from the text or introducing metrics of similarity between terms for the extraction of semantic relations from written information in natural language.

However, many issues remain to be resolved as regards the reuse and integration of knowledge described by ontology and many languages propose several alternatives for the formalization of the concepts, using different theoretical approaches including those described in 4.2, each with its own advantages and disadvantages.

4.1 Semantic Web

The implementation of an adequate system of knowledge representation in a context such as that of the Web requires significant changes to the assumptions made in previous technologies proposed so far from the research of RC . The scenario of the Web and impose its large size:

- knowledge bases of considerable size, much larger than hitherto represented;
- a rapid evolution of information difficult to control ;
- lack of referential integrity which results in broken links i.e. portions of knowledge no longer accessible;
- distributed authority that involves the presence of information is not always reliable as opposed to what happens to traditional knowledge bases ;
- approximate reasoning methodologies that replace the inference valid and complete .

However, the obstacles faced by the Semantic Web are not only relative to the creation of languages for knowledge representation but also to provide support for the maintenance and the creation of semantic annotations, and thus enable the development of applications that take advantage of the new availability. Therefore the realization of the Semantic Web today is moving on three fronts:

- development of languages based on meta - information machine-understandable to explicitly express the semantics of a Web resource;
- development of tools and new architectures that use these languages and terminologies to provide research support , maintenance, presentation and access to sources of information;
- the creation of applications that offer innovative services to users of the Semantic Web

As part of these goals many research community have shown strong interest in stepping up activities in the study of overhead: Information Retrieval, Web Mining, Knowledge Discovery Ontology, Alignement, etc. It is important to note that the Semantic Web is not only interest of researchers. The adoption of this new resource includes the transition to a completely different paradigm that sees the setting of the Web as a mere repository of text and images for the benefit of creating a resource supplier of complex and reliable services. This attracts the interest of the industry in this new technology sees a major source of income.

The architecture of the Semantic Web is represented by the so-called Semantic Web Wedding Cake (or Layer Cake) by Tim Berners-Lee at the XML 2000 conference. The description generated from the outset a large interest from the community who work for the Semantic Web evidenced by the fact that the image in Figure 6 has now become an icon for the issues related to the Semantic Web is often used by other authors to portray the guidance of the same infrastructure.

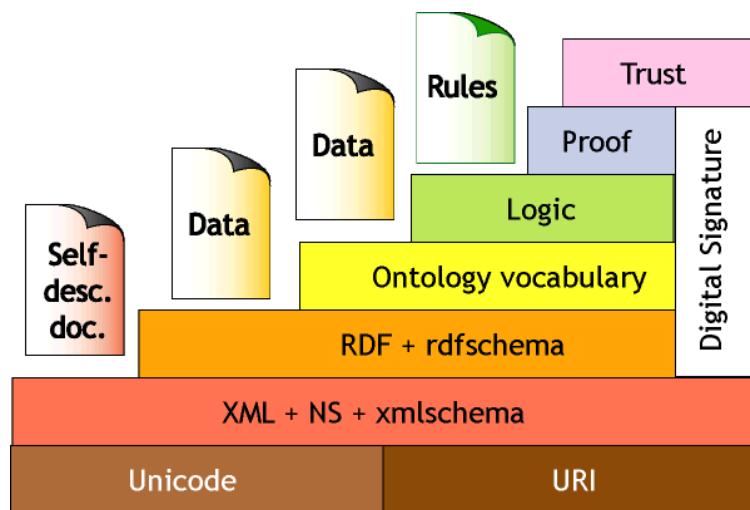


Figure 6 - Semantic Web Wedding Cake

There are two guidelines that govern the evolution of the Web toward its future form increasingly clear that the separation of the information content from its presentation and the passage from the interaction man - machine interaction Web - machine.

According to Figure 6, the Semantic Web can only be achieved through a multi-tiered layering , each of which will feature its own language which will have the task to extend and complement the services provided by the layer immediately below , showing the upper level new features. The Semantic Wedding Cake presented in figure puts particular emphasis on protocols and languages designated for each specific layer of the architecture of the Semantic Web.

At the base of the architecture we find the standard already widely recognized and popular, such as Unicode is a standard for representing characters in 16 -bit universally adopted as the direct successor dell'ASCII, and the URI (Universal Resource Identifier) , which are the irreplaceable foundation of the Web among the factors for its success. With the URIs the Web was able to identify the resources accessible through its own protocol , HTTP, and with all other existing protocols (FTP, Telnet, Gopher, etc.). The main point to which other systems had not arrived was a universal syntax, protocol-independent , and easily memorized or exchangeable for , with which to identify network resources .

Although the image is sufficiently explanatory of objectives and layers of abstraction where the infrastructure of the Semantic Web will face , however, is a fairly informal species program for the arrangement of the upper layers of the architecture : Ontology terms and definitions are alleged in a given language , the inferential logic relates to the processing of information, the Proof is derived from the need to keep track of logical deductions , while the Trust refers to the need to check the reliability of the resources examined .

All of these elements are not necessarily related in terms of which depict semantic (i.e., in a necessarily overlap). There can be no trust logical deduction, deduction without an ontology, and so on. The following describes in more detail the levels of most interest to this project.

4.1.1 XML and RDF levels

XML (eXstensible Markup Language) arises from SGML (Standard Generalized Markup Language), the international standard for the description of the structure and content of electronic documents of any type, but not with respect to this complex has its own features and benefits of ownership suitable for the Web is a meta-language that allows the definition of customized markup language and then the realization of domain - specific terminology. Among the design goals of this language we state the fundamental ones:

- efficiency and adaptability to a distributed environment ;
- ease of use ;
- predisposition to the development by other application ;
- legibility of the documents;
- flexibility.

XML, along with a number of technologies related to it has benefited from a substantial industry support, and has resulted in a profound revolution in the software world (not just in the Web). An XML document is a simple text file that is platform-independent. The main difference is that compared to HTML tag names used in it are not predetermined but can be specified arbitrarily.

In order to facilitate the exchange of XML data between different applications and avoid possible problems of name collision, in a context in which the interpretation of the data is not unique, we introduced the concept of namespaces (namespace) . The idea is to declare the use of a namespace by specifying a unique name through the mechanism of URI (Universal Resource Identifier).

The most important innovation introduced by this language is the separation of the description of the data from the visual formatting thus achieving independence from the browser (functionality ensured by XSLT). XML is a first step towards structuring the informational content of the resource, however, still far from the idea of a semantic structuring. The main advantages of XML are:

- broad applicability;
- mainly used as an exchange format on the Web;
- distinguish the information content from its presentation, etc..

However, XML is ineffective due to various shortcomings inherent in its definition:

- semantics implied, since it does not formally defined;
- ambiguity of interpretation of information (see example in Table 3);
- the need to share the interpretation of messages exchanged off-line or otherwise in the design phase ;
- is a technology primarily aimed to support syntactic interoperability
- lends itself to the representation of knowledge in adherence to a model exclusively classification Table 3:

<pre><author> <uri>page</uri> <name>Daniele</name> </author></pre>	<p>But also in this way:</p> <pre><document href="page"> <author> Daniele </author> </document></pre>	<p>or this way:</p> <pre><document> <details> <uri>page</uri> <author> <name>Daniele</name> </author> </details> </document></pre>
--	---	--

Table 3 -- Example of Ambiguity representation

In XML you can provide the specification of the data structure through the use of other languages such as DTD (Document Type Definition) and XSD (XML Schema Definition). In particular, XSD has been specially developed by the W3C to allow you to describe the structure of XML documents,

more expressive than DTD and in a manner best suited to the needs of developers, its particularity is to be described, in turn, through XML syntax.

It is important to note that to a same XSD document, in general, relate more XML documents: the latter are the real containers of data and are called instance documents, in contrast with the document schema that is a pure descriptor of their structure. Drawing a parallel with the Object Oriented Programming, we might say that a class describes an object such as a schema describes a document instance.

Based on XML was designed RDF (Resource Description Framework), a fundamental model to describe the associative information between distributed resources in the network. These associations are expressed in machine-understandable manner and coded into a set of triples, where each triple is the same Subject, Verb, Object Complement of an elementary sentence : it is said that particular things (e.g., *a book*) have properties (e.g. "*have as author of*") with certain values (eg, a certain person).

The mechanism of the Universal Resource Identifier (URI) is heavily used for the identification of resources to which the assertions. The assertions defined in RDF are serializable in a variety of syntactic forms, the most convenient of which is adopted and XML. RDF Schema (or RDFS) extends RDF and provides a mechanism to describe groups relate resources and relationships to which they are subject. The role of RDFS is comparable to that DTD or XML Schemas play more appropriately for XML. More details on RDF and RDFS will be provided in section 4.2.1.

4.1.2 The level of ontology languages

An ontology language is the tool that allows you to define an ontology as a starting point and has an RDF Schema which adds a richer set of terms to complete the level of abstraction and solve certain linguistic ambiguities. The basic requirements that must possess an ontology language are the following:

- must be particularly intuitive, the current trend is to use the paradigms of type frame-based or object-oriented who have recently found much success;
- must have a well-defined formal semantics with established reasoning properties in terms of completeness , correctness and efficiency;
- It must be easily connected to the lower levels of the architecture of the technologies of the Semantic Web to ensure interoperability with RDFS, RDF and XML.

The creation of an ontology language also needs to address several issues related to the environment in which it is introduced, the Web is necessary to ensure interoperability in a distributed environment, manage the polysemy and synonymy, supporting the development of ontologies, implement the operator equivalence between resources referenced by different names, manage scalability, allowing the integration of different ontologies, simplify automated reasoning , and so on.

Many ontology languages have been proposed: Cycl [32], KIF³ (Knowledge Interchange Format Genesereth), Ontolingua⁴, Simple HTML Ontology Extensions⁵. The first to fully comply with all the requirements was OIL (Ontology Inference Language) subsequently integrated with DAML-ONT funded by the Defense Advanced Research Projects Agency⁶ (DARPA) giving rise to the most well-known : DAML + OIL⁷, which fills in the gaps of the two resulting the instrument far more suitable.

However, recently a new language has been proposed by the W3C OWL⁸ (Web Ontology Language) is particularly important because of its simplicity and compatibility with RDF and RDFS. More details on OWL and its successor OWL2 are provided in section 4.2.2. We can easily conclude that what is missing today are not the tools of RC but the support is still poor relative to the upper layers of the architecture , among other things, are the basic tools for the usability of the semantics provided in the Semantic web

4.1.3 Subsequent levels

The levels of logic, of Demonstrations (Proof) and Trust (the Trust) are still in a very preliminary stage, and the debate on how to be concretely realized is still wide open. With regard to the level of logic, if one side is certainly positive to have systems that recognize the basic concepts of subclass, inverse property, etc. But it would be even more profitable you can instruct the computer on any logical principle, allowing it to reason, by inference, using these principles.

Recall that with "*inference*" means a deductive procedure by which, starting from one or more premises, is obtained, by logic a conclusion. In order for the Semantic Web can become sufficiently expressive to help in a wide range of situations, self- extracting useful information from the vast amount of semantically annotated web documents , it is essential to build a powerful logical language to make inferences , to the level of ontologies, in fact, there is no inference, but only knowledge representation .

The layer of logic will make what was a declarative language with limited expressive language in a "*Turing-complete*" with inferences and functions, enabling applications to connect together different RDF. Given that the number of logical rules to find on the Web could also be huge (thousands, if not millions of links to cross before reaching the conclusions), the inference engine must necessarily be based on heuristics.

To add logic to the Web is a problem with significant complications, math and technology, arising from the desire to find an acceptable compromise on the expressive power, trying to avoid as much as possible that this could lead to paradoxes. To this is added the fact that the deductive systems, able to make deductions from the layer of logic, may be multiple and not necessarily interoperable.

³ <http://www.ksl.stanford.edu/knowledge-sharing/kif/>

⁴ <http://www.ksl.stanford.edu/software/ontolingua/>

⁵ <http://www.cs.umd.edu/projects/plus/SHOE/>

⁶ <http://www.daml.org/>

⁷ <http://www.w3.org/TR/daml+oil-reference>

⁸ <http://www.w3.org/TR/owl-features/>

Instead of designing a single comprehensive system to support reasoning, the approach proposed by the Semantic Web is to define a language for representing the demonstrations. The systems could then export to other systems deductions made which, in turn, could be used for further deductions. A demonstration (Proof) is a sequence of formulas, each of which is derived by applying the rules of inference, axioms definitions, or formulas to it earlier in the sequence.

A *Proof Language* (language of communication aimed at demonstrations) will allow software agents to exchange assertions together with the chain of inference through which they have been obtained by deduction from other statements deemed acceptable by the receiving agent.

This will allow the creation of generic validation engine, which act as a nucleus for more specific applications (e.g., Applications for the control of access rights).

Even when there will be a Proof Language, given the total freedom of expression of the Semantic Web, there will still be a big problem to overcome: Who can trust the answers of a system that allows us to say anything and everything? An answer to this question goes to a wider dissemination of so-called digital signature (digital signature).

The **Digital Signature** is of significant importance in different layers in abstract model of the Semantic Web. The public key cryptography is a technique known for some years, but has not yet spread widely as might have been expected. In the vision of Berners -Lee, an element that could have played against the spread of this technique is that it is a " coarse-grained ", imposing a binary choice between trust or not trust (trusted / not trusted) , and would require an infrastructure in which the parts can be recognized and accepted as credible only in specific domains.

With a finer granularity such as this, the digital signature could be used to establish the provenance of ontologies and deductions, as well as data. Agents should therefore be skeptical about the claims on the Semantic Web, tracking until you have verified the reliability of the source.

In other words, the RDF statements must be marked with the digital signature of the person or organization that produced them (or at least claims to agree with them): digital signature, using cryptographic techniques, guarantees the authenticity of various assertions and allows you to discover their origin.

It will then instruct the user to the software in your computer such as digital signatures to trust or not. However, the risk you take by proceeding in this way is to cut out a slice too big of available knowledge considering it unreliable because only asserted by persons / organizations not directly known by the user. Just to address the issue of how to increase the number of entities to be trusted for a given user that has been proposed the idea of the Web of Trust considers that the trust as transitive, so that if user A user declares trust B, which in turn claims to trust the user C, we get that A also trusts C, creating a Network of Trust that, structuring a graph, binds together the various users. In order to better modulate the level of trust, it could then calibrate weighing it with respect to some parameter, such as the topological distance in the graph of trust.

Superimposed on this, you could then also create a similar network of Mistrust: Note that in fact have explicit directions unreliability of a source can be very useful as our agent software, meeting in

navigating the Web a document which does not have an explicit proof of confidence or lack of confidence, it would be led by default to consider more reliable than a document for which there is instead an explicit proof of distrust .

Will therefore be multiple factors that the computer will have to take into account when allocating the degree of reliability relative to a certain piece of information: the user must also be placed in a position to choose whether to make this process transparent (being informed of all the factors involved in the decision) or opaque (e.g., setting at will, and once and for all, an appropriate parameter to adjust the level of the application in charge of filtering the information obtained on the web).

4.2 Defining Languages for Ontologies

To this day, there is a large number of formalisms for the representation of ontologies and their implementation to conform to the dictates of the Semantic Web. In the next sub-sections, we will illustrate the currently most widely used languages in order to structure the information and support interoperability within and between different applications.

First we will analyze in more detail the RDF language (already introduced in 4.2.1) which is a fundamental part of the architecture of the Semantic Web You will then pass to the description of OWL and OWL2, two different languages for the definition of ontologies defined by W3C. It will conclude with SKOS, a family of formal languages that provides a simplified model to describe conceptual schemes.

4.2.1 RDF and RDFS

RDF (Resource Description Framework) is a standard defined by the W3C⁹ for describing the semantics of an XML document and, in this sense, it is placed at a level of abstraction higher than the latter. RDF is in a sense an application of XML: if XML is an extension of the document, RDF can be seen as an extension of the data introduced by the XML. The basic model of RDF data consists of the following elements.

- **Resources:** This term refers to anything that can be described. A resource can be a Web page, a portion of this or a collection of pages (e.g. an entire Web site). In addition, a resource can be understood as a material object is not directly accessible via the Web or as an abstract entity (the property "*have the author*"). In other words, a resource is any entity reporting the object of the RDF statement. The mechanism of the URI (Universal Resource Identifier) is used to uniquely identify each resource, so that even its information content.

⁹ <http://www.w3.org/RDF/>

- **Properties:** A property is a feature, a report describing a resource. The meaning, the set of values it can take, the types of resources that can refer are all available information from the RDF schema in which it is defined.
- **Assertions:** An assertion is a sentence with a fixed structure, comprising: a subject (the described resource), a predicate (the property) and an object (the value assigned to the property), where the object can be a simple string or another assertion. An assertion is then a Triple of the form (subject, predicate and object). In RDF, the three components of the triple are called, respectively, subject, predicate, object. The first two components are mandatory resource, while the third (i.e., the value of the property) can be both a resource that a literal.

The interesting thing about this way of representing knowledge is that the object O can in turn be seen as the subject of a new triple S. This technique allows you to represent situations also very articulate. For example, the phrase "Paul owns a bicycle black" can be decomposed into two statements: "*Paul has a certain bike*", i.e. (*Paul owns a bike*), and "*That bike was black*" or (*bike, hasColor, black*).

This representation also has the great advantage of being incremental in the sense that you can easily add new knowledge without being forced to change the triple constructed previously.

A further example of using the RDF model can be provided by the following statement: "*Daniele's phone_ number is 2322*", which can be summarized graphically as in Figure 14. The previous statement is formalized in RDF / XML as shown in Figure 7. Syntactically, the concepts expressed by RDF are then serialized using XML.

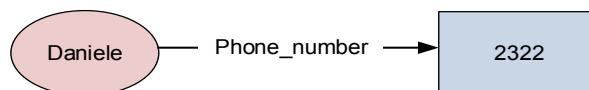


Figure 7 - A simple document RDF / XML

The mechanism for describing RDF resources is neutral (general purpose) with respect to the application domains, that is not addressed to a particular application context, let alone, a priori defines the vocabulary and semantics of a certain area.

The essential purpose is just to have a tool adaptable to the description of information about a any context. In RDF, however, there are no levels of abstraction: there are resources and their relationships, all organized in a graph dish. In other words you cannot define types (or classes) of resources with their specific properties overcome this limitation, RDF has been enriched, by means of **RDF Schema**¹⁰, with a simple type system (reminiscent of the type systems of object-oriented programming languages).

¹⁰ <http://www.w3.org/TR/rdf-schema/>

A resource may, for example, be defined as an instance of a class (or multiple classes) and classes can be organized in a hierarchical manner, allowing you to derive, by inheritance, new knowledge.

Another shortcoming of RDF (RDF Schema which is intended to remedy) is to provide no mechanism for declaring the property, in order to define the constraints of applicability, to organize them hierarchically. Through an RDF Schema is possible to define all the terms that will be used in RDF statements by assigning them a specific meaning, as in a sort of vocabulary; these vocabularies, freely created by individual users (or user community), reside in the necessary documents, available directly on the semantic Web and which can be accessed by software agents which human beings.

RDF Schema uses the RDF model in order to define the type system of RDF resources and providing a set of default properties that can be used to define classes and properties at the user level. You can also define constraints domain and codomain (range) on the properties and some types of reports (including those of the subclass of a resource and subtype of a property). The combination of these elements is called the RDF Schema vocabulary.

The specification language is a declarative language RDFS little expressive but very simple to implement. It is expressed through the syntax of RDF serialization / XML, and uses the XML namespace mechanism for the formulation of URI that uniquely identifies the resources (as defined in the schema itself) allowing the reuse of terms defined in other schemas. Concepts and properties already declared for a domain can be used again or set out to meet the needs of a particular community of users.

Figure 8 shows the relationship between a set of RDF statements and the patterns that they use. The objective of the use of RDF(S) is to make available to the user a set of terms as possible to achieve exhaustive descriptions most disparate, with the intention that the vocabularies provided clues are influential in determining the content Treaty from single RDF descriptions. The role that plays in the architecture of RDFS Semantic Web is the transition between those languages related to specific descriptions of resources and those aimed at the representation of concepts in a more abstract way: the ontology languages . In fact, although RDFS offers greater expressive power than RDF and XML, which introduces elements are not sufficient to complete the knowledge representation for the Semantic Web, other extensions are therefore necessary.

The ontology languages are therefore an extension of RDFS and complete its instrumentation for the description of knowledge, introducing new formalisms meaningful and enriching semantics with a complete collection of axioms of interpretation. The main difference between these two instruments is that, while RDFS semantics provides a mostly text, the ontology languages define a clear formal semantics essential to make content machine-processable.

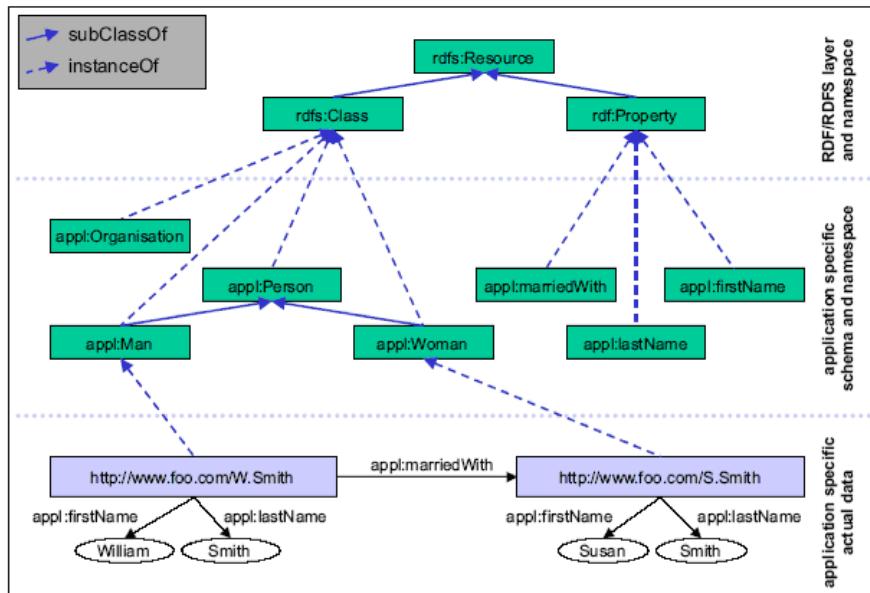


Figure 8 - Example of the relationship between level information RDF and those at RDFS

4.2.2 OWL and OWL 2

OWL (Web Ontology Language)¹¹ is placed in the architectural scheme of the Semantic Web layer of Ontologies ("Ontology Vocabulary") and relies on the syntax and basic ontological primitives provided by RDF / RDF Schema, which occupy the lower layer . While RDF is intended for integration and association of data and distributed resources, OWL aims to enable potential deductive reasoning on distributed data. OWL introduces more expressive potential compared to RDF , provides independent views of the data allowing an easier development and building of knowledge bases complex and is equipped with a machine-understandable semantics.

OWL is the result of considerable effort in the development of a set of logical constructs computational and flexible as it is the most recent of a series of languages with similar objectives. In particular, is the direct successor of DAML + OIL: the integration of DARPA Markup Language¹² and Ontology Inference Layer¹³ (or Save) 0, and finds some of its roots in the language SHOE¹⁴ (Simple HTML Ontology Extensions) created to incorporate descriptions machine- readable hypertext Web. OWL contains a family consisting of three languages OWL Lite, OWL DL, and OWL Full, respectively, each of which has a greater potential inferential and expressive. In particular:

- **OWL Lite** is designed for those users primarily needing a classification hierarchy and do not feel the need to express complex constraints. For example , OWL Lite allows you to express cardinality constraints , but the values are only 0 or 1

¹¹ <http://www.w3.org/TR/owl-features/>

¹² <http://www.daml.org/>

¹³ <http://www.ontoknowledge.org/oil/>

¹⁴ <http://www.cs.umd.edu/projects/plus/SHOE/>

- **OWL DL** (whose name is due to its correspondence with Description Logics) is aimed at users who want to have the maximum expressiveness possible while maintaining the computational completeness (i.e., having the assurance that the system will extract all the conclusions of inference) and decidability (i.e., having the guarantee that all processes will finish in finite time). OWL DL includes all language constructs of OWL Full, but with the limitation that they can only be used under some restrictions, for example, such a class is allowed to be subclass of several classes, but not to be an instance of a ' other class.
- **OWL Full** is designed for those users who require maximum expressiveness and the syntactic freedom of RDF whole , without any guarantee computational , for example, in OWL Full, a class is allowed to be treated simultaneously as a collection of individuals, both as an individual in its own right. OWL Full allows an ontology to augment the meaning of the vocabulary (RDF or OWL) default. To date, there is no automatic reasoner can fully support all the features of OWL Full.

Each language is thus an extension of its simpler predecessor, both in what can be expressed in a formal way, in terms of the validity of the conclusions that can be deducted. Developers adopting OWL ontologies should consider which sublanguage best suits their needs. The choice between OWL Lite and OWL DL depends on the level required by users of the most expressive constructs that are provided from OWL DL. The same goes for the choice between OWL DL and OWL Full : When using the OWL Full OWL DL respect , support the reasoning is less predictable given that a full implementation , as already mentioned, does not yet exist.

OWL 2¹⁵ is an ontology language for the Semantic Web arising from OWL. The language provides classes, properties and individuals, such as storing all the documents for the Semantic Web ontologies written in OWL 2 can be used together in RDF documents because they themselves are exchanged as RDF documents. In addition, OWL 2 supports data types defined in XML Schema Definition Language. Like its predecessor OWL, OWL 2 also has three sub- languages :

- **OWL 2 EL** is particularly indicated for the management of ontologies that contain a large number of properties and / or classes. It has been demonstrated that dedicated algorithms of reasoning for this profile may be developed in a fully scalable. The EL acronym reflects the profile of the bases in the EL family of description logics [EL ++], which provides the only existential quantification
- **OWL 2 QL** is aimed at applications that use large volumes of data and where the reasoning is based on a query mechanism. You can make queries combined using conventional relational DB. The expressiveness of the language is limited even if it includes the main features of conceptual modeling such as UML diagrams and ER diagrams. The QL acronym reflects the ability to implement a relational query language standard.

¹⁵ <http://clarkparsia.com/http://www.w3.org/TR/owl2-overview/>

- **OWL 2 RL** flagship applications that require scalability without sacrificing the expressiveness of the language. The systems of reasoning can be implemented based on the definition of rules, hence the acronym RL (Rule Language).

The differences between the languages OWL and OWL 2, with its sub-languages, in terms of levels of expression are shown in Figure 9. In particular, note how the original level in OWL2 DL has been further detailed modeling to improve the expressiveness of ontologies.

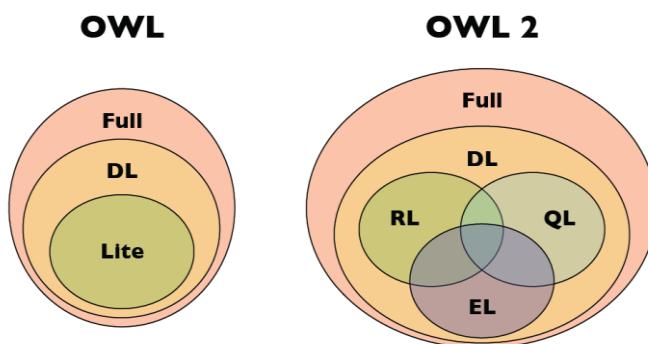


Figure 9 - Comparison between OWL and OWL2

4.2.3 SKOS

Simple Knowledge Organization System (SKOS)¹⁶ area of work developing specifications and standards to support the use of knowledge organization systems (KOS) in the context of the Semantic Web. Specifically, it is a family of formal languages that provides a model for expressing the basic structure and the content of conceptual schemes such as glossaries, classifications, taxonomies and any type of structured vocabulary.

It represents a semantic model (defined by RDF and RDFS) developed initially by July 2003 (like open source project) within the project SWAD-E (Semantic Web Advanced Development for Europe)¹⁷ that aimed to define a model for thesaurus compatible with the ISO most important (ISO 2788, ISO 5964). Since September 2004, the responsibility for the subsequent development is passed to the W3C that supports the evolution of the relevant Working Draft.

The conceptual basis of SKOS, SKOS Core Vocabulary, was designed with the idea of being easily extensible. In particular, the SKOS Core Vocabulary plays the role of a connector between the more traditional systems of knowledge organization (KO) used in libraries, museums, archives, such as thesauri and classification systems and new structures designed for the Web such as open-directories.

SKOS Core Vocabulary consists of a series of RDFS classes and properties of RDF that are used to represent the content and the basic structure of so-called concept-scheme. The possibility of being extended to other RDF vocabularies such as DCMI Metadata Terms¹⁸ and FOAF (Friend of a

¹⁶ <http://www.w3.org/2004/02/skos/>

¹⁷ <http://www.w3.org/2001/sw/Europe/>

¹⁸ <http://dublincore.org/documents/dcmi-terms/>

Friend)¹⁹ greatly increases the potential in that, in those cases in which the instruments of SKOS Core not comply with the requirements or be sufficient, it is possible define new classes.

For this reason SKOS Core Vocabulary offers a natural flexibility that provides a basis for interoperability even in situations in which the concept- scheme have been developed using different specifications. Another feature that inherits from RDF SKOS Core Vocabulary is the mechanism provided by the sub -classes and sub -property of RDFS. To support this type of extension properties of SKOS Core Vocabulary is grouped into the following families:

- property for the lexical labels;
- property for the labels of documentation;
- properties for semantic relations;
- properties for symbolic labels.

The properties within these families are organized in a hierarchical manner so that it is possible to extend the most appropriate for their needs. So SKOS can be used both for transfer of knowledge organization systems existing in the world of the Web to build from scratch simple conceptual schemas for the Web in order to support research , classification, modeling taxonomies specific to a certain domain .

SKOS Vocabulary considers "*Concept*" the basic unit of any concept scheme. The class Concept allows to model a particular resource in order to express this as a concept. A concept can be thought of as something that can be defined or described. Each concept can have only one description or rather a single "*preferred term*" and can have unlimited alternative descriptions or "*alternative tags*". The easiest way to define a concept in RDF is assign it a URI and use the form *skos:Concept*. We have seen in the previous sections how to express the statements using RDF graphs. Figure 10 shows another example made with SKOS. The serialization of this example is shown in Figure 11.

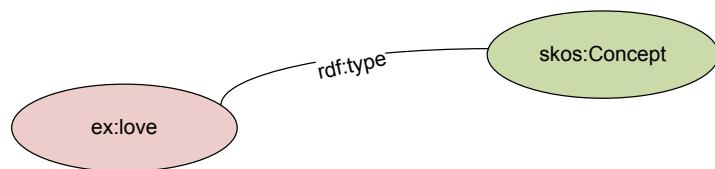


Figure 10 - Example of serialization concept of love

```

- <rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://www.example.com/eg#"
  <skos:Concept rdf:about="http://www.example.com/concepts#love"/>
</rdf:RDF>
  
```

Figure 11 - Serialization concept of love

¹⁹ <http://www.foaf-project.org/>

We note that in this example there are two namespace (URI) referenced by the prefixes '*rdf*' and '*ex*' that provide a method to identify unambiguously the semantics and conventions that govern the use of property by identifying the authority that manages vocabulary. In addition, in this example, the triple love *rdf:type Concept* indicates that love is an instance of concept that the resource itself , in this case love is a concept.

In SKOS labels are used to denote resources using the common language. There are different types of labels:

- *skos:prefLabel* , which is the preferred term for a certain concept and other synonyms;
- *skos:altLabel*, which is the most commonly used term for a certain concept and other synonyms;
- *skos:hiddenLabel*: a label "hidden" that is not normally visible but allows access to the content to applications that are text-based searches ; Normally this tag is used to include syntactically incorrect variants of the same term : e.g. eletricity, electricity etc..
- *skos:prefSymbol*: a label that lets you include a symbol or an image of a certain resource;
- *skos:altSymbol*: a symbol or an alternative image for a resource.

Also, there are 7 properties that can be used to add information to the description of a concept. These properties have as superclass *skos:note* : it can be used to provide a description of any type useful for any purpose . The subclasses are instead:

- *skos:definition*: useful to give a complete and accurate explanation of the resource. For example, " fruit of the plant family Bromeliaceae " ;
- *skos:scopeNote*: used to restrict or increase the description of the concept. E.g.: " Microwave Frequencies: 1 GHz to 300 GHz"
- *skos:example* : used to provide examples of the use of the term or concept
- *skos:historyNote*: used to indicate significant changes of meaning . Ex: "*Pears was incorporated as more specific concept of vegetables rather than fruit*"
- *skos:editorialNote*: used to provide administrative information . Ex: "*Contact Dr.Miles for a more complete definition of the concept* "
- *skos:changeNote*: used to keep track of the various changes that occurred . Ex: "*Moved by the class pear fruit to vegetables. Once the 15/12/2004 by Dr.Miles .*"

4.3 General Purpose Ontologies of Interest

The Semantic Web is a "*Web of data*". The combined technologies of the Semantic Web (e.g., RDF, OWL, SKOS, SPARQL, etc.). Provide an environment in which applications can query the data, draw conclusions using vocabularies, etc. However, to make the "*Web of data*" a reality, it is important to have available, in a standard format, the huge amount of data on the Web as well as the relationships

between the data itself (as opposed to a mere collection of data). This collection of interconnected data on the Web can be referred to as Linked Data²⁰.

Just as hypertext , the Web of data is constructed with documents on the Web , however , unlike the conventional Web hypertext , where links are anchors in hypertext documents written in HTML, for Web data links between arbitrary objects are described by a common RDF format , to allow conversion or on-the-fly access to existing databases (e.g., relational, XML, HTML, etc.). Tim Berners -Lee has outlined a set of ' rules ' for the publication of data on the Web so that all published data from becoming part of a single global data space:

- Use URIs to identify things;
- Use HTTP URIs so that these things can be consulted and improved by people and user agents ;
- the need to use standards such as RDF and SPARQL to explore a URI and retrieve information ;
- Include links to other URIs to improve discovery of other related information on the Web

These have become known as the "*Linked Data principles*", and provide the basic recipe for the publication and access data using the infrastructure of the Web The adoption of these principles has resulted in the creation of a global space parallel to what collects the documents, in which data from different knowledge domains are connected between them, the so-called " Web of data". However, in a presentation held in 2009, the same Berners -Lee revisits and summarizes the principles of Linked Data in three basic rules:

- each resource due to a real-world object is characterized by an identifier that begins with HTTP ;
- the main purpose for which they are used the Linked Data is to get information. For these reasons , the data must be expressed in a standard format so that they can be useful and reusable by users ;
- the information to be obtained is not linked , such as weight, height , or date of birth of a person, but concerns the set of relations it has established with all the other elements of the real world . In addition, when these relationships are defined and declared in an explicit way, the resources (or objects) placed between these relations should be given an identifier that begins with HTTP.

Even within the new guidelines is defined as the need to use standard formats for representing objects and their connections, without explicit specific formats, meaning that the elements of openness and sharing of Linked Data is also reflected in the technology.

One of the most important examples of the adoption of the principles of Linked Data is the Linking Open Data project. The main objective of this project is to serve as a catalyst for the Web container and data, identifying the set of data available under open licenses, published in RDF format (respecting the principles) and disseminated through the Web.

²⁰ <http://linkeddata.org/home>

Figure 12 is shown a portion of the diagram of the design of Linked Data, where each node represents a set of data published as linked data, and the arcs indicate the connections between the various data sets. Currently, starting from the statistics collected by the LOD community in the ESW wiki: The Web of Data would be composed of 19 billion RDF triples interlinked.

The design and development of software systems should take into account the semantic oriented paradigm of linked data and, where possible, use some of the many ontologies defined in this context. The following sections describe some of the patterns in the Ontological Semantic Web, Linked Data on the present considered relevant to the project objectives including:

- FOAF, Semantic Web is an application that can be used to describe people, their activities and relationships with other people. In particular, FOAF has been designed to allow you to structure information relating to persons of a social network in a machine-readable.
- SIOC, is a vision for the semantic representation of information produced by the on-line community: the RDF language describes the user-generated content sites such as blogs, forums, wikis and social networks and interactions that have aroused.
- SCOT, ontology which aims at the description of the characteristics of folksonomies. In particular, it aims to describe the structure and semantics of data tagged and their relationships, either explicitly through RDF / OWL, and enable interoperability for social sharing and reuse of metadata semantic tags from different sources.
- MOAT, defines a Semantic Web Model to define the meaning of the tags in a machine-readable way. GeoNames, is a free project for the creation of a database of geographic world. Its purpose is to provide the tools to translate the name of a mountain or a city in which the data represent : latitude, longitude, elevation, population, postal code, etc.

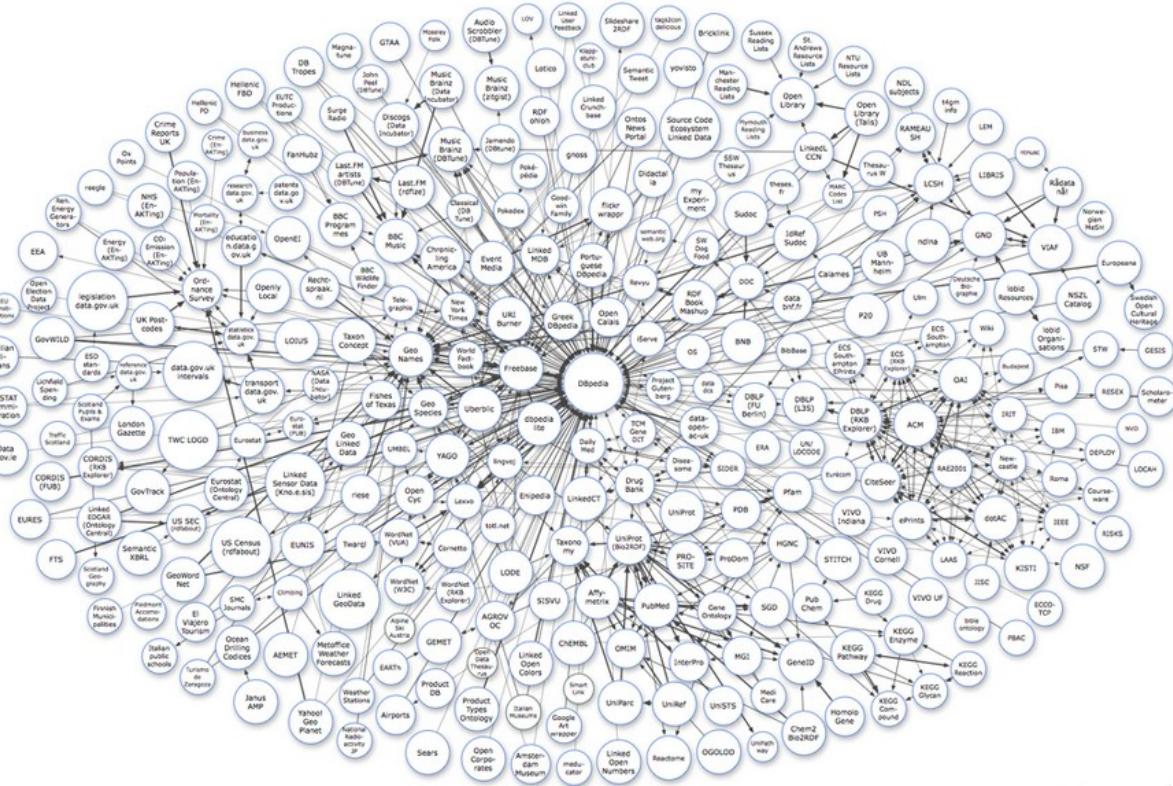


Figure 12 - Linked Data Cloud diagram of the project

4.4 Conclusions

Overcoming the limitations of Enterprise 2.0 in the management of the data produced by the various collaborative tools, it is essential to implement strategies using tools that cover formalisms capable of making machine-understandable the information produced. For this purpose, the technologies underpinning the Semantic Web act as primary and indispensable tool to manage tacit knowledge expressed or implied by the users of social networks and stored in different silos of information. The ability to model semantic information, to conceptualize and relate them is the "cornerstone" of any outsourcing activities that can have good results, especially when referring to areas well defined and restricted, as can be improved a determined business process through best practices.

5. Enhanced Enterprise 2.0

Introducing Enterprise 2.0 tools in the enterprise and implementing a KM system makes generation easier and the sharing of information content.

The develop of the size of the online community makes more difficult the exploitation of these shared resources , such as analysis and ranking of proposals, and support users in selecting the best proposals.

In general, the social tools create problems of knowledge management [33]. In particular, one must deal with the following issues:

- **Fragmentation of information and heterogeneity of the data.** The Enterprise 2.0 tools are often an aggregation of different services not communicating with each other. Employees of different divisions may also prefer different tools, such as focusing on a few blogs and other use of Wiki. This leads to the creation of silos of information, information sets that remain in a specific area of infrastructure software and it is not visible to the entire company. It is possible that data relating to a specific entity of relevance to the company (e.g. a project) are located in different silos and then maybe stored in different formats. Instead, you need to have any time all information generated for a specific topic.
- **Extraction and reuse of knowledge.** The concept of knowledge is closely linked to the possibility of its effective utilization. However, the information arising from the use of collaborative tools , but also those in simple documents , are difficult to find and to be re-used by software agents. This is because it lacks a formal and machine-readable representation of the knowledge produced, which enables to applications an easier interpretation of the contents of the documents.
- **Metadating of resources.** Add the descriptive elements of a resource is a fundamental operation for the catalog created content. The tagging, the most common form of metadating, is commonly used to catalog resources created by different users in order to make them more accessible. In the context of Web 2.0, an important feature is that of the social tagging or sharing tags between multiple users. The collection of these tags leads to the creation of a folksonomy. The limitations of tagging are mainly related to the ambiguity of the tags themselves and the lack of homogeneity. For example, referring to the concept of e-mail, some users might use the tag "*email*", others as "*e-mail*", causing an apparent discrepancy It would be desirable to provide tools to homogenize the metadating.

To overcome all these problems, it was necessary for the introduction of a layer of semantic annotation systems Enterprise 2.0, according to the line drawn in the definition of the Semantic Web, leading to the definition of the architecture known as Enhanced Enterprise 2.0.

The evolution towards the Enhanced Enterprise 2.0 goes along the evolution of progress made on the web. Web 1.0 was an environment that allowed the simple connection of information between the server ,the Web 2.0 (ie, the Social Web) and Enterprise 2.0 are focused instead in connecting users via social applications and the Semantic Web is the result ,as opposed,of an increased connectivity between knowledge . Leveraging both social connections and knowledge connections you get the Ubiquitous Web , which is based on the Enhanced Enterprise 2.0. Enhanced systems Enterprise 2.0 typically have the following components [34]:

- **Manufacturers of semantic data:** semantic tools or add-ons to existing Enterprise 2.0 tools , with the task of converting the Social Data in extensible and interoperable formats, using ontology based on RDF.
- **Models:** Taxonomies and ontological domain-specific patterns that give a structure to the data produced.
- **TripleStore:** repository where all data RDF are stored and where they generate , when it is possible, new information via inference on the data held.
- **Consumers of semantic data:** applications that exploit the data present in the system finding resources, browsing semantics or the generation of semantic mashups, aggregates of resources linked by semantic concepts . The service consumers can make use of the SPARQL query language to obtain information from TripleStore .

The Enhanced Enterprise 2.0 can be introduced into a system where they are already active Enterprise 2.0 tools : just giving all the social tools of connectors that translate the data produced in RDF format. The main Enterprise 2.0 tools (SLATES) become so semantic tools (SemSLATES) [33]

	SLATES	SemSLATES
<i>Search</i>	Plain-text search	Semantic search
<i>Link</i>	Hyperlinks between documents	Relationships between typed objects
<i>Authoring</i>	Documents	Data and metadata
<i>Tags</i>	Tagging	Semantic indexing based on ontologies
<i>Extension</i>	Hyperlinks navigation	Graph-based navigation
<i>Signals</i>	RSS feeds	Semantically-indexed RSS feeds

Figure 13 - From [33], the difference between instruments and SLATES SemSLATES

The semantic tools described by the acronym SemSLATES can be defined as follows:

- **Semantic Search:** searches based on concepts that solve the problems of ambiguity as resources are uniquely identified by a URI.
- **Semantic Links:** You can give semantic meaning to the connections between documents (ex, identifying a link as a relationship - person project). These semantic information may be used for research or suggestion related content.
- **Semantic Authoring:** textual content, blogs, wikis and so on. It can be stored in the semantic and machine-readable format.

- **Semantic Tags:** tags are saved in semantic.
- **Semantic Extension:** The user can not only reach new documents using hyperlinks, but also taking advantage of the knowledge graph of RDF, which allows tracking of documents linked to the concepts semantically related to the concepts of the source document.
- **Semantic Signals:** RSS Feed or posts of blogging can be filtered based on the concepts they contain, in order to avoid problems of information overload.

In Section 5.1 describes the ontological domain-specific patterns that enable SemSLATES, while 5.2 will describe the semantic tools of Enhanced Enterprise 2.0.

5.1 Domain-Specific Schemes

5.1.1 SKOS

The ontology SKOS (Simple Knowledge Organization System) is a family of formal languages designed to represent glossaries, classifications, taxonomies and any type of structured vocabulary, so it is particularly suitable to represent folksonomies.

The conceptual basis of SKOS, SKOS Core Vocabulary, was designed with the idea of being easily extensible. In particular, the SKOS Core Vocabulary plays the role of a connector between the more traditional systems of knowledge organization (KO) used in libraries, museums, archives, such as thesauri, classification systems and new structures designed for the Web such as the open-directories.

SKOS Core Vocabulary consists of a series of RDFS classes and properties that are used to represent the content and the basic structure of so-called concept- scheme. The possibility of being extended to other RDF vocabularies such as DCMI Metadata Terms and FOAF (Friend of a Friend) (Section 5.1.2) greatly increases the potential in that, in the cases in which the instruments of SKOS Core not comply with the requirements or be sufficient, you can define new classes SKOS Vocabulary considers "*Concept*" the basic unit of any concept scheme. The class *skos:Concept* allows to model a particular resource in order to express this as a concept . A concept can be thought as something that can be defined or described. Each concept can have only one description or rather a single "*preferred label*" while it may have unlimited alternative descriptions or "*alternative label*".

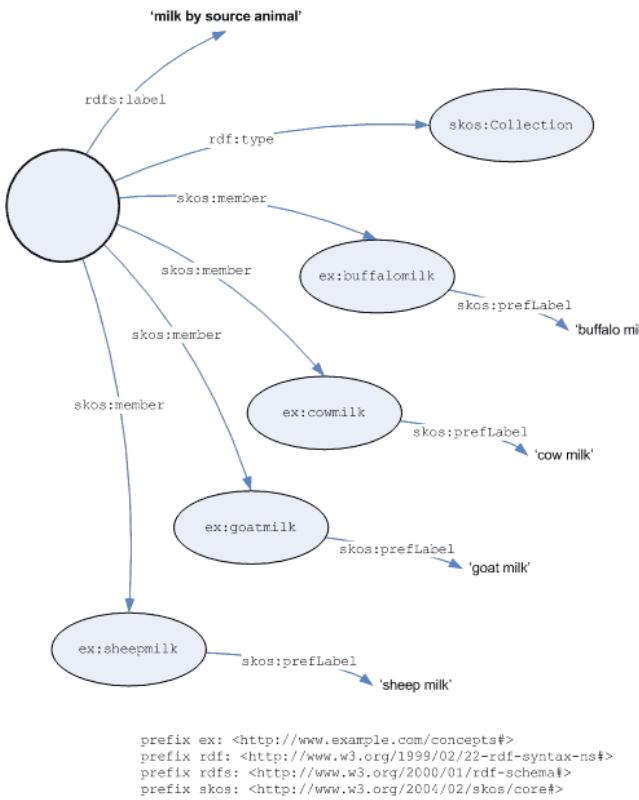


Figure 14 - Example of an RDF graph created using the SKOS

In SKOS labels are used to denote resources using the common language. There are different types of labels:

- *skos:prefLabel*, which is the preferred term for a certain concept and other synonyms;
- *skos:altLabel*, which is the most commonly used term for a certain concept and other synonyms;
- *skos:hiddenLabel*: a label "hidden" that is not normally visible but allows access to the content and to the applications that are text-based searches; Normally this tag is used to include syntactically incorrect variants of the same term : e.g. eletricity, electricity etc..

There are 7 properties that can be used to add information to the description of a concept. These properties have as superclass *skos:note* , class that can be used to provide any type of description.

The subclasses are instead:

- *skos:definition*: useful to give a complete and accurate explanation of the resource;
- *skos:scopeNote*: used to restrict or increase the description of the concept. Ex: "Microwave Frequencies : 1 GHz to 300 GHz "
- *skos:example*: used to provide examples of the use of the term or the concept
- *skos:historyNote*: used to indicate significant changes of meaning.
- *skos:editorialNote*: used to provide administrative information.
- *skos:changeNote*: used to keep track of the various changes that occurred.

5.1.2 FOAF

FOAF (Friend of a Friend) ontology is designed for the description of persons, groups, organizations, together with their interests, relationships and activities. FOAF has been designed to allow to structure the information relating to persons of a social network in a machine-readable. The major classes (see Figure 24) and FOAF ontology properties are divided into categories:

- **Core:** This category contains the classes and properties that make up the core of FOAF. They describe the characteristics of individuals and social groups that are independent from time and technology. In addition, FOAF defines classes to describe other entities such as: Project Organization and the Group.
- **Social Web** - in addition to the classes and properties of the core FOAF , there is a number of classes used to describe Internet accounts , web addresses, and other activities on the Web
- **Linked Data Utilities** - FOAF project initially defined as ' RDFWeb ' , over time it has become a widely adopted model for the publication of simple real data through a network of RDF documents linked . FOAF is an attempt to use the Web to integrate real data with human- oriented information in documents (ex, videos, books , spreadsheets , 3d models, etc)

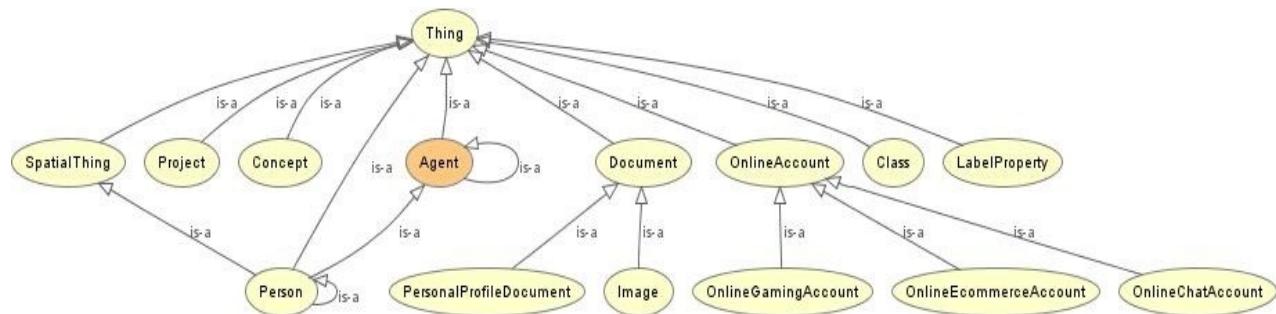


Figure 15 - FOAF ontology classes

The **class foaf:Agent** is used to identify the actors that are described in FOAF. These Agents can be natural persons (**Person**) , groups (**Group**) and organizations (**Organization**).

Can be associated attributes to a person to describe it, such as your name, email (foaf : mbox) and a personal image. Some of these attributes are designed to support the Social Web , such as web page (foaf : homepage), a reference to the online account that owns (foaf : OnlineAccount) and openid.

The presence of the classes **Organization** , **Group** (one Person can be member) and **Project** (connected to a Person through properties **CurrentProject** or **pastProject**) also allows to manage design areas . Also the interest (**foaf:interest**) could be used to map the skills of the user. It is, however, only basic information (for example projects are provided only with the name and the

homepage), so FOAF can be seen as a starting point and requires the extension with other ontologies , as we shall see in the following paragraphs.

The most interesting property of this ontology is foaf : knows , that links together two people who know each other, then analyzing all the properties of this type it is possible to identify the social network in which the individual moves.

In FOAF you can not use identifiers to identify the true and proper Person, but you define the inverse functional property: if an analyzer FOAF meets two people who have one of these properties in common (foaf can be in the email, or website) considers them as one and the same person .

5.1.3 ResumeRDF and DOAC

It is an ontology strongly diffused ,but FOAF is not able to express all the information relating to a person and for this needs to be extended with other ontologies . You can not , for example , track a person's skills , which can be mapped using instead ResumeRDF or DOAC.

Both **ResumeRDF**²¹ and DOAC (Description of a Career)²² ontologies are designed to express the semantic information contained in a CV ,and they have several points in common. Both use FOAF to describe the basic information on the people , but also personal ResumeRDF takes data from vCards and adds additional information such as the place of birth. DOAC is a format compatible with the European curriculum and describes in more detail the formative experiences : There are subclasses as PrimarySchool and SecondarySchool .

ResumeRDF is designed to be easily queried and it manages to capture a greater number of semantic data such as detailed information on the organization for which a person has worked . In ResumeRDF each competency has associated with it a name, a level and a number of years of experience , while in DOAC only four types of skills are possible (language , social , organizational and technical) .

5.1.4 SIOC

SIOC (Semantically Interlinked Online Communities) is an ontology created to interconnect the online community, representing semantic technologies with information about the structure and content of the community. The ontology is composed of a set of classes and properties (Figure 16) such as:

- **Site:** Defines the location of an online community or set of communities .
- **Forum** is a space for discussion , hosted on a website .
- **Post** can be an article, a message , an audio or video clip. A post is written by an "author", has a " topic" specifically, a "content" , "external links " , etc. .

²¹ <http://purl.org/captolo/resume-rdf/0.2/cv>

²² <http://ramonantonio.net/doac/0.1/>

- **User** is the account of a member of the online community.
- **Usergroup** is a collection of accounts of users interested in a common topic.

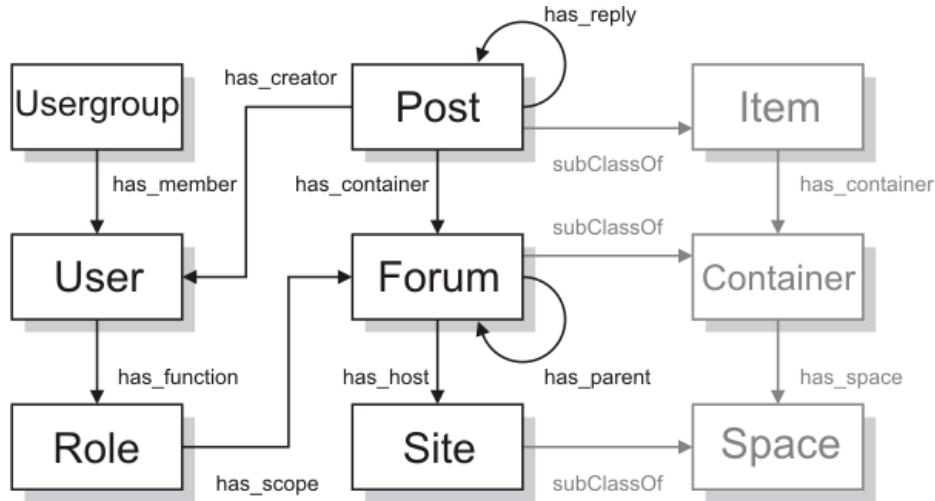


Figure 16 - SIOC ontology classes and properties.

SIOC is commonly used in conjunction with the FOAF vocabulary describing the relationships between people and information, and social networking with the SKOS model for the organization of the data. In particular (Figure 17), an instance of the Person of FOAF can have multiple online accounts (multiple instances of the User SIOC), while the topic of the post of SIOC can be expressed using the concepts of SKOS.

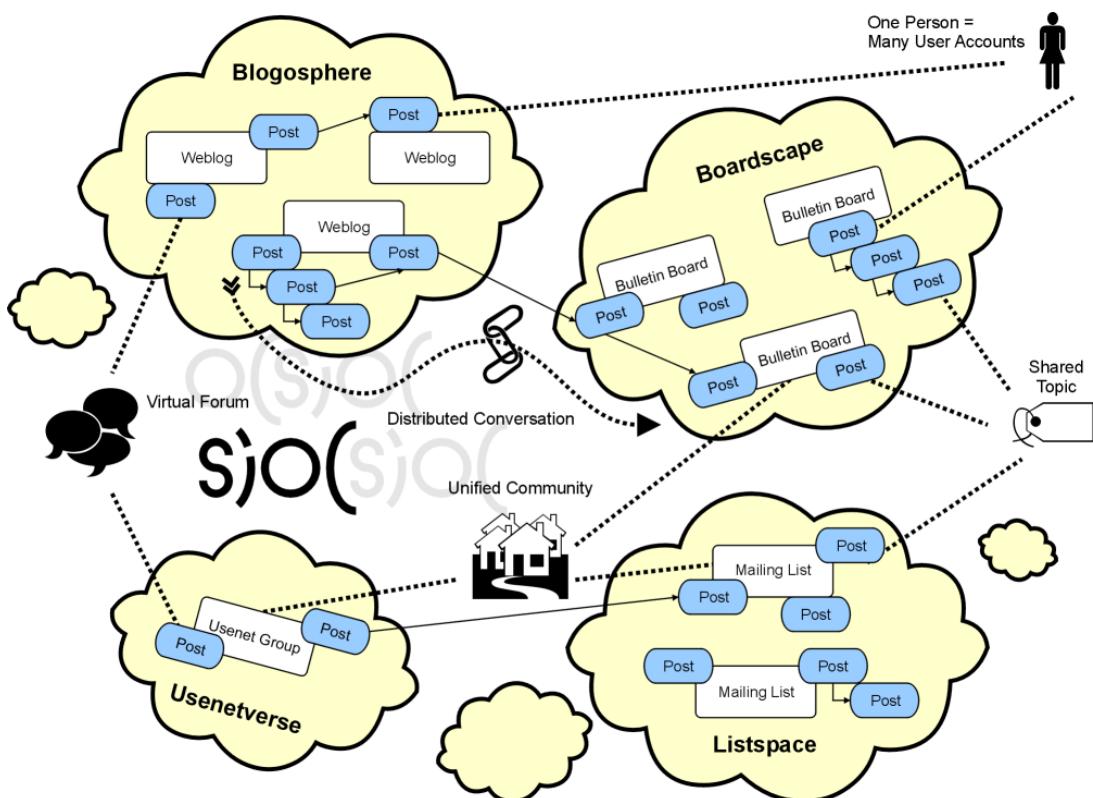


Figure 17 - Ontology connection between SIOC, SKOS and Foaf

5.1.5 SCOT

SCOT (Social Semantic Cloud of Tags) is an ontology describing folksonomies. In particular, it aims to describe the structure and semantics of data tagged and their relationships, and enable interoperability for social sharing and the reuse of metadata semantic tags from different sources. SCOT is an element of a tuple (U, T, R, Y) , where U is the set of users that participate in tagging, T is the set of tags, R denote the set of resources, each identified by one link called permalink. Y , however, is a ternary relationship between U , T and R , which represents the activity of tagging. It's also assumed that the tags are identified by a unique address (URI). This is true for systems such as Delicious and Flickr.

The users, tags and resources are represented using existing ontologies based on specific RDF (Resource Description Framework) for users FOAF, SIOC and SKOS resources for tags.

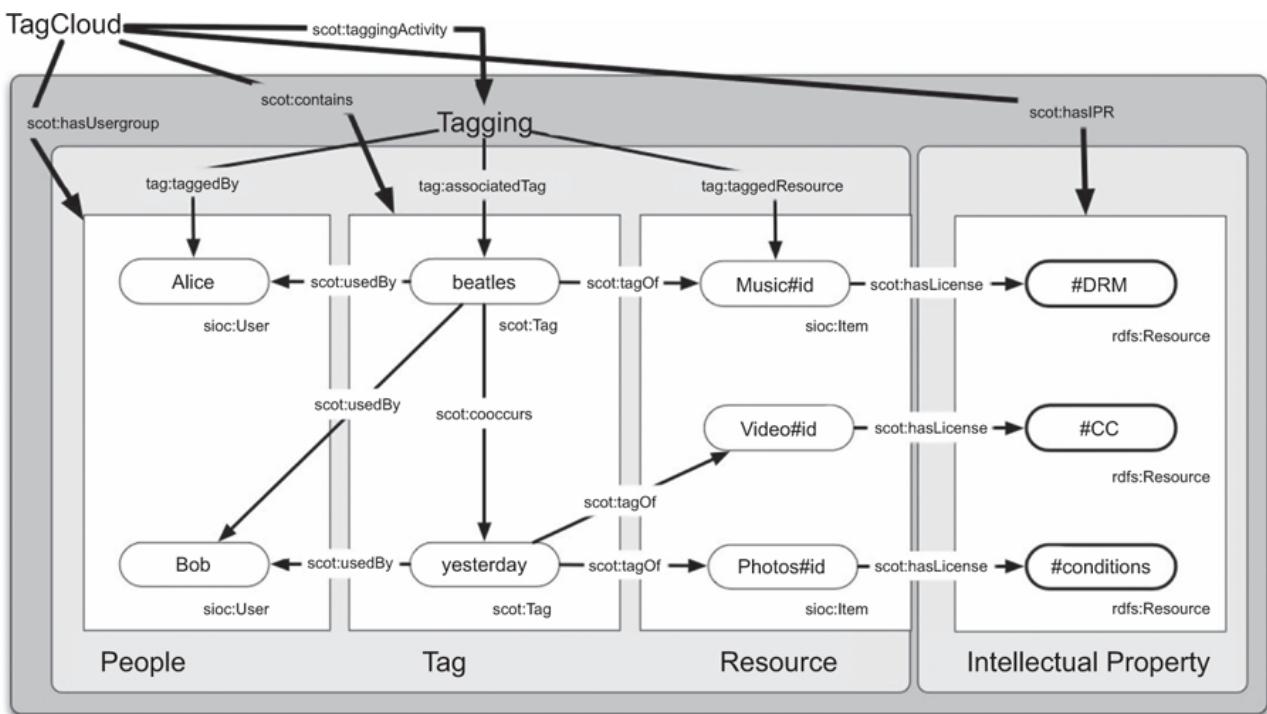


Figure 18 - Skos Classes and Properties

5.1.6 MOAT

MOAT (Meaning Of A Tag) [35] is another scheme, defining the semantics of folksonomies. Inside the ontology²³ the concept of tagging is expressed by the following tuple:

Tagging (User, Resource, Tag, Meaning)

Compared to the other definitions Tagging, the concept of meaning is introduced, which indicates the meaning of the tags in that particular instance. In particular, MOAT defines:

²³ <http://moat-project.org/ns>

- *the overall meaning of a tag*, the set of all the meanings that may be related to a tag in a folksonomy full ("apple" may refer to the fruit, the company or the record company - with the URIs of these concepts);
- *the local meaning of a tag*, or the meaning of a tag by a specific action of tagging (ex, with "apple" refers to the record in an action of tagging particular).

Meanings (Tag) = {(Meaning, {User})}}

One meaning is identified by a URI, often corresponding to the link to the relevant page in DBpedia. All users (shown in FOAF format) that associate meaning to that particular tag, they are shown.

```

<moat:Tag rdf:about="http://tags.moat-project.org/tag/apple">
<moat:name> <! [CDATA [ apple ]] > </ moat : name>
<moat:hasMeaning>
  <moat:Meaning>
    <moat:meaningURI rdf:resource="http://dbpedia.org/resource/Apple_Records"/>
    <foaf:maker rdf:resource="http://apassant.net/alex"/>
    <foaf:maker rdf:resource="http://example.org/user/foaf/1"/>
  </ moat : Meaning >
</ moat : Tag >
```

In the example above, two different people associate the word "apple" meaning "Apple Records", clarifying that have used that tag to refer to the company Apple.

The MOAT ontology reusing existing ontologies inside: FOAF to identify users, an extension of the Tag Ontology to identify the tags. In addition, MOAT may also be related to SIOC, defining the resource is marked as an instance of SIOC, but also directly linking to the tagged resource URI that represents the local meaning.

5.1.7 Online Presence Ontology

The OPO (Online Presence Ontology) [39] is an ontology for modeling the dynamic aspects related to the presence of an online user.

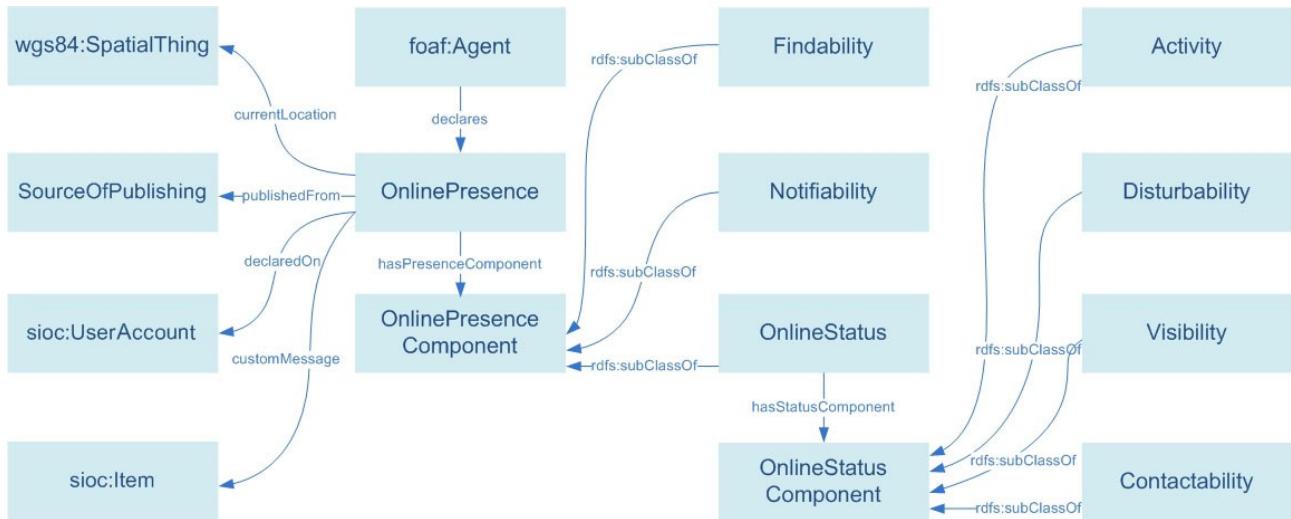


Figure 19 - Structure of OPO

A user, identified by FOAF, can manage some properties related to his online presence: visibility, activity, willingness to be contacted, the will to be disturbed. The user can choose what he wants to receive such notifications and profile information he wants to share.

You can specify the action that the user is doing at the moment (reading a document , participation in a project etc.) . Some features under development concern the faceted presence, the ability to give different groups of users different information about their online presence, and the definition of rules (e.g., if they are online and I am participating into the project " Semantic Web " then the members of the project team "Semantic Web" can contact me).

The faceted presence implies not only to give certain groups access to certain information, but also to control the granularity of the data (ex my friends can see the exact location from which access as strangers just my city) and render the data completely different to different groups (ex different status messages and varying availability depending on the group) .

5.1.8 GeoNames

Geonames is a free project for the creation of a database of geographic world . Its purpose is to provide the tools of translation of the name of a mountain or a city in which the data represent: latitude, longitude, elevation, population, postal code, etc..

Other geocoding systems are governed by precise rules and restrictions while Geonames is licensed under Creative Commons : anyone is free to use as he prefers the enormous amount of information available , provided that the source is acknowledged . Integrating geography in the Semantic Web becomes a task within the reach of any webmaster.

Today, the database contains over 10,000,000 geographical names corresponding to over 8,000,000 unique characteristics. All features are classified into one of nine classes and more not uncategorized features in one of 645 codes . Beyond the names of places in various languages are different data stored as latitude, longitude, elevation, population, administrative subdivision and postal codes. All coordinates use the WGS84 (World Geodetic System, standard for use in cartography , geodesy, and navigation) .

The names of the states are encoded in compliance with the standard ISO3166. The following describes the most important fields of Geonames ontology :

- Classes (in)
 - Code: code feature;
 - Feature: unique geographical object defined within Geonames;
 - Map: allows the display of the map;
 - RDFData: a document which contains a description of one or more characteristics;
 - WikipediaArticle : Wikipedia articles referring to the country that you are referencing;
- Datatype Properties (all having as domain a Feature)
 - Name
 - postalCode
 - population
 - ...
- Object Properties
 - childrenFeatures: links to an RDF document containing the description of the characteristics of children
 - incountry : country code taken from ISO list
 - locationMap map centered on the feature

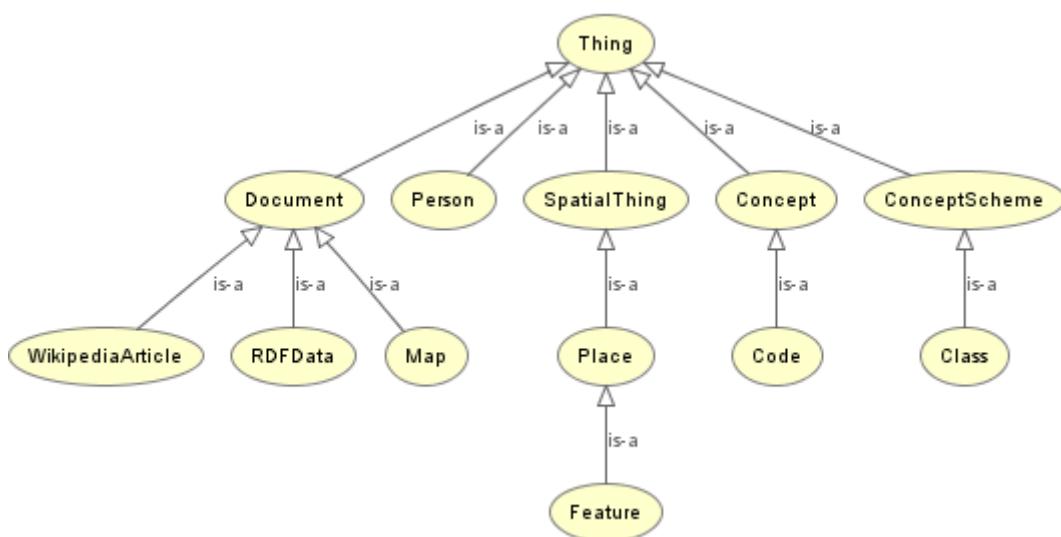


Figure 20 - Classes Geonames Ontology

The features of Geonames provide a categorization based on the selection of a class selected from a taxonomy available in the same ontology . For example , a feature relating to a geographical hotel will be labeled in accordance with the data which identifies the taxonomic class buildings (class S) and in particular from the code that identifies a hotel (i.e., S.HTL), so as, to differentiate Geographic annotation for the hotel from the others.

5.1.9 DOAP (Description of a Project)

DOAP (Description of a Project) is an RDF schema, describing software projects (in particular open-source projects) . You can map concepts such as the home page of the project, the developers, the programming language and the operating system on which the software is addressed. DOAP facilitates the creation of registers of projects by allowing sites " aggregators " to extract records of project from different sources and combine them into a single database. Reporting a project to an aggregator just create a DOAP file and publish it so that it is accessible via http and https requests . Updating the record it lasts that the project owner changes the DOAP file stored locally (on your website) and the aggregator will be immediately updated changes.

DOAP can be one of the possible candidates for mapping information on the projects , bearing in mind that This ontology can not be used to express the structure of the project and its other concepts of project management , such as the activities and work packages .

5.2 Tools of the Enhanced Enterprise 2.0

5.2.1 Semantic Search

The traditional search engines have some limitations. First of all, they are not able to understand the intrinsic meaning of the user query . Both to the query and the textual content of the resources are applied some techniques to manage the morphological changes of the words (stemming) , or to eliminate unnecessary words (stopwords removal) , but they are not managed cases of synonymy and polysemy: the algorithms are not able to understand the sense in which a word is used in the sentence, nor they are able to enrich the query with synonyms. Even the order of words in the sentence is not taken into account , and so this makes the simple full-text search unreliable.

In addition, the search results are often displayed as a simple list of resources , while having semantic information about the type of resource returned ,the display may be customized depending on the type of data.

From here the introduction of semantic search engines that develop the traditional system of Information Retrieval (IR) from a simple system of Document Retrieval to a system of the Entity and Knowledge Retrieval , on the other hand , improve the conventional methods of IR , under different points of view: the meanings of words can be formalized and represented in a machine-processable format using ontology languages such as RDF and OWL (a resource can be described as an

ontological class , with a set of attributes , relationships with other entities , constraints , etc. .) . With the logical representation (Description Logic) of resources, a semantic search system is able to recover significant results through a process of inference on the query and on the Knowledge Base (KB).

A semantic search engine is a search engine that aims to understand the intentions of the person who make a search and the contextual meaning of the terms entered. Typically these applications have the following characteristics [40] :

- **Management of natural language queries:** the user must be able to enter his question in natural language, without using Boolean operators or complex constructs .
- **Match between the concepts:** once identified the key elements of the user's question , it is necessary to find matches with the concepts present within the domain ontologies to perform semantic search .
- **Knowledge Base:** it is necessary to possess a knowledge base , as thoroughly as possible , which helps the matching between the concepts identified in the application and documents of the system, providing, for example ,synonyms and relationships to all of the terms most relevant .
- **Management of morphological variations:** as for full-text search engines , the semantic engine must be able to handle all the morphological variations of a word (plurals, abbreviations, etc.).
- **Management of synonyms and word meanings:** the original query should be expanded to include synonyms of the original word. The documents that contain synonyms must be taken into account only if the meaning of the word used in the document actually corresponds to the meaning of the original word , avoiding the problems of polysemy .
- **Management of generalizations:** it is necessary to exploit the hierarchical relationships between terms in the ontology knowledge base to improve the results. For example, a query that contains a specific concept must be expanded with its concepts it must be generic or can answer to a question expressed in generic form identifying the specific concept matching .
- **Operation unsupervised:** the search engine needs to use only the analysis of the contents of a document and its associated metadata. Clearly, this mode of operation (data- driven) introduces a major complication in terms of the quality of the extracted result , but allows greater applicability in different domains because the analysis is not affected by what you want to search (pattern) as usually happens in the case of traditional expert systems .
- **Identification of performance:** every search result must be associated with a higher level of reliability based on the degree of relevance with the concepts sought . In this way it is possible to make a ranking of the results and exclude those with reliability too low.

The semantic research has also found a place in the field of Enterprise ,currently , are present products like CloudView²⁴, Sinequa²⁵, SmartLogic²⁶, OpenText²⁷ consisting of search-based applications that provide access to structured and unstructured data in the system. In particular SmartLogic is based on the use of controlled vocabularies (taxonomies and ontologies) to allow the automatic classification of information and the navigation of the search results.

5.2.2 The semantic wiki

The Semantic Wiki combine to semantic technologies the ease of use of Wiki technology , as it is of Wiki the construction of which occurs according to a modeling ontological below.

Wiki 's normal all the knowledge is contained within texts and multimedia file, so it is easily accessible by humans but is not able to support advanced searches, or to allow combinations of existing information.

In Semantic Wiki , however, the presence of patterns OWL (Web Ontology Language) that guide the construction of knowledge, allows to capture or identify the information contained within the pages and the relationships between pages , creating patterns that can be easily queried and analyzed to identify new information , ensuring internal consistency of the information and allowing interoperability between different applications. Regarding the schema SemSLATES, the Semantic Wiki exploit the instrument **Semantic Links**.

An example of a Semantic Wiki is the Semantic Media Wiki , in which every single wiki page corresponds to an ontological element , which is easily mapped into OWL : the pages that describe the elements of the domain of interest become individuals , the categories (classifications elements) become OWL classes , the relations between pages become property .

²⁴ <http://www.3ds.com/products/exalead/uri/products/cloudview/>

²⁵ <http://www.sinequa.com/en/page/product/sba-platform.aspx>

²⁶ <http://www.smartlogic.com/>

²⁷ <http://www.opentext.com>

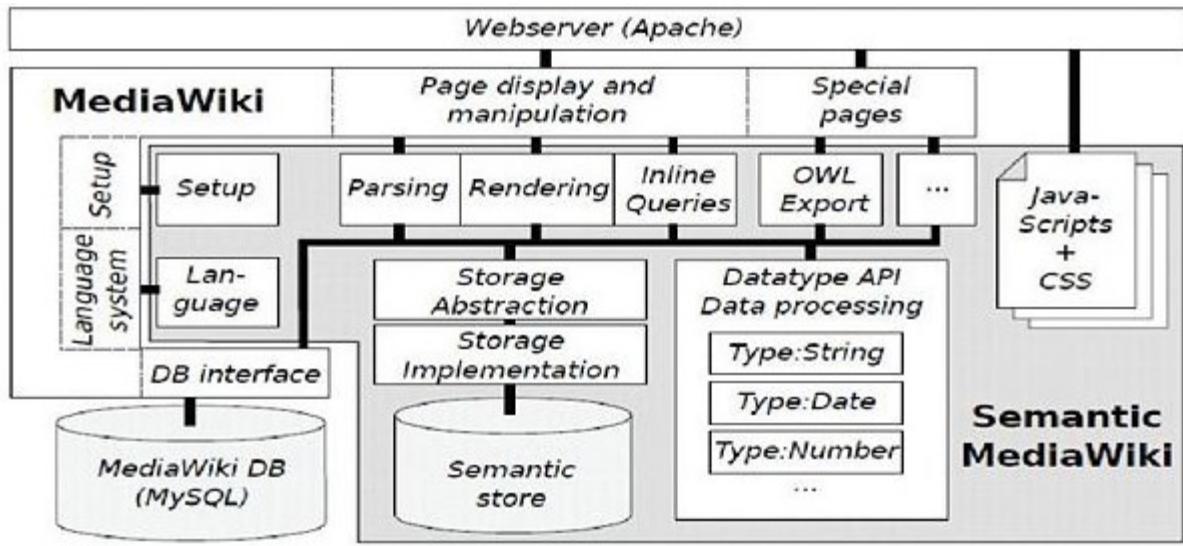


Figure 21 - Architecture of Semantic Media Wiki, which also applies to the Semantic Wiki in general

As you can see in Figure 21, a Semantic Wiki maintains a database of articles and an RDF Triplestore containing semantic information. The content of Triplestore is used for research and statistical aggregations and can be exported in OWL format to be used in other applications.

5.2.3 Semantic Social Tagging

The Semantic Social Tagging consists in modeling semantics of folksonomies . Among the domain-specific schemes designed for the representation of the tags (Sections 5.1.5 and 5.1.6) the most complete is MOAT ontology. MOAT differs from other tagging ontology as an entity tagging that includes an additional concept : the meaning.

The meaning is specified using URIs of instances of domination or resources on the public Knowledge Base , as DBpedia and the resources of the Linking Open Data project. By assigning a meaning to each tag avoids the problems of polysemy , as the meaning of a tag is disambiguated via the URI, and synonymy, as in more words indicating the same concept is associated with the same URI. MOAT also allows to identify the author of the tag, mapped using FOAF , and the resource tagged , mapped using SIOC.

Using the Social Semantic Tagging you can create a classification of all the resources in the community enterprise , as you can group items tagged with a particular concept identified by a specific URI , regardless of type (blog , wiki, etc. .) and origin. The tags may have also created semantic relations with other tags , as they exploit the semantic relationships between the concepts associated with them.

5.2.4 Semantic Social Networks

Using ontologies such as FOAF and XFN and hCard microformats you can describe demographic information of a person , his profile (interest, etc. .) and relationships with other people. The use of

semantic technologies helps in identity management , as it allows people with multiple accounts on different platforms to unify their identity information and it provides a standard model that allows you to uniquely identify a person not only inside the ' Enterprise Social Network but also in other social networks in which it belongs. In this way , you can get the content that a person has produced within the platform.

The connections between people can be identified not only by explicit links (by browsing the graph property foaf : knows), but even watching the links implicit : two people can be considered related if they appear in the same photo , tag the same documents or respond to each other their blog posts. Extending FOAF with ontologies as ResumeRDF or DOAC (Section 5.1.3) can also be mapped to curriculum information , having available to every employee a corporate semantic profile.

The semantic representation can be extended to the whole Social Network account using the SIOC ontology to represent the activities of community enterprise and its contents. Using SIOC trying to solve the problem of lack of integration between social software and other systems within the enterprise intranet , storing all the data in a triple store common ,making it easier to search across multiple resources Enterprise.

Use SIOC and FOAF profiles enriched with ResumeRDF or DOAC can be used to look for the right people within the community to assist the employees in the performance of their duties. The connections (explicit and implicit) between the person performing the search and other users ,together with the information on employees skills , are analyzed to suggest users to contact. SIOC also helps to identify relevant topic to the research and the individuals who have

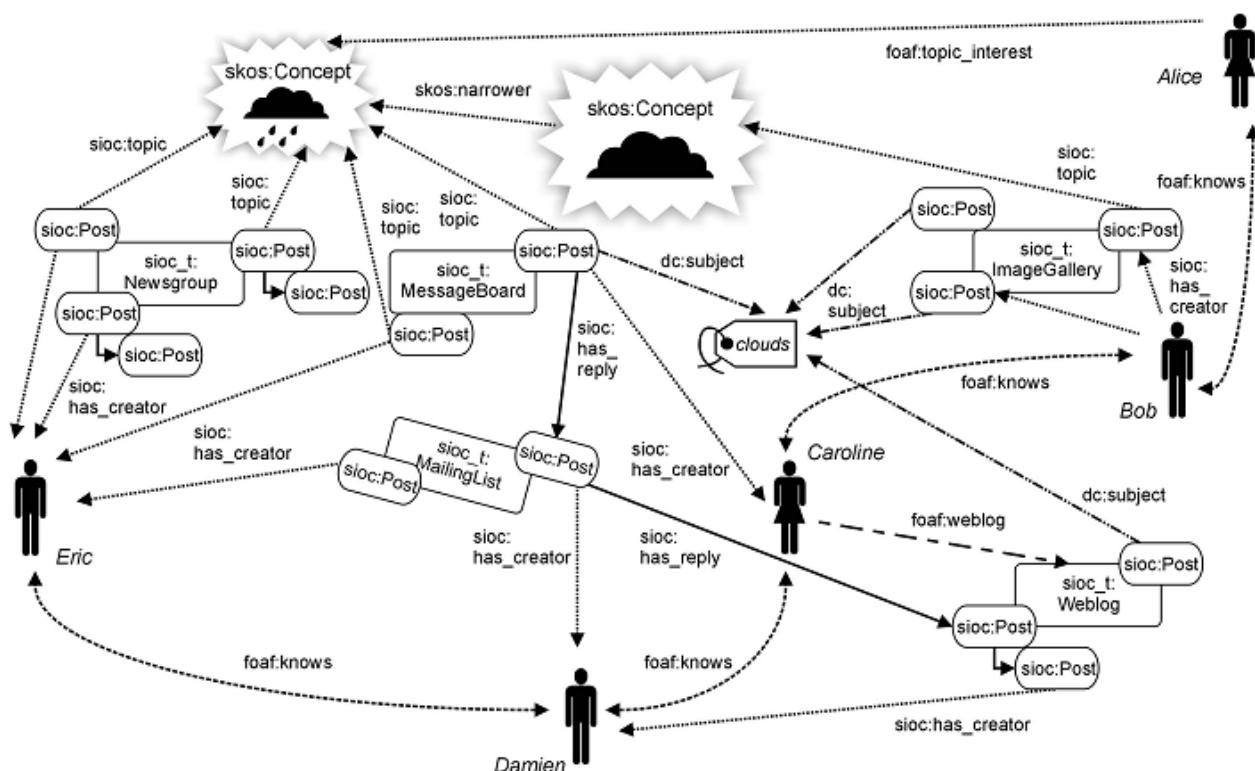


Figure 22 - Navigation of resources and users by using the concepts of SIOC , FOAF and SKOS .

Figure 22 shown an of semantic navigation of resources and users. Alice expresses an interest in the skos : example [36] of realization Concept " rain " and wants to find experts in the domain. At the moment it is connected (report direct knowledge) to Bob , whose posts and comments of image galleries indicate that has an interest in the *skos:Concept* "clouds" (semantically related to " rain") but not in the rain.

None of the members of the social network has extended to Alice as foaf : topic_interest "rain" , but Alice still manages to find a message board that Carolina has created a SIOC : Post which has as its topic " rain."

Alice can see the shortest path between Caroline and her inside knowledge of the graph , and finds that both have Bob as friend in common, so she asks Bob to be present Caroline.

Looking at the message board where it is inserted the post on "rain" , Alice finds another user Eric that has published several posts on the subject. Alice can view posts by Eric on any community, so through his post finds a Usenet newsgroup where the topic " rain " is discussed in greater detail.

5.2.5 Semantic Microblogging

The Signals of SLATES paradigm , the RSS Feed and the posts of the Microblog , are particularly suitable in the field of Enterprise to keep users up to date on what is happening within the company and the community. Growing to the size of the online community is born, however, a problem of information overload, employees receive huge amounts of updates. You need to create a filtering system updates based on the user's profile in the community, and this can be done using the Semantic Web technologies (implementing the Semantic Signals of SemSLATES) .

Among the projects related stands SMOB (Semantic Microblogging) [41], a framework for semantic microblogging that lets make blogging activities through the use of Semantic Web technologies and Linked Data.

SMOB is based on:

- **Ontologies** , used to define a common semantics to represent the post of blogging, so that they can be reused by each service capable of consuming data RDF (S) / OWL.
- **Hub distributed**, used to publish the data and exchange information using ontologies .
- **Connections between the components**, which make the post part of the Linking Open Data cloud²⁸.
- **Technology faceted**, which makes it possible to read only the status message corresponding to

²⁸ <http://linkeddata.org>

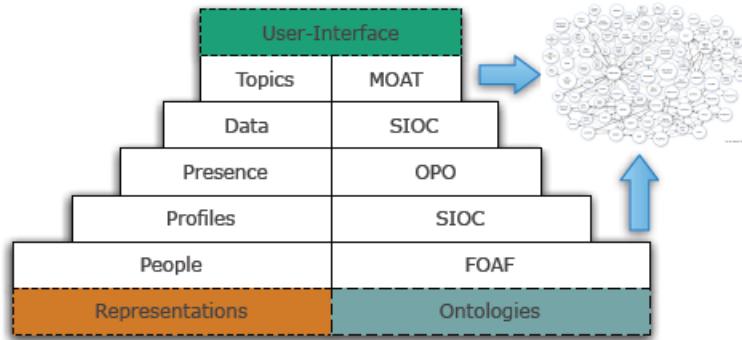


Figure 23 - The ontologies used by SMOB

It is precisely faceted technology to make the filtering of data requested within the Enterprise (and not only).

SMOB uses FOAF to define the people and the relationships between them , while the update of the microblog are managed through an extended version of SIOC, which adds classes **sioc:Microblog** and **sioc:MicroblogPost** and SIOC properties: follows (for notifications from people that the user follows or from which below) and **SIOC:addressed_to** (to identify to whom addressed the post) .

It is used the Online Presence Ontology to describe the user's status : online presence, visibility, activity, willingness to be contacted , the will to be disturbed. The user's geographical coordinates are expressed using GeoNames and OPO. MOAT (Meaning of a Tag) is used to model the tags and get information from the Semantic Web SMOB automatically suggests users the URL to use in the operations of tagging by reference to Sindice²⁹ (the Semantic Web Index) or DBpedia³⁰ (the version of RDF Wikipedia) .

5.3 Conclusions

The integration of semantic technologies within the Enterprise 2.0 tools leads to increased availability of content and to enhance their value . We solve the problem of synonymy and polysemy identifying concepts with unique URI . It avoids the problem of information overload because the information flows as RSS feed and post -blogging are filtered according to the context in which the user is located.

These semantic technologies should also improve the process of creating Best Practice, as representing semantically proposals and their content, they can be grouped by these concepts, making easier the detection regardless of the collaboration tool with which they were produced.

²⁹ <http://sindice.com>

³⁰ <http://dbpedia.org>

The analysis and conceptualization of Best Practice is in fact carried out by the same semantic tool. Using SPARQL queries and techniques for Semantic Search in general is also easier the identification and the selection of proposals for improvement.

Part II: Methodologies & Applications

6. Methodology for Knowledge Extraction and Classification

In this chapter, shall be defined methodologies useful for the indexing of the content of the application domain of interest. The research work focuses on methodologies and processes to enhance the knowledge of the organization, in particular the tacit, to improve competitiveness and growth.

Will be defined a methodology for Knowledge Classification able to associate the resources available to known categories, belonging to a classification created and updated based on the contents of the input resources. The latter can be characterized by unstructured data (documents, presentations, etc. available within the organization), but also by instances of the models defined on the domain of interest.

The chapter presents an overview of techniques for indexing the contents with respect to models of domain specific knowledge with particular reference to classification techniques suited to the purposes of the project. Therefore, will be described first the methodology of Knowledge Classification and subsequently, will be defined Knowledge Extraction and Ontology & Instance Matching methodologies used to support the methodology for knowledge extraction.

6.1 Methodology for Knowledge Classification

The methodology for the Knowledge Classification is based on a hybrid approach based on two algorithms:

- **Rule Based:** Focused on pattern matching algorithms that classify a resource in a certain category based on a set of criteria that define the conditions of belonging to it. The pattern matching algorithms are able to perform the classification without a phase of training.
- **Instance-based:** algorithms focused on K-Nearest Neighbour technique that classify objects according to what are "close" to the demands of the training set. The classification is made through the identification of the K nearest instances [36]. These classification algorithms must be trained on the training set. The metrics to assess the closeness between resources is given by the similarity measure defined by the Euclidean distance. The evaluation of the ambiguity of the results of classification and error estimation are key elements for the implementation of a new activity training.

Instance-based algorithms perform the task of classification on the basis of the training set. The training activities will be supported by a Knowledge Extraction methodology which will be detailed later. The purpose of this methodology is the identification and extraction of concepts characterizing the input resources (structured and unstructured) by means of data-driven approaches and unsupervised. The results produced by the process of extraction of knowledge are two: Unsupervised Conceptualization (UC) is a hierarchy of categories is not known, and content classification with respect to the categories of resources extracted. This classification uses the instance-based classifier trained during the process of Knowledge Extraction. Since the process of

Knowledge Extraction is expensive in computational terms, will be performed in off-line mode in the initial phase and periodically to train the classifier instance-based on new input resources. In order to classify a resource with respect to a set of predefined categories (SKOS Vocabulary) introduces a methodology & Ontology Instance Matching aims to align the conceptualization produced by Knowledge Extraction (Unsupervised Conceptualization) the classification of the default categories (SKOS Vocabulary). Figure 24 show the interaction between the processes of Knowledge Extraction, Ontology & Knowledge Instance Matching and Classification in accordance with those previously described.

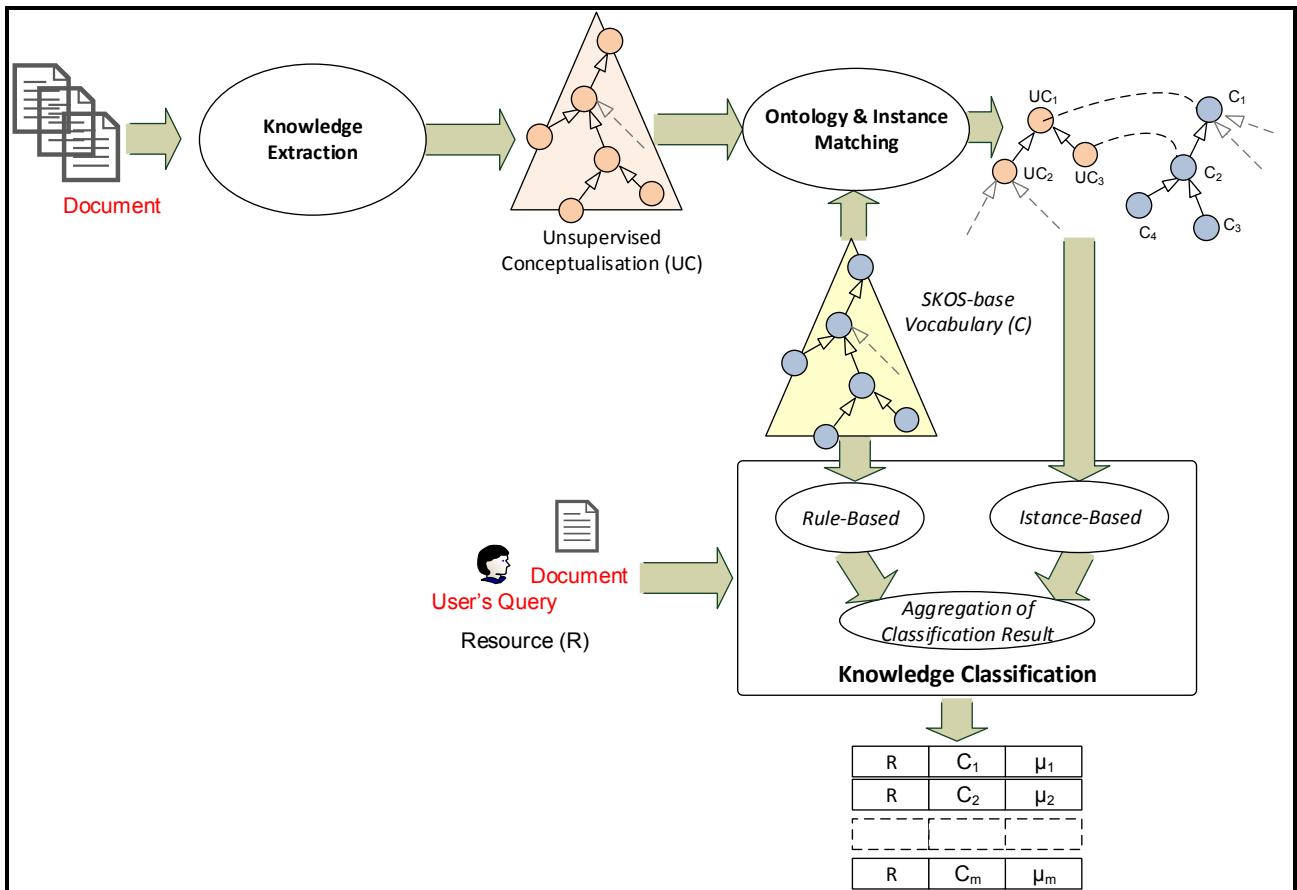


Figure 24 - Processes interaction KE, IOM and KC

The process of Knowledge Classification combines the classification results obtained by the algorithms Instance-Based and Rule-Based. Table 1 show the input and output of the Knowledge Classification methodology.

Knowledge Classification	
INPUT	OUTPUT
<ul style="list-style-type: none"> Resources (for example, user queries, documents, etc.); SKOS-based vocabulary; 	Set of concepts in the SKOS Vocabulary and classification degree of the input Resource

<ul style="list-style-type: none"> Matching between concepts and concepts <p>Unsupervised conceptualization of SKOS vocabulary.</p>	
--	--

Table 4 - Input/Output of Knowledge Classification methodology

The two components, Instance-Based and Rule-Based are detailed in the following figures.

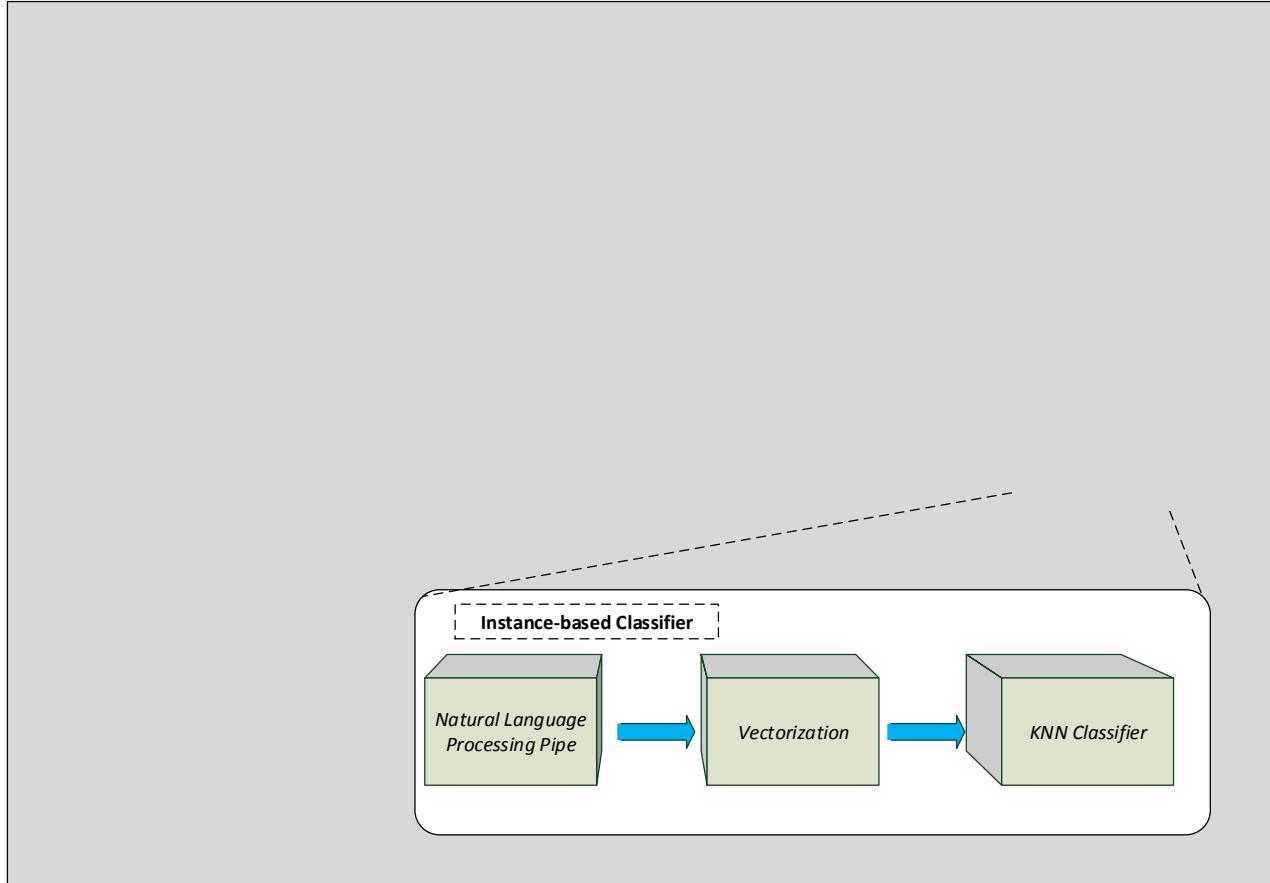


Figure 25 - Knowledge Classification Pipe Phase

Figure 25 illustrates the overall process of Methodology for Knowledge Extraction. It is essentially composed of the following phases:

- **Pipe Natural Language Processing:** is executed to retrieve relevant terms from the content of the resource given as input (Section);
- **Vectorization** is performed for the construction of the vector model of the content of the resource given as input, according to the features extracted during the extraction of Knowledge (more details on this will be provided in the form of Knowledge Extraction);
- **KNN Classifier:** The classifier trained during the extraction of knowledge is performed on a model of the resources provided in the input vector in order to extract the concepts Unsupervised Conceptualization associated resources. The results of the KNN classifier express the degree of membership of each of the resources provided as input to the concepts of the Unsupervised Conceptualization.

The Rule-Based Classifier (Figure 26) is characterized by a phase of pattern matching, this activity is based primarily on the recognition of a pattern within a string made taking into account the definition of concept available in SKOS vocabulary and provided input into the content of the resource . Some regular expressions are made to find a concept in the text using the SKOS properties, such as: *prefLabel*, *alternativeLabel*, etc., Activity identifies the degree of membership resources classified SKOS vocabulary and concepts provided in the input.

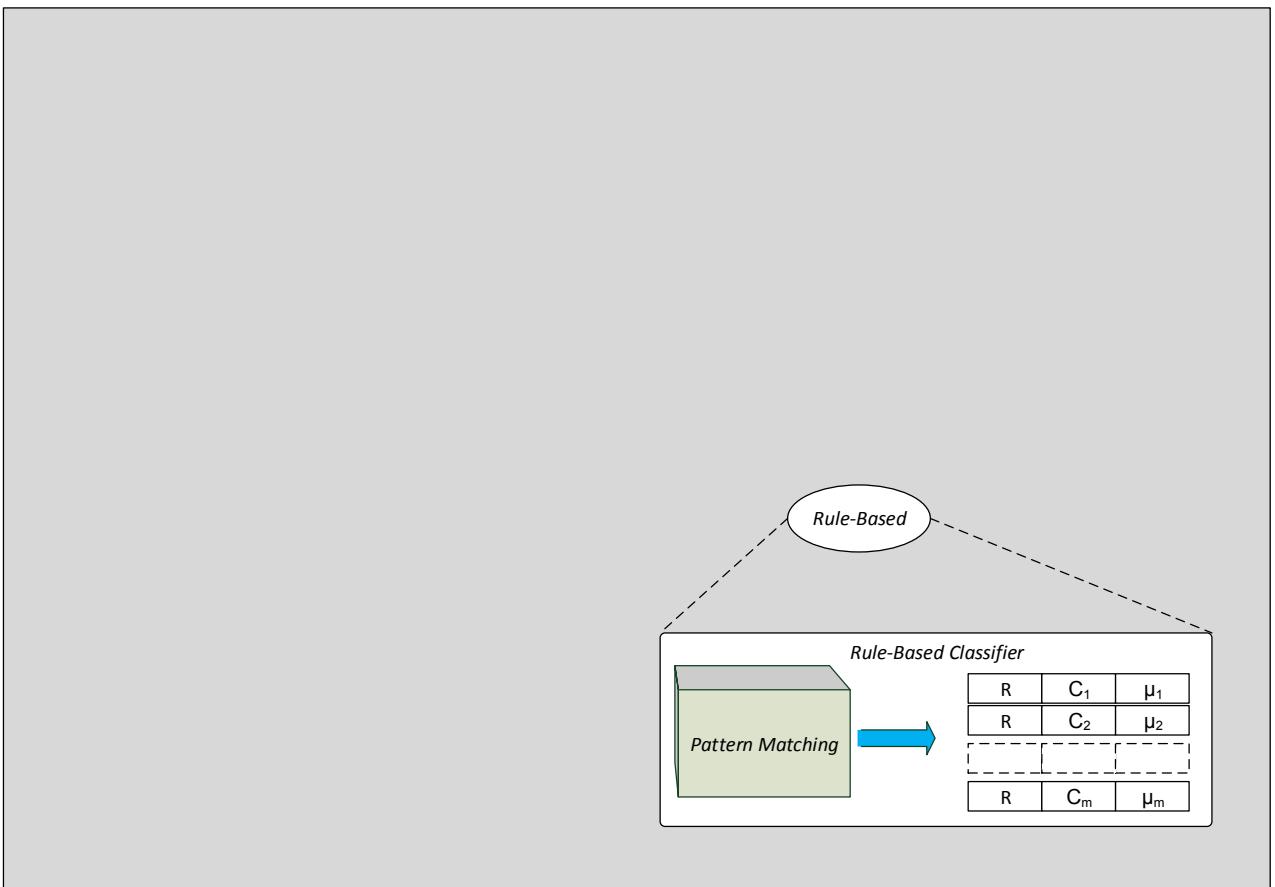


Figure 26 - Rule Based Classifier Phase

Finally, the Aggregation Classification Result Module (Figure 27) combines the results of the classification Instance-based and Rule-based system for classifying resources with respect to the set of concepts available in the SKOS vocabulary with adjoining measure of classification.

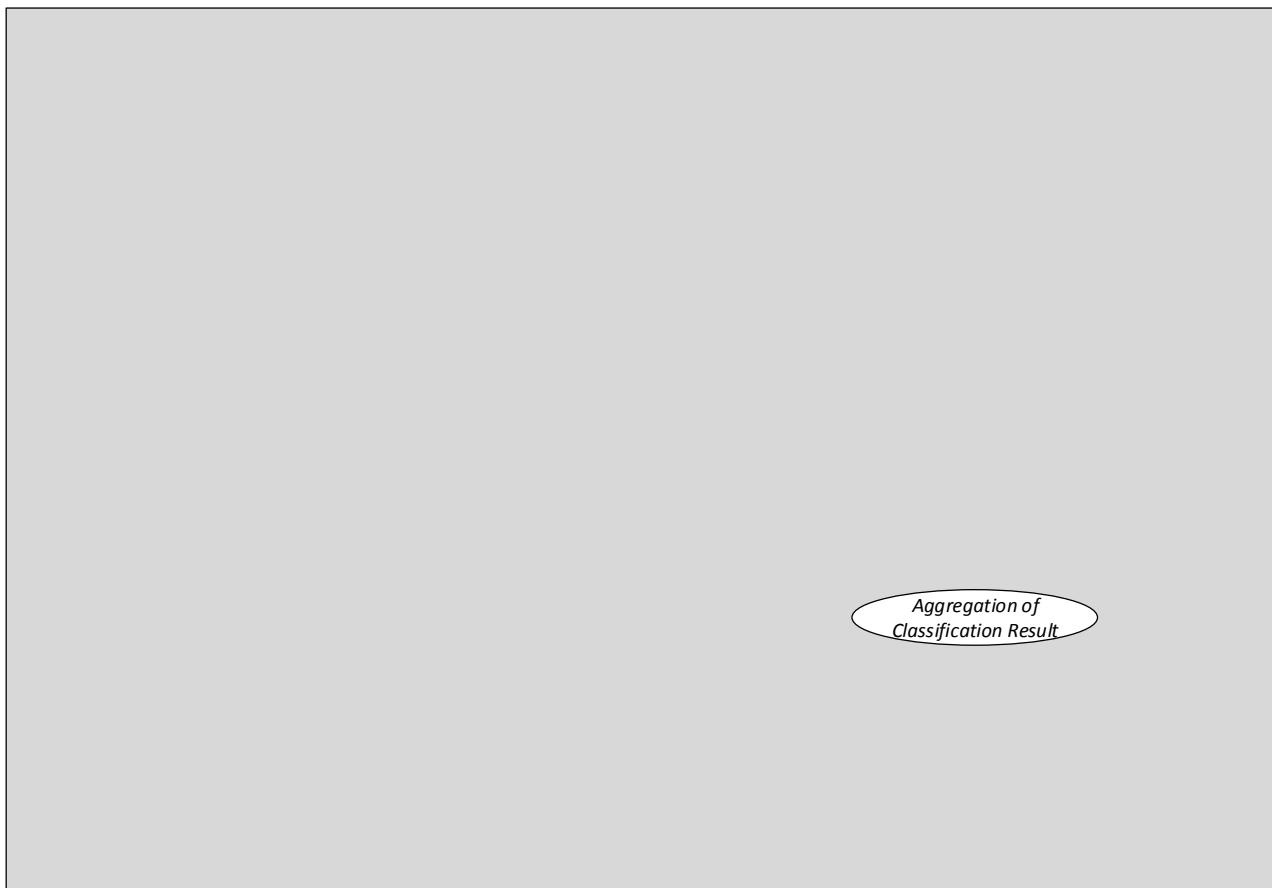


Figure 27 - Aggregation Classification Result Module

6.2 Methodology for Knowledge Extraction

In this section we will describe the methods used for the extraction of knowledge. The latter is characterized by automatic techniques for identification of concepts and relationships in the domain of interest through the analysis of structured (e.g., relational databases, XML, etc.) and unstructured (e.g., text, documents, etc.) sources. The use of these techniques may be relevant to solve various organizational problems. Firstly, the design and construction of domain ontologies and taxonomies is a laborious process that requires substantial resources, in terms of cost and effort [37]. In addition, it is necessary to take advantage of the life cycle of knowledge (from tacit knowledge to explicit knowledge) in order to update the classification of the organization's domain of interest [38].

The methodology of knowledge extraction is a process aimed at enabling the semi-automatic construction of taxonomies from existing repositories and organizational data sources in order to obtain useful information, for example, get information on how your employees use their skills using data from Web 2.0 tools (Enterprise Wiki, Corporate Blog, etc.) and text data documents. This is done in order to summarize the Enterprise Contents and give them a hierarchical structure, according to specific needs.

In particular, the methodology for the extraction of knowledge will be applied to the textual content in order to obtain an intrinsic meaning of resources. Many types of textual input were taken into

account from the knowledge extraction methodology. Specially, the methodology handles the following types of digital resources:

- Working documents produced by individuals or by the collaboration of several people within the organization (i.e., design documents, publications, etc.);
- User -generated content (i.e., blogs, wiki entries, etc.)
- Curriculum Vitae (CV), Excel sheets (containing the best practices of establishment);

The output of Knowledge Extraction are represented in a readable format to an electronic computer through the languages of the semantic web, such as: RDFS³¹, OWL³² and SKOS³³. These technologies are compatible with models Automotive. The final results that are to be produced through a methodology of knowledge extraction are:

- **Unsupervised Conceptualization**, hierarchy carried out extracting concepts taking into account resource's content.
- **Resource Categorization**, that means the weighted association between resources' content and concepts belonging to Unsupervised Conceptualization.

6.2.1 Process Definition

Before defining this methodology, it is necessary to identify and characterize the input and output of extraction of knowledge process, which are described in Table 5.

Knowledge Extraction “Processing Pipe”	
INPUT	OUTPUT
<ul style="list-style-type: none"> • Digital resources of heterogeneous nature: documents and deliverables produced by employees, resources from Web, scientific papers, CV, forums, blogs, wikis, etc. 	<ul style="list-style-type: none"> • Hierarchical conceptualization of resources' content represented by exploiting semantic technologies, such as: RDF, RDFS, OWL, SKOS.

Table 5 - Input/Output of Knowledge Extraction process

To obtain a hierarchical conceptualization of resources, you first need to represent all their text content in a mathematical model. In particular, each resource is encoded in the vector space model by using a representative set of relevant terms (feature) weighted according to the occurrences in the text. These carriers are analyzed in order to obtain the concepts and relationships between them (for example the relation of subsumption). The analysis produces hierarchies of concepts, employees of the mathematical model, encoded in an appropriate RDF model in order to store them in a semantic repository. Figure 28 illustrates the process for the extraction of knowledge. It essentially consists of the following phases:

³¹ Resource Description Framework Schema (RDFS): <http://rdf.org/resume-rdf/>

³² Web Ontology Language (OWL): <http://www.w3.org/TR/owl-features/>

³³ Simple Knowledge Organisation System (SKOS): <http://www.w3.org/2004/02/skos>

- **Natural Language Processing** – extracts bag of words from the text contained in User Generated Content (Section 6.2.1.1);
- **Vectorization** – selects a feature set, weights it and stores the results in an index (Section 6.2.1.2);
- **Concept Data Analysis** – extracts concepts hierarchies represented by means of a mathematical model (Section 6.2.1.3);
- **Semantic Technologies Mapping** – encodes the hierarchy of concepts extracted with semantic languages can be interpreted by an electronic computer (Section 6.2.1.4).

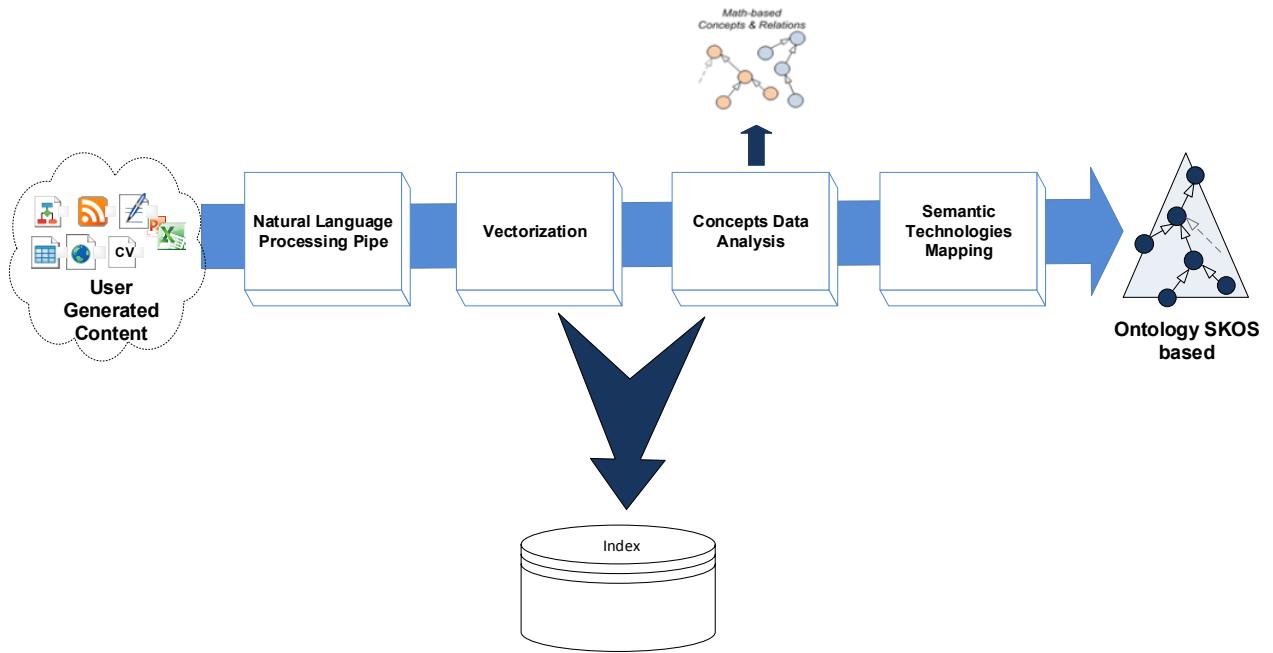


Figure 28 - Knowledge Extraction Process

The following sections provide a detailed overview and an example of the execution of each phase of the Extraction of Knowledge methodology.

6.2.1.1 NATURAL LANGUAGE PROCESSING

The first phase of the process begins with the execution of a Pipe NPL. This process consists in analyzing textual information in natural language to extract and disambiguate terms in the context in which they are used. As shown in Table 6, the main result of the Pipe NLP is a set of disambiguate terms.

Natural Language Processing Pipe	
INPUT	OUTPUT
<ul style="list-style-type: none"> Digital resources of heterogeneous nature: documents and re-sources from Web, scientific papers, forums, blogs, wikis, etc. 	<ul style="list-style-type: none"> A set of most relevant terms (i.e., features), their associated senses and digital resources in which they appear.

Table 6 - Input/Output of NLP Pipe phase

Figure 29 shows a detailed view of the NLP Pipe. In particular, Figure 29 highlights activities of NLP Pipe (e.g., PoS Tagging and Concept Disambiguation) that take place during text parsing in order to extract the right set of terms in the resource's content.

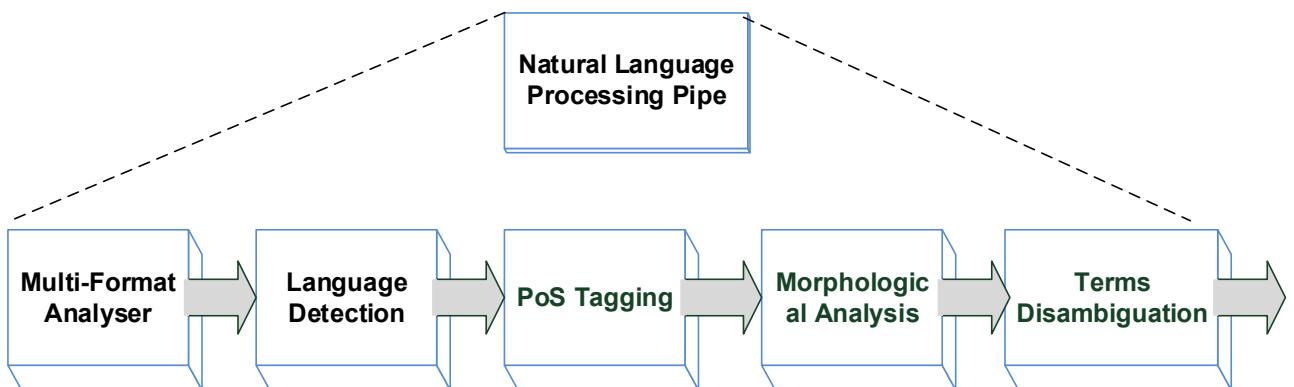


Figure 29 - Knowledge Extraction: NLP Pipe

For the task of NLP Pipe we consider the following steps:

- **Multi Format Analyser:** understanding the format (PDF, doc, HTML etc.) of the input documents and extracting text from them.
- **Language Detection:** understanding the language in which the text is written, in order to configure language dependent tasks properly.
- **Part of Speech Tagging:** the classification of words into lexical categories. There are many implementations of lexical parsers, released with different kind of licenses.
- **Morphological Analysis:** it is the identification, analysis and description of the structure of morphemes and other units of meaning in a language such as words, affixes, parts of speech, intonation/stress, or implied context.
- **Terms Disambiguation:** the activity of automatically assigning the most appropriate sense of a polysemous term by analysing the context in which it is used. Terms Disambiguation is useful because different senses of a polysemous term can be treated as different terms of the feature set.

6.2.1.2 VECTORIZATION

The vectorization process is aimed at the conversion of the relevant terms (features) that describe a resource in a vector representation. Table 7 shows the input and output of the process of vectorization.

Vectorization	
INPUT	OUTPUT
<ul style="list-style-type: none"> A set of most relevant terms (i.e., features), their associated senses and digital resources in which they appear. 	<ul style="list-style-type: none"> A term-document matrix containing vectors associated to each input resource, in which columns are resources themselves and rows are representative features.

Table 7 - Input/Output of Vectorization phase

6.2.1.3 CONCEPT DATA ANALYSIS

Concept data Analysis applies techniques of data analysis to extract the concepts and the relationships between them through the analysis of the resources and of the terms in a given domain. Table 8 show the input and output of the process of Concept Data Analysis.

Concept Data Analysis	
INPUT	OUTPUT
<ul style="list-style-type: none"> A set of most relevant terms (i.e., features), their associated senses and digital resources in which they appear. 	<ul style="list-style-type: none"> A term-document matrix containing vectors associated to each input resource, in which columns are resources themselves and rows are representative features.

Table 8 - Input/Output of Concept data Analysis micro-phase

The main result of the Concept Data Analysis phase is the hierarchical structure of the content of resource. This process takes as input the term-document matrix created in the previous step. Considering this term-document matrix, the Concept Data Analysis aims to develop the digital resources (which are called objects) and features (which are called attributes) according to a shared meaning. Intuitively, we are interested in grouping together the maximum number of objects that share the same set of attributes, and vice versa. In the literature, there are many techniques that can be used to group the data and features: Clustering, Hierarchical Clustering [39], LSA [40], FCA, Fuzzy FCA. The most appropriate technique is selected taking into account the following aspects:

- **Hierarchical Conceptualization:** it refers to the structure of the conceptual groups obtained by applying the specific technique. In this sense, Clustering and LSA return flat clusters.

- **IR Support:** it refers to the Information Retrieval³⁴ support provided by the output of specific technique. Table 9 highlights that all techniques provide useful support for Information Retrieval of grouped resources;
- **High cost of updating the structure:** it focuses on the cost of updating the structure obtained from a specific technique by adding daily new resources to the enterprise repository. In this sense, Table 9 highlights that Clustering, Hierarchical Clustering and LSA are expensive. On the other hand, there are many low cost algorithms aimed to update the lattice structure obtained from FCA and Fuzzy FCA.
- **Support for classification of new data:** it focuses on the availability of heuristic approaches to classify new incoming resources according to the previous extracted structure. Table 9 highlights that only Hierarchical Clustering does not possess this feature.

	Clustering	Hierarchical Clustering	LSA	FCA &Fuzzy FCA
Hierarchical Conceptualisation		✗		✗
Supporto all'Information Retrial	✗	✗	✗	✗
Costi elevati di aggiornamento della struttura	✗	✗	✗	
Supporto alla classificazione di nuovi dati	✗		✗	✗

Table 9 - Main features of Data Analysis techniques

The comparison results are reported in Table 9. The technique of Formal Concept Analysis was selected and exploited to perform the Concept Data Analysis. In particular, it has been applied to the extension of the fuzzy Formal Concept Analysis (Fuzzy FCA) (Fuzzy FCA) [5], [6], [41].

Furthermore, to optimize the execution of the FFCA was used the methodology Fuzzy C-means [42] in order to reduce the size of the input array. Clustering is, however, an optional component of the Pipe CDA, since it is applied only when the size of the repository of the UGC exceeds a specific value.

6.2.1.4 SEMANTIC TECHNOLOGIES MAPPING

Knowledge Extraction needs to be represented by the standard. In fact, the models previously defined using the standards typical of the Semantic Web, such as: RDF, RDFS, OWL and SKOS. Then, the phase of Semantic Technologies Mapping has the purpose of representing the extracted

³⁴ **Information retrieval** is the activity of obtaining information resources relevant to an information need from a collection of information resources.

knowledge based on this technological layer. Table 10 shows the input and output of the process of Semantic Technology Mapping.

Semantic Technologies Mapping	
INPUT	OUTPUT
<ul style="list-style-type: none"> A concept lattice representing a mathematical modelling of the extracted knowledge 	<ul style="list-style-type: none"> Hierarchical conceptualization of resources' content represented by exploiting semantic technologies, such as: RDF, RDFS, OWL, SKOS, etc.

Table 10 -Input/Output of Semantic Technologies Mapping micro-phase

6.2.2 Evaluation Model for Knowledge Extraction Methodology

In order to demonstrate the validity of the approach described in the steps of the evaluation are constituted by:

- **Using the right metrics to evaluate the extraction of conceptualization.** In particular, the performance evaluation is performed both on the evaluation of the content, ie content analysis products from conceptualization, both on the evaluation of the structure, to understand the value of the generated structure;
- **Description of the experimental results according to the selected dataset.** Is performed a qualitative analysis of the results obtained on sample datasets.

This section presents first the metrics adoperate. Next, show the results collected from the activities of the testing of methodologies used in the Knowledge extraction process.

6.2.2.1 SELECTED METRICS

We are going to concentrate on metrics that measure the Ontology Extraction quality. These will be detailed in the following paragraphs.

6.2.2.1.1 Micro-Average Precision

The metrics described in this paragraph are used to evaluate the performance of the document categorization.

The simplest method of evaluation of the ontology population task described in the previous section is based on Precision, Recall and F-Measure [43], which are the most widely used metrics in Information Extraction evaluations like MUC (Message Understanding Conferences) [44]. These metrics have also a very long-standing tradition in the field of Information Retrieval.

Precision: measures the number of correctly identified items as a percentage of the number of items identified. In the Ontology/Taxonomy Extraction field it is the percentage of concepts retrieved that are correctly extracted and classified.

$$\frac{\text{Correct Retrieved Concepts}}{\text{Retrieved Concepts}}$$

Recall: measures the number of correctly identified items as a percentage of the total number of correct items. In the Ontology/Taxonomy Extraction field it is the percentage of correct concepts that are retrieved.

$$\frac{\text{Correct Retrieved Concepts}}{\text{Correct Concepts}}$$

There must be a trade-off between precision and recall, for a system can easily be made to achieve 100% precision by identifying nothing (and so making no mistakes in what it identifies), or 100% recall by identifying everything (and so not missing anything). For this reason the F-Measure metric was created.

F-Measure: weighted average of Precision and Recall.

$$F\beta = \frac{(1 + \beta^2) * (\text{Precision} * \text{Recall})}{(\beta^2 * \text{Precision} + \text{Recall})}$$

Where β ranges from 0 to 1.

Precision and recall are single-value metrics based on the whole list of documents returned by the system. For systems that return a ranked sequence of documents, it is desirable to also consider the order in which the returned documents are presented. For this reason we will use another metric, the **Micro-Average Precision and Recall**.

By computing a precision and recall at every position in the ranked sequence of documents, one can plot a precision-recall curve, plotting precision $p(r)$ as a function of recall r . Average precision computes the average value of $p(r)$ over the interval from $r = 0$ to $r = 1$:

According to the micro-averaging [45] of recall and precision (at the generic step λ), is defined as follows:

$$Rec_\lambda = \sum_{Q_i} \frac{|R_{Q_i} \cap B_{\lambda, Q_i}|}{|R|} \quad Prec_\lambda = \sum_{Q_i} \frac{|R_{Q_i} \cap B_{\lambda, Q_i}|}{|B_\lambda|},$$

where R_{Q_i} is the set of relevant resources for a given query Q_i , B the set of retrieved resources at the step λ and B_{λ, Q_i} is the set of all relevant resources, retrieved at the step λ , for the query Q_i .

6.2.2.1.2 Computation of metrics: Classes to Clusters Algorithm

In order to compute the previously described metrics, the generated classification and the right one (**gold classification**) needs to be compared. This is difficult because the two classifications use

different labels, so to align them we will use **Classes to Clusters**, an algorithm used by Weka³⁵ to perform cluster evaluation.

This algorithm sees the generated classification sets as clusters and recursively calls itself to assign the gold labels to these clusters in order to minimize the average error.

For each cluster, it finds the classes of the documents contained in the cluster and tries to assign each of these classes to the clusters. All the combinations are tried until no further improvements to the average error are possible.

Classes to Clusters:

0 1 <-- assigned to cluster

242 442 | 3

22 77 | 2

Cluster 0 <-- 2

Cluster 1 <-- 3

Incorrectly clustered instances : 319.0 40.7407 %

Table 11 - Example of classes to clusters algorithms.

In the example in Table 11 Cluster 1 is assigned to Class 3 because it contains the greater number of elements of Class 3, then Cluster 0 is assigned to Class 2.

6.2.2.1.3 Structural Quality Metric

The **Structural Quality Metric** is used to evaluate the structure of the ontology created, i.e. how much the generated taxonomy corresponds to the right one.

Calling T_{gold} the taxonomy of the relevant set and T_{gen} the generated taxonomy, SQM says that when two labels appear in a parent child relationship in T_{gold} , they should appear in a consistent relationship (parent-child or ancestor-descendant) in T_{gen} or vice versa. Based on the above discussion, let

$$pcLinks(T) = \{<a,b> \mid a \text{ is parent of } b \text{ in } T\}$$

$$adLinks(T) = \{<a,b> \mid a \text{ is ancestor of } b \text{ in } T\}$$

³⁵ Weka, Data Mining Software in Java: <http://www.cs.waikato.ac.nz/ml/weka/>

$$adLinks(T) \supseteq pcLinks(T)$$

SQM-P: This measures the precision, i.e., the percentage of parent-child relationships in T_{gen} that appear consistently in T_{gold} .

$$SQM - P = \frac{|pcLinks(T_{gen}) \cap adLink(T_{gold})|}{|pcLink(T_{gen})|}$$

SQM-R: This measures the recall, i.e., the percentage of parent-child relationships in T_{gold} that appear consistently in T_{gen} .

$$SQM - R = \frac{|pcLinks(T_{gold}) \cap adLink(T_{gen})|}{|pcLink(T_{gold})|}$$

6.2.2.1.4 Example Datasets

To assess the methodology of Knowledge Extraction were used sample datasets and ontology / taxonomy resulting from the process of Knowledge Extraction was compared both from the point of view of the structure that the contents with respect to the classification gold sample datasets. In our case they do not have a dataset of documents from the royal domain of Automotive methodology for the extraction of knowledge has been evaluated on a sample. In particular, reference was made to the dataset for example , human classified repository of Open Directory Project (ODP) , also known as Dmoz (from directory.mozilla.org) . ODP is a multilingual open content directory of links on the World Wide Web ' owned by Netscape but it is constructed and maintained by a community of volunteer editors . ODP uses a hierarchical ontology scheme for organizing lists of sites. Ads on similar topics are grouped into categories, which can, in turn, include smaller categories. Figure 30 shows the home page of the site (<http://www.dmoz.org>).

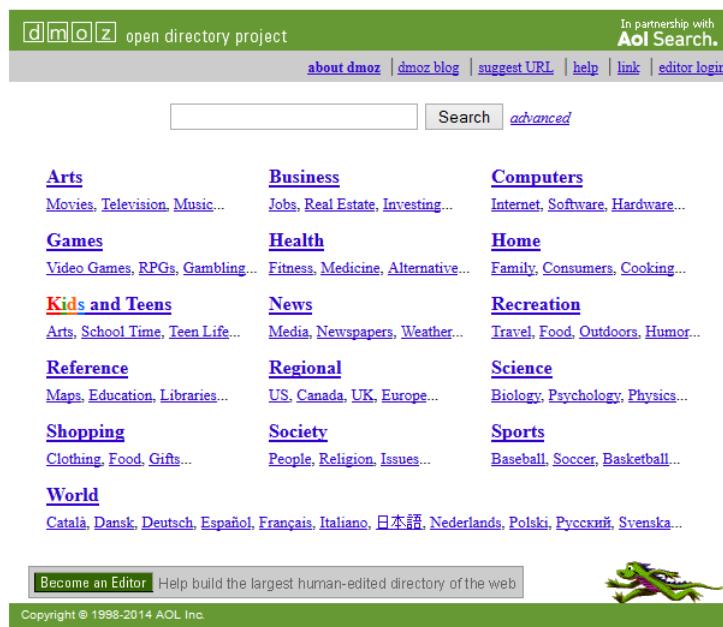


Figure 30 - DMOZ homepage

Top: Computers (92,580)

-
- | | |
|---|--|
| <ul style="list-style-type: none">• Computer Science (1,847)• Hardware (5,026)• Internet (23,552) | <ul style="list-style-type: none">• Security (2,515)• Software (27,169)• Systems (2,514) |
|---|--|
-
- | | |
|--|--|
| <ul style="list-style-type: none">• Algorithms (297)• Artificial Intelligence (1,156)• Artificial Life (202)• Bulletin Board Systems (88)• CAD and CAM (702)• Computer and Technology Law@ (100)• Data Communications (906)• Data Formats (1,697)• Desktop Publishing (109)• E-Books (157)• Emulators (376)• Fonts@ (260)• Games@ (26,894)• Graphics (1,196)• Hacking (161)• Home Automation (64) | <ul style="list-style-type: none">• Human-Computer Interaction (260)• Intranet (36)• MIS@ (364)• Mobile Computing (509)• Multimedia (2,362)• Newsgroups@ (268)• Open Source (689)• Operating Systems@ (8,128)• Parallel Computing (312)• Performance and Capacity (41)• Programming (13,965)• Robotics (700)• Speech Technology (367)• Supercomputing (34)• Usenet (268)• Virtual Reality (306) |
|--|--|
-

Figure 31 - Top category and sub-categories inside DMOZ directory

ODP data is made available through an RDF-like dump that is published on a dedicated download server, where an archive of previous versions is also available. New versions are usually generated weekly.

```
<?xml version="1.0" encoding="UTF-8"?>
<RDF xmlns:r="http://www.w3.org/TR/RDF/" xmlns:d="http://purl.org/dc/elements/1.0/" xmlns="htt
<!-- Generated at 2011-10-09 00:03:26 EST from DMOZ 2.0 --&gt;
&lt;Topic r:id=""&gt;
  &lt;catid&gt;1&lt;/catid&gt;
&lt;/Topic&gt;
&lt;Topic r:id="Top/Arts"&gt;
  &lt;catid&gt;381773&lt;/catid&gt;
&lt;/Topic&gt;
&lt;Topic r:id="Top/Arts/Animation"&gt;
  &lt;catid&gt;423945&lt;/catid&gt;
  &lt;link1 r:resource="http://www.awn.com/"&gt;&lt;/link1&gt;
  &lt;link r:resource="http://animation.about.com/"&gt;&lt;/link&gt;
  &lt;link r:resource="http://www.toonhound.com/"&gt;&lt;/link&gt;
  &lt;link r:resource="http://enculturation.gmu.edu/2_1/pisters.html"&gt;&lt;/link&gt;
  &lt;link r:resource="http://www.digitalmediafx.com/Features/animationhistory.html"&gt;&lt;/link&gt;
  &lt;link r:resource="http://www-viz.tamu.edu/courses/viza615/97spring/pjames/history/main.htm
  &lt;link r:resource="http://www.spark-online.com/august00/media/romano.html"&gt;&lt;/link&gt;
  &lt;link r:resource="http://www.animated-divots.net/"&gt;&lt;/link&gt;
&lt;/Topic&gt;
&lt;ExternalPage about="http://www.awn.com/"&gt;
  &lt;d>Title>Animation World Network</d>Title>
  <d>Description>Provides information resources to the international animation community. Fe
  <priority>1</priority>
  <topic>Top/Arts/Animation</topic>
</ExternalPage>
<ExternalPage about="http://animation.about.com/">
  <d>Title>About.com: Animation Guide</d>Title>
  <d>Description>Keep up with developments in online animation for all skill levels. Downloa
  <topic>Top/Arts/Animation</topic>
</ExternalPage>
<ExternalPage about="http://www.toonhound.com/">
  <d>Title>Toonhound.com</d>Title>
  <d>Description>A comprehensive guide to animation, featuring news, reviews, and interviews with
  <topic>Top/Arts/Animation</topic>
</ExternalPage>
```

Figure 32 - RDF dump structure

Each fragment “<ExternalPage>...</ExternalPage>” contains important information about the web site:

Title: the title of the classified web site (a Dublin Core element)

Description: a short description of the classified web site (a Dublin Core element)

Topic: the path inside the DMOZ directory (it could be thought as taxonomy).

We developed a tool that exploiting DMOZ's RDF dumps, produces a database catalogue (in a csv file format that can be read and stored in any database or index) and a file system hierarchy that mimics the DMOZ directory categories and sub-categories.

In practice, each category is a folder into the file system and we use subfolders to represent sub-category structures. So, for example, the DMOZ path "Top/Arts/Anime" will produce a nested folder hierarchy "C:\dmoz_dataset\Top\Arts\Anime" (in windows).

6.2.2.1.5 Experimental Results

This section provides the results of the methodology described for the knowledge extraction on some portions of the example dataset ODP valued according to the model described above. The test was run on a small set of 525 documents belonging to the categories shown in Table 12.

Document	Summary
Top/Business/E-Commerce/Developers	90
Top/Business/E-Commerce/Customer_Relationship_Management	64
Top/Business/E-Commerce/Consulting/Software_Developers	38
Top/Business/E-Commerce/Developers/Database_Applications	30
Top/Business/E-Commerce/Consulting	207
Top/Business/E-Commerce/Education_and_Training/Courses	23
Top/Business/E-Commerce/Education_and_Training/Centers	20
Top/Business/E-Commerce/Developers/B2B	15
Top/Business/E-Commerce/Conferences	11
Top/Business/E-Commerce/Developers/Web_Applications	9
Top/Business/E-Commerce/Developers/Managed_Sites	6
Top/Business/E-Commerce/Directories	4
Top/Business/E-Commerce/Associations	4
Top/Business/E-Commerce/Associations/International_Associations	3
Top/Business/E-Commerce/Developers/Tools	2
Top/Business/E-Commerce/Developers/Sectors	2
Top/Business/E-Commerce/Education_and_Training/Degree_Programs	2
Top/Business/E-Commerce/Employment/Resumes/Individual_Resumes	1
Top/Business/E-Commerce/Employment/Executive_Search	1
Top/Business/E-Commerce/Developers/Sectors/Transport-logistics	1

Table 12 - Dataset used to evaluate methodology of Knowledge Extraction.

In particular, Figure 33 highlights better results obtained for specific category.

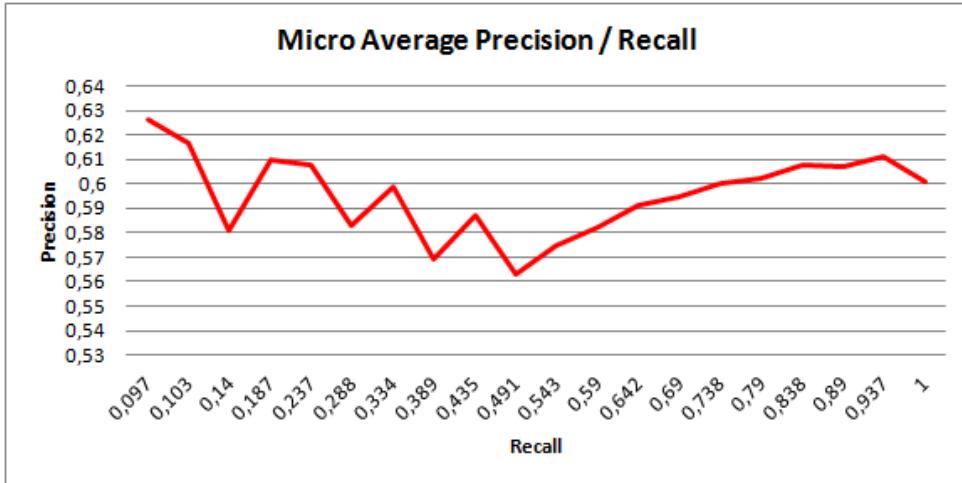


Figure 33 - Micro-Average Precision Recall.

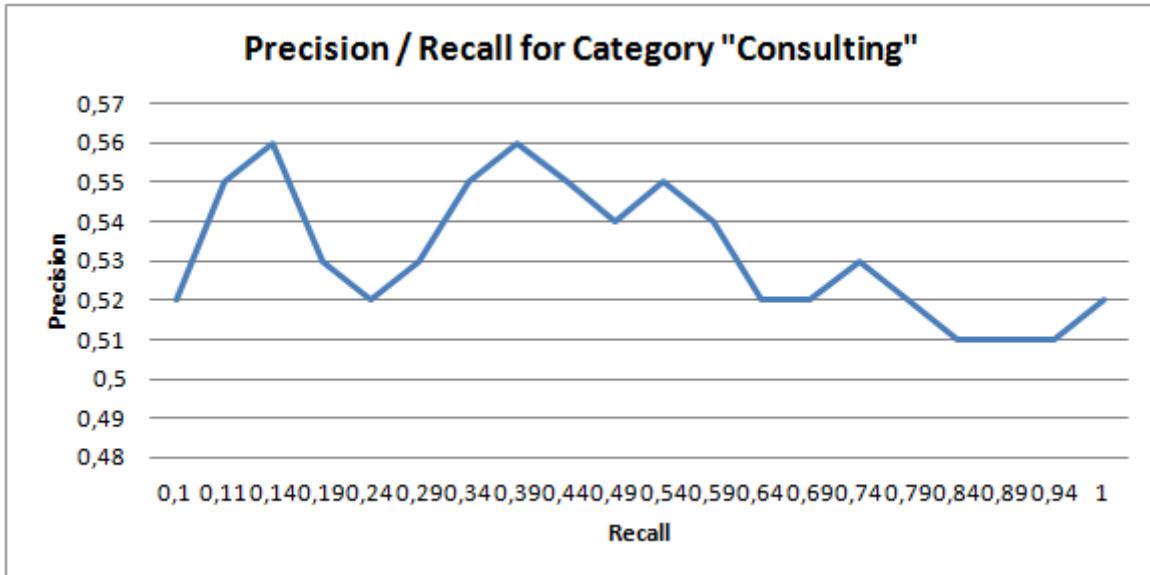


Figure 34: Precision/Recall for specific category (i.e., Consulting).

These categories have been used as “Correct Values” while computing the performance metrics. By performing the methodology for knowledge extraction evaluation of structural metrics have been obtained. The results are shown in Table 13.

SQM-P	SQM-R
0.88	0.83

Table 13 - Structure Quality Metric: Precision and Recall (see Section 4.2.2.1.3 for definitions).

The performances are obtained on the subset of ODP dataset. In particular, only a brief description (less than 20 words) has been exploited in the analysis process.

6.3 Methodology for Ontology and Instance Matching

The ontology matching is a methodology that enables semantic interoperability between systems that use different ontologies. This general objective has been framed according to the Automotive project in order to address specific goals. In this chapter, we refer to the Ontology Matching as an integrated approach to recover the correspondence is both ontological instance. In the literature [46] this objective falls within the areas of Ontology and Instance Matching (IOM). Intuitively, the Ontology Matching regards the matching process is that the correspondence between the ontological concepts. According the Ontology Matching [47], can be performed on:

- **Schema-based Matching Level**, where schema elements are the basis to find correspondences between concepts.
- **Instance-based³⁶ Matching Level**, where instances are the basis to find correspondences between concepts.

Instead, Instance Matching concerns the matching process that finds correspondences between two descriptions and retrieves same (or similar) *entities* of a specific domain [48], independently from its representation [47].

Under the project Automotive, it is useful to use both techniques . In fact, the method of ontology matching is part of various processes. Some of these processes take advantage of the matching methodology to support the preparation and maintenance of domain ontologies SKOS specif -based, while others exploit the matching methodology to assess the correspondence between instances of the models, in order to allow the " provisioning" of specific suggestions . The use of different matching methods will be detailed in the following sections through the definition of a strategy Ontology and Instance Matching. In order to support the creation and maintenance of the knowledge base will be used a semi-automatic process, IOM is the start-up and during the life cycle of the project Automotive to align the results obtained by the process of extraction of knowledge (i.e., Unsupervised conceptualization) with respect to specific ontologies tomorrow SKOS-based. These ontologies are essentially instances of SKOS schemes on which the IOM methodology must identify the matches by exploiting the subsumption relationships expressed through properties such as Broader and narrower in SKOS. In the literature, the Ontology Matching mainly exploits the relations of subsumption to align the concepts. In this case, you must define an integrated approach and Ontology Instance Matching.

Regardless of the strategy used for matching, the input of the instance and Ontology Matching is characterized by the entities (concepts or individuals) ontological on which the strategy of matching can be configured and executed to meet the needs of different scenarios In order to provide specific suggestions, the IOM methodology is used to evaluate the matching between instances of the models, that are represented in the same pattern. For example, consider the scenario where you

³⁶ Let us note that **Ontology Matching Instance-based** is not Instance Matching, in fact, the former aligns ontological concepts exploiting also their instances, and instead, the latter finds correspondences among heterogeneous instances.

want to suggest Best Practice similar on the basis of the industrial process or that describe the skills that are required to describe the process that the Best Practice wishes to outline. Specifically inputs are:

- Source Ontology/Population: according to the scenarios described above it may refer to Unsupervised Conceptualization, required competences, selected task, etc.;
- Target Ontology/Population: according to the scenarios described above it may refer to Organization Ontologies.

Final results of Ontology and Instance Matching methodology is the set of weighted (i.e., membership in the range [0, 1]) correspondences between source and target ontologies and population. More formally, let e_s and e_t be entities of Source and Target Ontology/Population, respectively, methodology for Ontology and Instance Matching carries out triples, such as:

$$\langle e_s, e_t, n_{st} \rangle,$$

where n_{st} is a mapping degree of trust (confidence) that is a measure of the trust in the fact that the mapping is appropriate.

Following subsections are organised as follows. First of all, the definition of the Ontology and Instance Matching process will be argued. Subsequently, according to the process definition two matching strategies will be described in order to address the aims introduced above. Finally, evaluation model and results will be given.

6.3.1 Process Definition

This section describes the process of Ontology and Instance Matching previously introduced to support the methodology of Knowledge Classification. The process is based on the strategies defined in a number of existing tools and the Ontology Instance Matching, as RIMOM [49] ASMOV [50], , and Link Discovery in Linked Data used to align instances from different sources, such as Silk [51] and LIMES [52]. The IOM tools are useful for their combinatorial strategies, however, the tools of Link Discovery are useful for their configuration strategies. In particular, the process of IOM has been customized on taken into domain under consideration. Table 14 shows the input and output of the process of Ontology and Instance Matching.

Ontology and Instance Matching Processing Pipe	
INPUT	OUTPUT
<ul style="list-style-type: none"> • Source Ontology/Population and Target Ontology/Population 	<ul style="list-style-type: none"> • A list of matching between pair of instances of input ontologies in form of triples: $\langle i_s, i_t, n_{st} \rangle$ where $i_s \in O_s$, $i_t \in O_t$, n_{st} = similarity value.

Table 14 - illustrates the overall of Ontology and Instance Matching process.

Figure 35 illustrates the overall process of the methodology and Ontology Instance Matching

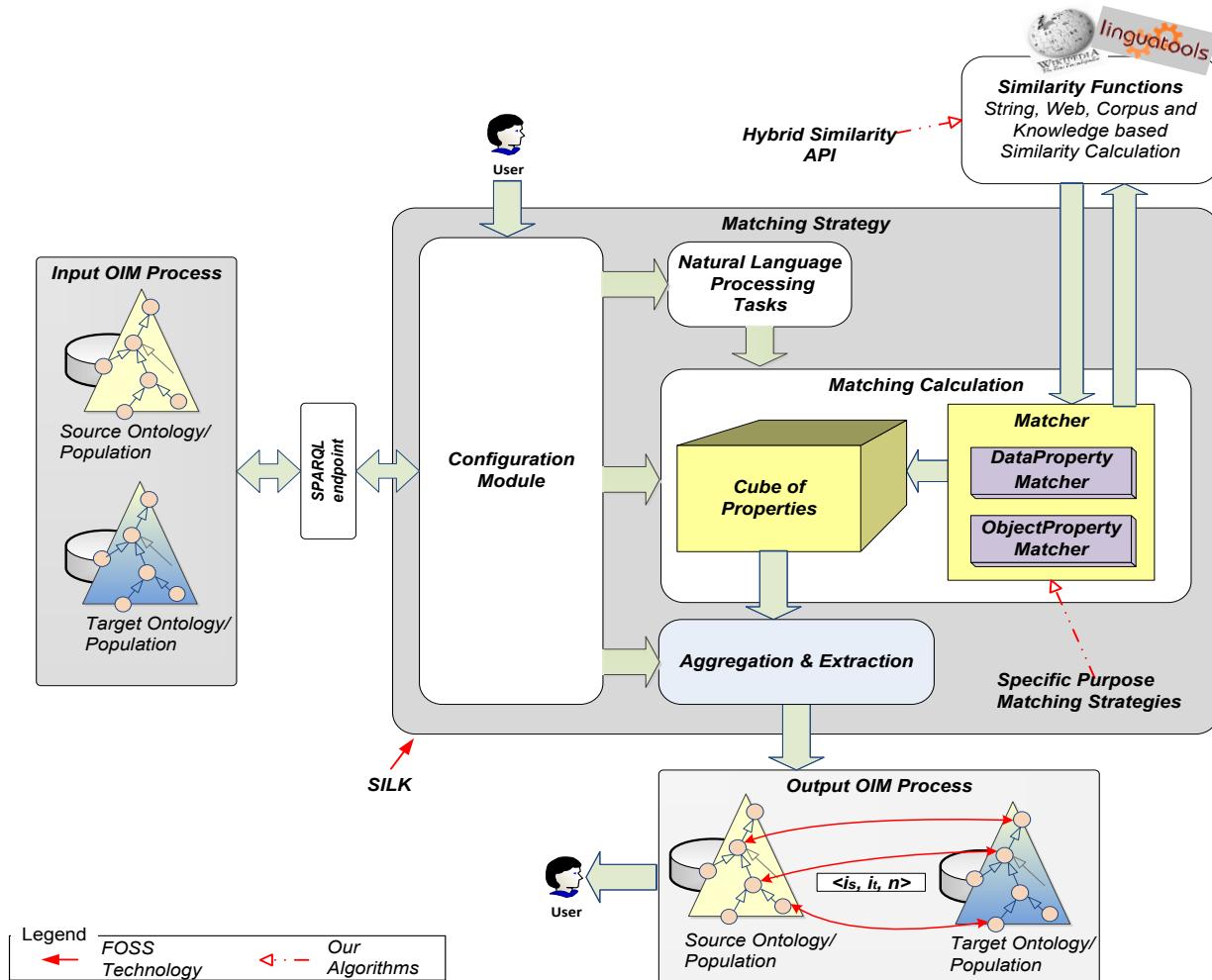


Figure 35 - Ontology and Instance Matching Process

Matching Strategy module is the core of IOM process. It consists of the following modules:

- **Configuration Module:** It is used by an experienced user to configure the whole process of IOM,
- **Natural Language Process Task:** Pipe provides several task of NLP,
- **Matching Calculation:** provides several matcher,
- **Aggregation and Extraction:** aggregates in a single similarity value of all the results returned by the matcher module Matching Calculation.

6.3.1.1 CONFIGURATION MODULE

The Configuration Module describes the entire process of IOM. It is used by an expert user is able to configure all the necessary components. Table 15 shows the input and the output of the Configuration Module.

Ontology and Instance Matching Processing Pipe	
INPUT	OUTPUT

<ul style="list-style-type: none"> • Source Ontology/Population and Target Ontology/Population 	<ul style="list-style-type: none"> • A list of matching between pair of instances of input ontologies in form of triples: $\langle i_s, i_t, n_{st} \rangle$ <p>where $i_s \in O_s$, $i_t \in O_t$, n_{st} = similarity value.</p>
---	---

Table 15 - Input/Output of the Configuration Module

In particular, configurable components are the following:

- **Classes**, type of instances, of Source and Target Ontology respectively, which are involved in the of IOM process (e.g. SKOS Concept).
- **Object Property**, name of one or more Object Property of the entities used them in phase of the similarity evaluation.
- **Date Property**, name of one or more Property Data of the entities used in the evaluation phase of the similarity (e.g., label, description, etc.).
- **NPL**, one or more activities NPL tasks to apply on Object Properties and Data Properties (e.g., lowercase, removing spaces, Stemming , Tokenize , etc.).
- **Similarity Function**, function of similarity to be applied on pairs of Object Properties and / or Data Properties belonging to source and target, ontology classes aforesaid. The function of similarity can be of type String -based (e.g., Levenshtein or Jaccard) , type of Web-based (e.g., Wikipedia Link Measure - WLM) , Knowledge-based (for example , Wu & Palmer, Leacock and Chodorow), Corpus-based and type (e.g., Record, ESA) or Hybrid.
- **Aggregate Function**, is a type of aggregation policy that combines into a single value the results of similarity is that the Object Properties and Data Properties aforesaid (e.g., media, Euclidean distance).

6.3.1.2 NATURAL LANGUAGE PROCESS TASK

The process begins by performing some tasks of Natural Language Processing. The tasks are:

- **Stemming**: it applies on a string.
- **Tokenization**: it splits the string or text into tokens (symbol).
- **Lower Case**: Converts a string to lowercase.
- **Upper Case**: Converts a string to uppercase.
- **Remove Blank**: removes spaces from a string.
- **Stop Words Removal**: Removes stop words from a text.
- **Strip Prefix**: Removes the prefix in a string.

Table 16 describes the input and output of the NLP Tasks module

Natural Language Processing Tasks	
INPUT	OUTPUT
<ul style="list-style-type: none"> Configuration XML-based file produced by Configuration Module. 	<ul style="list-style-type: none"> Results of NLP tasks applied on Data Property and Object Property values according to specifics listed in configuration file.

Table 16 - Input/output of NLP Tasks

6.3.1.3 MATCHING CALCULATION

Matching Calculation provides a number of matcher to apply to the properties of instances of the classes of the ontologies to align. On each pair of properties can be invoked a metric of similarity or a combination thereof. Table 17 shows the input and the output of module Matching Calculation.

Matching Calculation	
INPUT	OUTPUT
<ul style="list-style-type: none"> Configuration XML-based file (returned by Configuration Module). Source and Target Ontology Instances elaborated by NLP tasks according to configuration file. 	<ul style="list-style-type: none"> A Pair-wise properties similarity adjacency cube in which each cell contains similarity values for all pair of properties belonging to instances specified in configuration file.

Table 17 - Input /Output of Matching Calculation

6.3.1.4 SIMILARITY FUNCTIONS

The Similarity Functions module calculates similarity values between entities of input ontologies on demand by matchers of Matching Calculation Module. Table 18 shows details about similarity metrics available in Similarity Functions Module.

Similarity Functions	
INPUT	OUTPUT
<ul style="list-style-type: none"> Values of a pairs of properties ($prop_s, prop_t$) belonging to class C_s in Source Ontology and class C_t in Target Ontology, respectively. 	<ul style="list-style-type: none"> A similarity value n_{st} between values of pair ($prop_s, prop_t$).

Table 18 - Input/Output of Similarity Functions

6.3.1.5 AGGREGATION AND EXTRACTION

The aggregation module and extraction combines the results of individual matcher in order to improve the quality of the matching and return a list of matching between the instances of ontologies provided in the input. There are different aggregation criteria such as, for example: Average, Maximum, Euclidean distance, and so on.

6.3.2 Matching Strategy: Alignment between different SKOS instances

The relationship between the form of Ontology and Instance Matching and the form of Knowledge Classification, highlights the need to align the concepts of Unsupervised conceptualisation (genarata the form of Knowledge Extraction) with the concepts of the SKOS-based vocabulary.

This strategy is defined by an appropriate configuration of the previously defined Configuration Module in the process of IOM. The following sections describe in detail how to configure the matching process between SKOS taxonomies and which features of similarity and aggregation apply.

6.3.2.1 CONFIGURATION MODULE

The following are the configurations of the components set forth in the Configuration Module described above:

- ***Classes, SKOS:Concept.***
- ***Object Property,*** configured on relations of Taxonomic Information as ***SKOS:narrower*** and ***SKOS:broader*** relations.
- ***Data Property,*** configured on properties of Descriptive Information as ***SKOS:prefLabel***, ***SKOS:altLabel*** and ***SKOS:hiddenLabel***.
- ***NLP*** configured on one or more NLP tasks between:
 - Lower case, Remove Blanks, Stemming, Tokenize, etc.
- ***Similarity Functions*** configured on:
 - hybrid similarity metric that takes into account a sum weighted of more types of similarity metrics for instances of Data Properties ***SKOS:prefLabel***, ***SKOS:altLabel*** and ***SKOS:hiddenLabel***,
 - a generalization of knowledge-based metrics for instances of ***SKOS:Concept*** in order to exploit relations of specialization (***SKOS:broader***) and generalization (***SKOS:narrower***) defined in the SKOS standard (see Similarity Functions for this Matching Strategy).
- ***Aggregate Function*** configured on an aggregate methods between
 - Average, Maximum, Euclidean Distance or Geometric mean.

6.3.2.2 NATURAL LANGUAGE PROCESS TASKS

According to defined configuration, NLP Tasks as ***Lower Case***, ***Stop Words Removal*** and ***Strip Prefix*** can be applied on Data Properties ***SKOS:prefLabel***, ***SKOS:altLabel*** and ***SKOS:hiddenLabel***.

6.3.2.3 MATCHING CALCULATION

The goal is to exploit relations of specialization and generalization, existing between instances of SKOS concepts, by means of *SKOS:narrower* and *SKOS:broader* relations, respectively, defined in the SKOS standard.

We define a similarity measure, between instances of SKOS concepts, that uses both *descriptive information* (e.g., label) or *taxonomic information expressed by SKOS:narrower a SKOS:broader* relations.

Descriptive information are addressed by SKOS Data Properties, that is, all properties having a *RDF literal* as value. For class *SKOS:Concept* these properties are:

- ***skos:prefLabel***, the most common term used for a specific concept.
- ***skos:altLabel***, most common synonymous used for a specific concept.
- ***skos:hiddenLabel***, a label normally hidden but used by application having functionality of search text-based. This label is typically used to list incorrect possibility of a term (e.g., electriciti, electricity, etc.).

Taxonomic information is addressed by SKOS Object Property, that is, all properties describing semantic information to classify and order instances of SKOS Concepts. These properties are:

- ***skos:broader***, to establish the meaning of a concept as more general than another.
- ***skos:narrower***, to establish the meaning of a concept as more specific than another.

Therefore, let two instances of SKOS concept, $i_s \in O_s$ Source Ontology and $i_t \in O_t$ Target Ontology, we define:

- $sim_{Node}(i_s, i_t)$ as similarity function on Data Property:

$$sim_{Node}(i_s, i_t) = w_{prefLabel} * sim_{prefLabel}(i_s, i_t) + w_{altLabel} * sim_{altLabel}(i_s, i_t) + w_{hiddenLabel} * sim_{hiddenLabel}(i_s, i_t)$$

with

$$w_{prefLabel} + w_{altLabel} + w_{hiddenLabel} = 1$$

$w_{prefLabel}$, $w_{altLabel}$ and $w_{hiddenLabel}$ are weights used to highlight the importance of the *prefLabel*, *altLabel* and *HiddenLabel* proprieties on the similarity function $sim_{Node}(i_s, i_t)$.

- $sim_{Taxonomy}(i_s, i_t)$ as similarity function on *skos:Concept* that uses the Object Properties *SKOS:narrower* and *SKOS:broader*:

$$sim_{Taxonomy}(i_s, i_t)$$

- $SKOSSim(i_s, i_t)$ as similarity function as follow:

$$SKOSSim(i_s, i_t) = w_{Node} * sim_{Node}(i_s, i_t) + w_{Taxonomy} * sim_{Taxonomy}(i_s, i_t)$$

With $w_{Node} + w_{Taxonomy} = 1$, w_{Node} and $w_{Taxonomy}$ are weights used to highlight the importance of *Data Property* and *Object Property* on the similarity function $SKOSSim(i_s, i_t)$.

6.3.2.4 SIMILARITY FUNCTION

This section explains how to calculate the similarity described previously in the phase of Matching Calculation.

Sim_{Node} derive da $sim_{prefLabel}$, $sim_{altLabel}$ e $sim_{hiddenLabel}$. Each of them can be computed by means of similarity metrics String-based (e.g., Jaccard or Levenshtein Distance [53]) implemented in SimMetrics³⁷ library, Corpus -based tool implemented in DISCO³⁸, or Web-based applications, such as WLM implemented in WikipediaMiner³⁹.

$sim_{Taxonomy}(i_s, i_t)$ can be calculated by means of a generalization of the knowledge-based metrics on two different ontologies SKOS.

In order to compute Similarity strategy to align different SKOS instances belonging to different SKOS ontologies we have generalized knowledge-based metrics. Calculation of traditional Similarity Knowledge-based metrics typically depends on path length between two concepts in the same taxonomy.

The metric Wu & Palmer measure the similarity between two concepts of a taxonomy combining the depth of the two concepts in the taxonomy with the depth of the Least Common Subsumer - LCS (or Last Common Ancestor - LCA) , which is the nearest common ancestor between the two concepts of taxonomy (modeled on a tree structure).

Figure 24 shows the LCS between two concepts in a taxonomy. The metric of Wu & Palmer is generalized to be used in two different taxonomies. The knowledge used to evaluate the metric Wu & Palmer is defined by ontologies SKOS that must be aligned and not from an external knowledge (such as WordNet) as usually happens in the literature. The intent is to follow a pattern matching - based approach to align disjoint SKOS taxonomies.

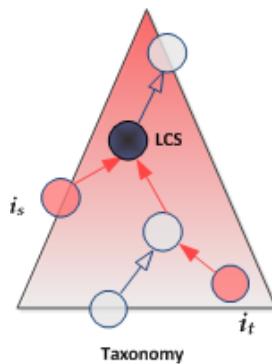


Figure 36 - LCS between two concepts in a Taxonomy

³⁷ SimMetrics library: http://sourceforge.net/project/memberlist.php?group_id=123463

³⁸ DISCO tool: http://www.linguatools.de/disco/disco_en.html

³⁹ WikipediaMiner: <http://wikipedia-miner.cms.waikato.ac.nz/>

We generalize **LCS**, between nodes of the same taxonomy, in **Most Similar LCS – MSLCS**, between two nodes of two different taxonomies. Specifically, let A_s the Ancestor Set of i_s in O_s and A_t the Ancestor Set of i_t in O_t , we define **MSLCS**, between two concepts $i_s \in O_s$ and $i_t \in O_t$, the pair $MSLCS_s, MSLCS_t$ where:

- $MSLCS_s \in A_s$,
- $MSLCS_t \in A_t$,
- $sim(MSLCS_s, MSLCS_t) = \max(sim(MSLCS_k, MSLCS_j)) \forall k \in A_s \text{ and } j \in A_t$

Figure 37 shows the MSLCS between two concepts i_s and i_t in two SKOS taxonomy.

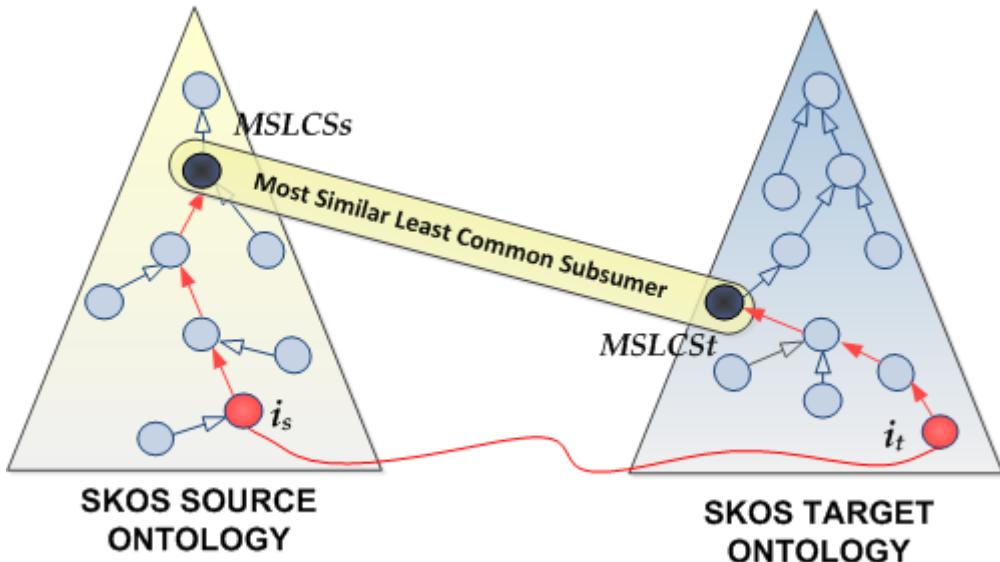


Figure 37: MSLCS between two concepts in two SKOS Taxonomies

MSLCS links together two taxonomies so that it is possible compute similarity by means of traditional Knowledge-based metrics.

6.3.2.5 AGGREGATION AND EXTRACTION

Different types of aggregation criteria can be chosen as: Average, Maximum, and so on. The best criterion of aggregation can be identified during the evaluation of the results and can be changed due to a change in the configuration file, in order to improve the results.

6.3.3 Matching strategy: Alignment between instances in the same schema

To prove the validity of the process of Ontology and Instance Matching described in Section 6.3.1, comes the need for a formal assessment of the solution proposed. The evaluation process is structured in the following way:

1. Selection of the most suitable parameters for evaluating the matching process.
2. Selection the dataset for testing.
3. Creation a Matching Reference for the selected dataset.
4. Evaluation of the results on the selected dataset.

Figure 38 shows an evaluation model for a matching process, as described in [54]. The matching process takes as input two ontologies, \mathbf{o} and \mathbf{o}' , and generates a matching M depending on the parameters (for example, the threshold) and resources (for example , the knowledge base and the domain ontology) . The evaluation process takes as input the matching M , previously generated , and a Matching Reference RM and assess the results according to the metric (e.g. Micro-Average Precision) evaluation choice.

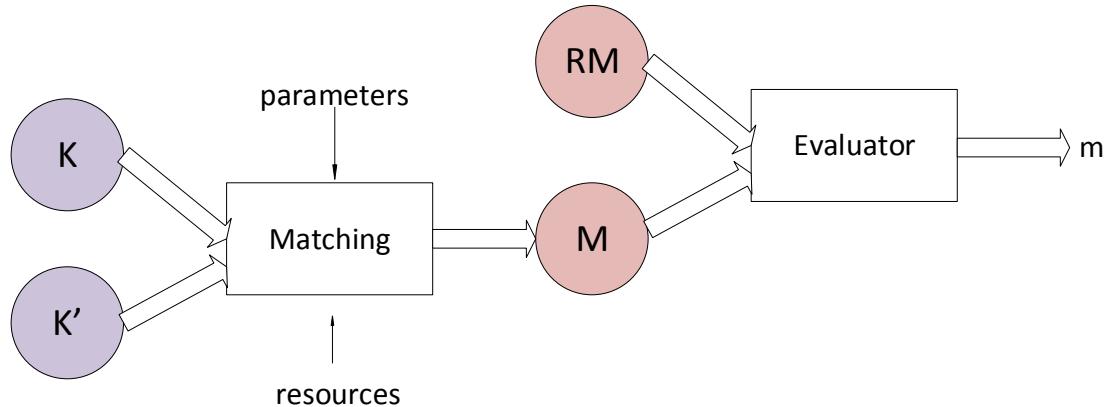


Figure 38 - Evaluation model for a matching process.

6.3.3.1 SELECTED METRICS

The metrics used in the literature for the evaluation of processes and Ontology Instance Matching are the Precision and Recall [55], [56], [50]. These metrics provide measures that can measure the level of compliance of the generated matching M with respect to the Reference Match MRI. These metrics have already been defined in Section 6.2.2.1.1, but in the context Ontology and Instance Matching they obtain the following meanings:

Precision: measure the number of matches correctly found as a percentage of the number of matches found. Both $\text{correctly_found_matche \#} = \text{number of matches found}$ and $\text{correctly_then_all_found_matche \#} = \text{number of all matches found}$ [57].

$$\text{Precision} = \frac{\#\text{correctly_found_matche}}{\#\text{all_found_matche}}$$

According to the definitions of matching M generated and Reference Matching RM , previously mentioned, the Precision can be expressed as [55]:

$$\text{Precision} = \frac{|RM \cap M|}{|M|}$$

Recall: measures the number of correctly found matches as a percentage of the total number of correct matches. Both $\text{correctly_found_matche \#} = \text{number of correctly found correspondences}$ and $\text{all_correct_match \#} = \text{number of all possible correct matches between source and target ontology}$ [57].

$$Recall = \frac{\#correctly_found_matche}{\#all_correct_matche}$$

Similarly to the Precision and according to the definition of matching generated M and Reference Matching RM, previously mentioned, the Recall can be expressed as [55]:

$$Recall = \frac{|\ RM \cap M |}{| RM |}$$

Precision and Recall measures are the most widespread and commonly used. Usually prefers to use only one metric of evaluation, however, the sun Precision and Recall are not sufficient to make the feedback you want. Indeed, the Recall can easily be maximized by returning all possible matches at the expense of a poor Precision. Similarly, a high Precision can be achieved at the expense of poor Recall by returning only a few matches (correct). For this purpose, as done for the evaluation metrics Knowledge Extraction , a curve is plotted Precision - Recall in which is a function of the Precision Recall, $p(r)$ in the interval $[0,1]$. Therefore, the measure of Micro -Average Precision and Recall, described in Section 6.2.2.1.1, is redefined to evaluate the results of the process of Ontology and Instance Matching.

According to [45], the micro-averaging of Precision and Recall at the generic step λ defined in Section 6.2.2.1.1 becomes:

$$Prec_\lambda = \sum_{C_i} \frac{|RM_{C_i} \cap M_{\lambda, C_i}|}{|M_\lambda|} \quad Rec_\lambda = \sum_{C_i} \frac{|RM_{C_i} \cap M_{\lambda, C_i}|}{|RM_\lambda|},$$

where RM_{C_i} is the matching set of a source ontology concept C_i in the Reference Matching RM , M_{λ, C_i} is the matching set at the step λ for the concept C_i , M_λ is the matching set at the step λ and RM_λ , is the Reference Matching set at the step λ .

6.3.3.2 SELECTED DATASETS

The choice of dataset in this section is intended to evaluate the matching process described by Matching Strategy (Section 6.3.2). This strategy is the need to align instances of SKOS concepts, then, you must select taxonomies as the test dataset. The main weakness of the evaluation process, described above, is related to the problem of acquiring the Reference Match between pairs of elements belonging to the two data sets provided as input. Despite the diverse range of datasets and Matching Reference available on the website of the initiative OAEI [55],, a project aimed at evaluating tools and Ontology Instance Matching, it was not possible to use taxonomies and related Reference Matching suited to the evaluation of previously Matching Strategy defined (Section 6.3.2).

In fact, although some of the datasets published on the OAEI⁴⁰ respond to the needs of the Matching Strategy matching, do not have the corresponding Reference Matching (because they are no longer available on the site). The problem of the acquisition of the Reference Match in the evaluation process is a known fact in the literature [58]. For these reasons, the Reference Matching is often created manually on small datasets. The example dataset chosen for the experimentation of the method of matching carried out under the project Aristotle, are two SKOS taxonomies derived from a repository classified by experienced users: Open Directory Project - ODP. The two taxonomies are identified by the nomiDmoz_Subset1 and Dmoz_Subset2. The first consists of fourteen SKOS concepts selected from the sub-categories in the Business category, the second consists of 50 SKOS concepts selected from the sub-categories of Business and Science. Figure 39 and Figure 40 shows the two taxonomies extracted.

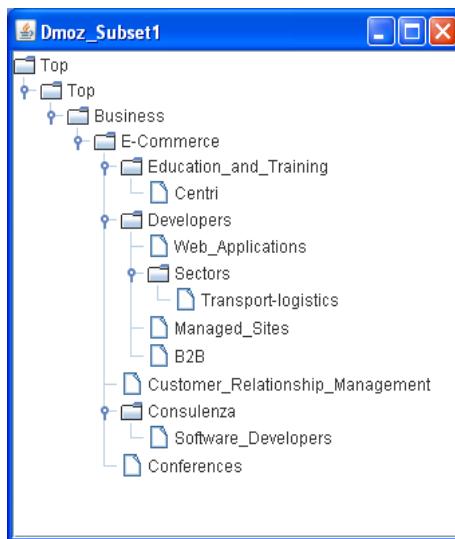


Figure 39 - Dmoz_Subset1

⁴⁰ At <http://oaei.ontologymatching.org/2009/> provides three large SKOS subject heading lists for libraries are available formalised with SKOS vocabulary. At <http://oaei.ontologymatching.org/2010/> provides real world taxonomies extracted from Google, Yahoo and Looksmart web directories.

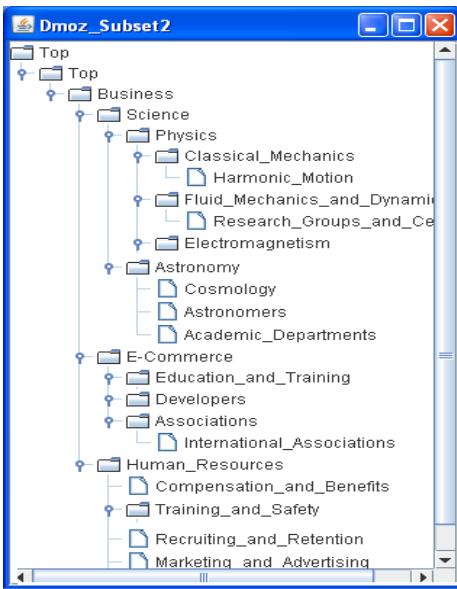


Figure 40 - Dmoz_Subset2

6.3.3.3 REFERENCE MATCHING

According to the evaluation model for the defined by IOM process and after the choice of dataset for Matching Strategy, it is necessary to create a Reference Match containing all possible correspondences between the source and target ontologies. The Reference Match was created through the semi-automatic process shown in Figure 41.

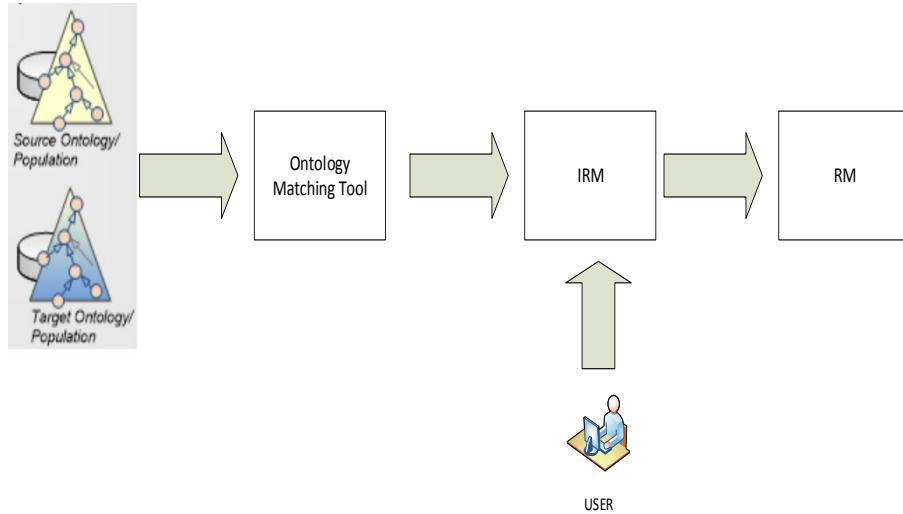


Figure 41 - Semi automatic process to generate a Reference Matching

The Initial Reference Matching (IRM), which is generated by a tool Ontology Matching available in the literature, was supervised by experienced users to validate the set of correct matches missing matching generated by adding and / or removing incorrect correspondences, and thus provide a

Reference Matching RM. The cardinality of the matching is 1: N, that is, for each source ontology SKOS concept are aligned with one or more target ontology concepts SKOS. Table 19 shows Reference Matching obtained by the semi-automatic process described in Figure 41 on Dmoz_subset1 and Dmoz_Subset2 taxonomies.

Source Ontology Concepts	Target Ontology Concepts
B2B	Outsourcing, Consulting, Services, Industrial, Journals, Newsletters, Strategy, E-Commerce, Marketplaces Marketing_and_Advertising
Business	Business, E-Commerce
Centers	Consulting, Research_Groups_and_Centers, Courses, International_Associations, Recruiting_and_Retention, Education_and_Training
Conferences	Consulting, Tools, Courses, Services Journals, Newsletters
Consulting	Associations, Strategy, Services, Consulting
Customer_Relationship_Management	Consulting, Database_Applications, Human_Resources, Employment
Developers	E-Commerce, Associations, Developers, Employment, Marketplaces Strategy, Technology, Tools, Research_groups_and_Centers, Database_Applications
E-Commerce	E-Commerce, Human_Resources, Business, Developers, Marketplaces
Education_and_Training	Associations, Education_and_Training, Employment, Strategy, Recruiting_and_Retention, Courses
Managed_sites	Database_Applications, Degree_Programs, Courses, Executive_Search, Industrial, Services, Journals Consulting, Tools, Strategy
Sectors	Associations, Developers, Strategy, Services,
Software_Developers	Database_Applications, Tools, Developers, Strategy, Research_Groups_and_Centers, Technology
Transport-Logistics	Associations, Tools, Industrial
Web_Applications	Consulting, Tools, Database_Applications, Technology, Strategy, Services

Table 19 - Reference Matching on Dmoz subsets

6.3.3.4 EXPERIMENTAL RESULTS

The results presented provide the values of Micro Average Precision and Recall obtained by Matching Strategy implemented and integrated in Silk (Link Discovery tool). In order to generate a matching 1: N, Silk has been configured on 1 as the threshold value.

shows the results of Micro Average Precision and Recall obtained by the execution of the Match Strategy on the test dataset.

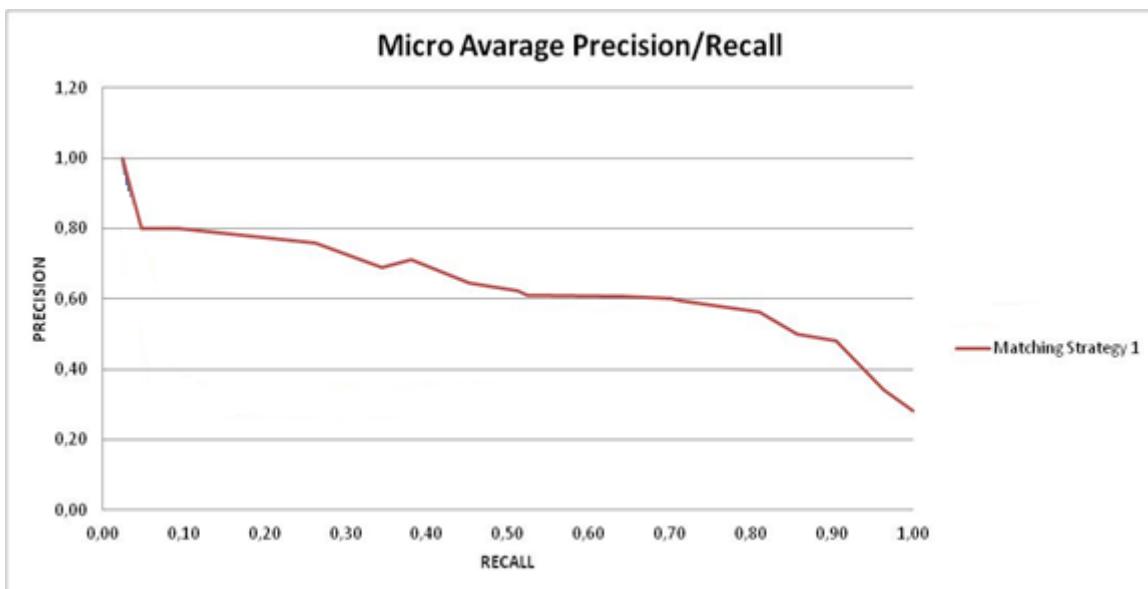


Figure 42 - Results of Micro Average Precision and Recall of Matching Strategy on test datasets.

Part III: Case Studies

7. Definition of models of knowledge in the domain Automotive

The objective of this chapter is to define a knowledge model for the representation of general purpose resources and ontological models of the domain you want to use in the research work. In particular, the approach will be defined which has the aim to pursue in order to correlate the available resources with the knowledge of the domain to be prepared. In conclusion will be described an integrated view of ontological models provided, their main relationships and roles of each of them.

The chapter is aimed at achieving the following objectives macro:

- ✓ Identification of the types of resources managed by the final system (characterizing the domain of interest) that are represented by means of ontologies ;
- ✓ Identification of the ontological models for representing knowledge in the business domain of reference, i.e., controlled vocabularies, taxonomies, etc..;
- ✓ Identification of the main metadata and classification criteria correlating the types of resources previously identified with the ontological domain models that characterize the business environment reference.

7.1 Modelling

The knowledge management needs to be modeled, structured and interconnected through ontological representation. Through this you can get a formal conceptualization of a particular domain, with the additional feature of being shared by the members of the organization. Among the various processes definition of corporate knowledge and its classification using ontological structures in the literature, has aroused particular interest is the Knowledge Meta Process [59], both for its organic, both for the feature to be integrated with the traditional Software Engineering processes.

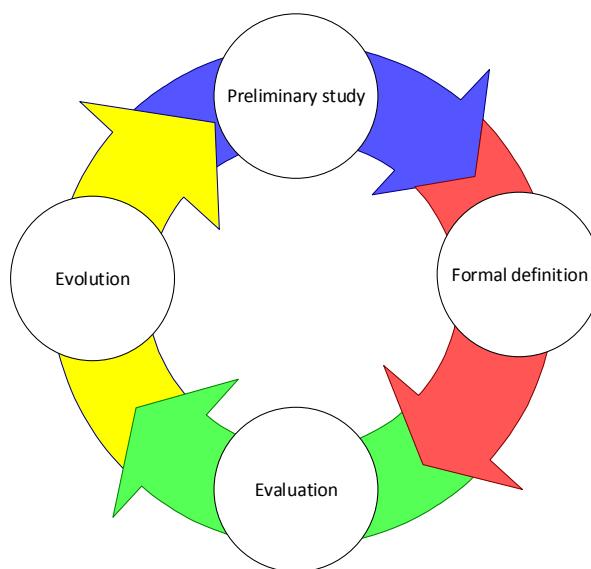


Figure 43 - Knowledge Meta Process

Figure 1 shown the process that consists of four main steps, each of which is characterized by the presence of several sub-steps:

- **Preliminary study.** At this stage begins the ontological description of the aspects characterizing the application domain, starting from requirements analysis. In particular, this description focuses in the identification of key concepts and the relationships between them intervening. The result of this step is a rough drawing of the ontology, which allows to identify the concepts and relationships between them, without the need for completeness and formalities. At the end of this phase it is possible to have an initial assessment, rather than empirical, the actual validity of the model. It is obvious that inconsistencies or errors cannot be detected already, to make up for that eventuality, the whole process is cyclical and will eventually be possible to return at this stage to enter, modify or delete concepts and relationships.
- **Formal definition.** Once you have obtained from the previous step a semi-formal model, we can distinguish two major approaches for its refinement: top- down and bottom-up. Similarly to what happens in Software classical, the top-down approach is to start with a general conceptualization that is gradually refined by adapting to the needs that are taken into account. In the literature it is possible to find ontologies defined at different levels of abstraction. According to this level it is possible to define three levels of membership for ontologies [60]:

The high-level ontologies are independent of the application domain and provide a base of knowledge from which may be derived from different domain ontologies. The concepts expressed are intentionally general so as to be adaptable to a wide range of applications as possible. In this sense, the concepts defined are somewhat abstract, so much so that we are talking about meta-concepts [61]. Examples of this type of ontologies are the Suggested Upper Merged Ontology (SUMO⁴¹), the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE⁴²) and Cyc⁴³. Ontologies serve as a reference interconnection between abstract concepts, defined by the top-level ontologies, and those specific to the business domain. In other words, these ontologies are useful to realize elements common to specific business domains, representing elements commonly defined and belong to an application domain of reference, such as people and their interests (FOAF⁴⁴), bibliographies and citations (BIBO⁴⁵), networks and social activities (SIOC⁴⁶), simple knowledge structures (SKOS⁴⁷).

⁴¹ <http://www.ontologyportal.org/>

⁴² <http://www.loa.istc.cnr.it/DOLCE.html>

⁴³ <http://www.cycfoundation.org>

⁴⁴ <http://www.cycfoundation.org>

⁴⁵ <http://bibliontology.com/>

⁴⁶ <http://sioc-project.org/>

⁴⁷ <http://www.w3.org/2004/02/skos>

An ontology of the business domain, however, specifies particular concepts of a particular business domain and the relationships between them. Ontologies of this level can be derived directly from those described in the upper level , with the aim of exploiting the semantic richness and relevance of the concepts they contain, allowing , in addition, to make the model as generic as possible and relevant standards of - facto. In other words, to adapt the model to another specific domain, it is sufficient only to adapt the domain ontologies, while leaving those defined in the other levels.

A characterizing aspect of the role of domain ontologies business is the ability to classify the resources in the domain of interest with the intent to conceptualize in terms of areas and sub-areas. In the specific case of the automotive domain, these ontologies can, for example, specify the type of processes that the organization is usual to implement. The most natural approach to classification is dictated by the definition of taxonomies which, in their most general meaning, in fact belong to the discipline of classification.

To encode domain ontologies it was decided to use SKOS, Simple Knowledge Organization Scheme, as it allows you to define a taxonomic representation so expressive and easy to implement. SKOS allows you to define synonyms and relations between the terms of a taxonomy, to add information about a concept and can easily be used to define controlled vocabularies. SKOS ontology every element is a concept, which can have a label (`skos:prefLabel`) and one or more synonyms (`skos:altLabel`) . The relationship between concepts can be made through properties that determine a different hierarchical level:

- relations between a concept and a more specific one (`skos:narrower`)
- relations between a concept and a more generic (`skos:Broader`)
- simple correlations (`skos:related`) .

To apply a bottom-up approach, we start from the domain ontologies and then find the possible reference ontologies and then high-level which can be compatible with what is defined . The peculiarity of this approach is that the concepts can be extrapolated in a semi- automatic by business documents, starting from the consideration, not always verified, that most of the conceptual structures characterizing an organization can be extrapolated from the set of such documents. The purpose of this step is, therefore, to formalize the ontology obtained in the previous step, whatever the chosen approach. The output is an ontology formally defined but yet to be evaluated in the next step.

- **Evaluation.** At this stage it should first check the validity of the properties of the ontology built. The aspects that have occurred thus involve the compliance of the language used for the representation (syntax) and the consistency of the model (semantics). The result of this phase is the ontological model ready to be included within the Knowledge Management system.
- **Evolution.** The introduction of ontological models within the company information system involves an evolution similarly to that suffered by the system itself. Specifically, the evolution

of ontologies goes hand in hand with the evolution of business knowledge, by this stage you can go back to the initial process to redesign elements of ontology.

7.2 The process applied to the domain Automotive

7.2.1 Preliminary Study

At this stage, on the basis of the requirements collected and analyzed, is carried out a first representation of the model, with the aim of identifying the main concepts and the relationships between them intercurrent, without the need to be formal. The models belong to the above reference ontologies and semantic structure are responsible for the company's competence in the context of an application domain that embraces several specific business domains. In this study, we chose not to involve high-level ontologies, as implicitly contained in the reference ontologies defined.

7.2.1.1 ***BEST PRACTICE MODEL***

Best Practices are employed for the management of production processes such as application management idea which asserts the necessity of a continuous search for more effective modus operandi to achieve a particular result or any other technique currently in use. In this way the quality of the processes is maintained in a state of secure valence. It is clear, once again, what is the central figure of the worker and its interactions on Social Network Company, on the basis of these interactions, in fact, it is possible to determine the new Best Practice or improve existing ones. A first draft of the model also takes into account the relationship between the essential and the Best Practice process modeled, considering the terms of its relationship with the business structure that manages it. They are, finally, considering all the other elements that characterize the Best Practice, whether they are descriptive or loss.

The result is a preliminary draft of the Best Practice Model (Figure 44):

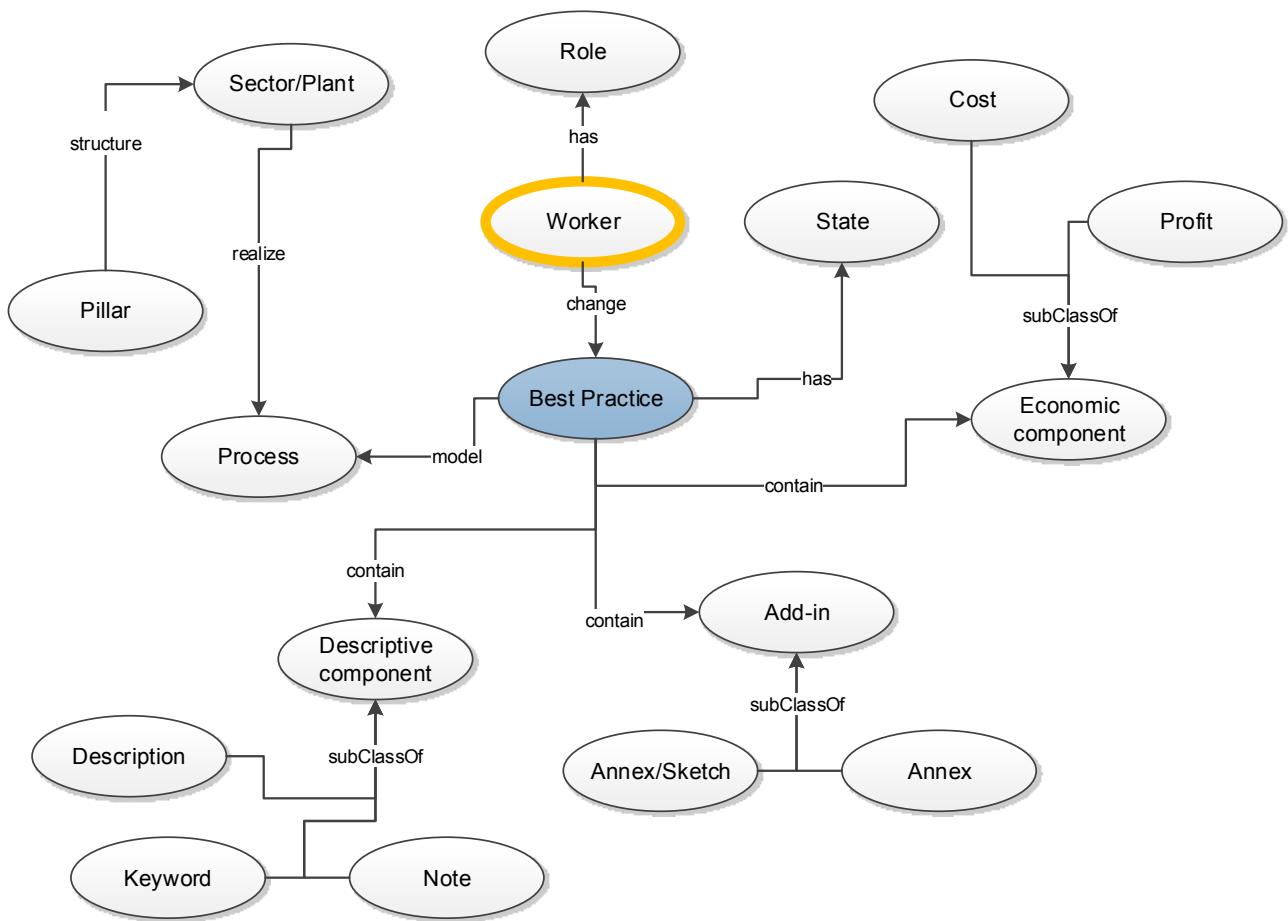


Figure 44 - Best Practice - Preliminary study

7.2.2 **Formal definition model**

In this phase, any operation is performed on the basis of formal modeling of the preliminary models defined in the previous step. First you define the reference ontologies. They are also choices of reference ontologies to extract reusable elements for modeling.

7.2.2.1 **BEST PRACTICE MODEL**

At this stage of modeling was not considered necessary to introduce new classes, but only search for and possibly take in the model of the elements known in the literature. The main purpose of such an introduction is to make not only more powerful model, but also to make it more "universal", that is adaptable to a wide range of organizations. The key point on which we focused was to represent the industrial process as a characteristic element of an organization. In this context, the organization itself must be contemplated in the modeling and for this purpose was used the approach suggested by the ontology ORG (Organization Ontology⁴⁸). This ontology is designed to represent the unique characteristics of an organization of any kind and is primarily characterized by the support structures decomposed into functional units, managing also the dislocation in different geographic locations. Another peculiar feature is to ensure that the modeling of roles and relationships between the

⁴⁸ ORG Ontology: <http://www.w3.org/TR/vocab-org/>

members of the organization. All these elements are perfectly apt with the findings from the previous phases of analysis. Another key consideration as regards the metadating resources carried out by means of a class that uses Metadata elements available in the literature and well known for their ductility in the metadating resources, i.e. projects Dublin Core⁴⁹ and MPEG- 7⁵⁰. The resulting model is depicted in Figure 45.

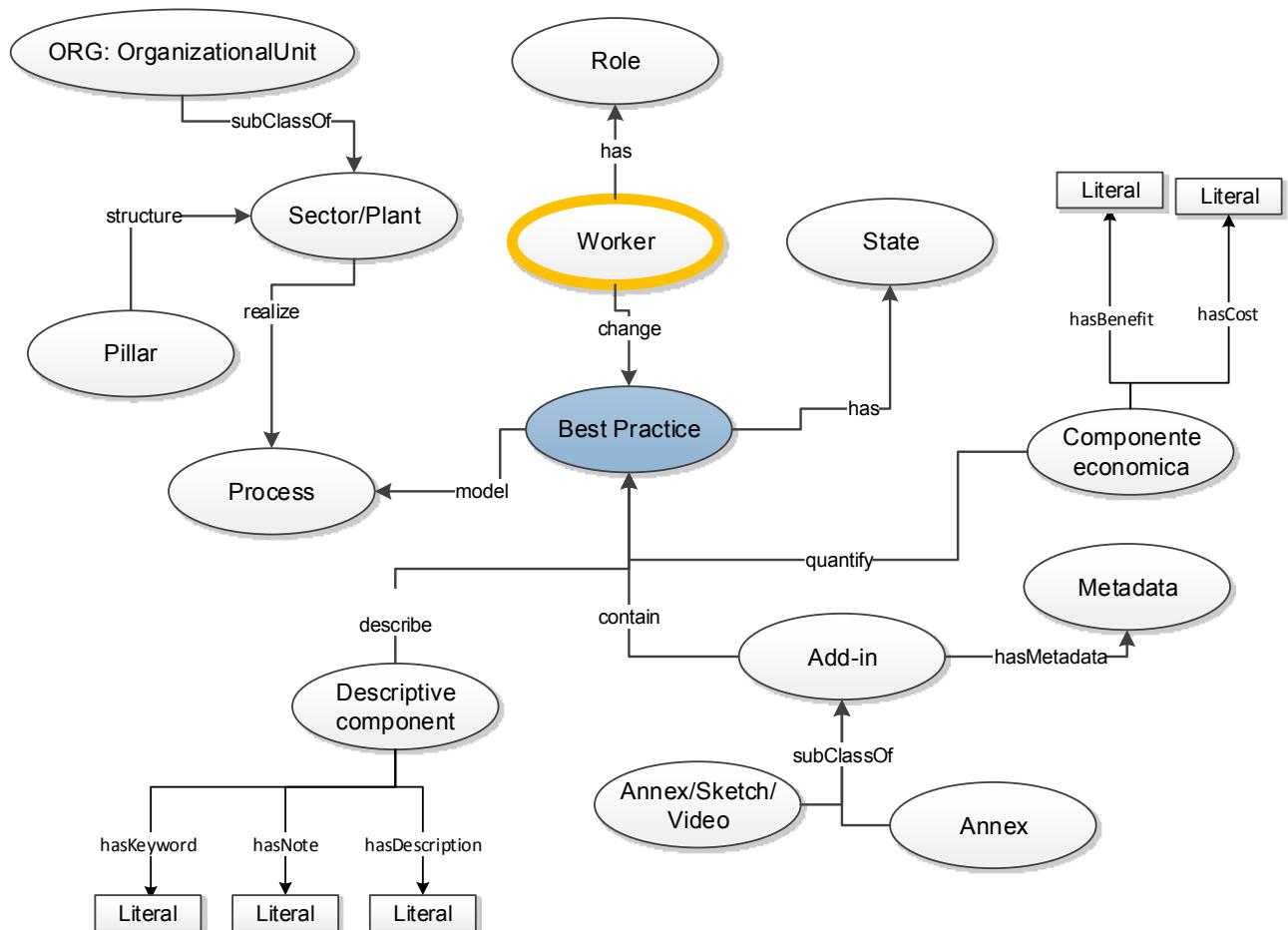


Figure 45 - Best Practice Model

7.2.2.2 DOMAIN BUSINESS ONTOLOGIES

As introduced above, the ontology of the business domain are meant to represent particular aspects of a particular business domain. More in detail, describe the structure of the various concepts that classify the specific knowledge management. To model the domain ontologies has chosen to use the standard SKOS, through which structure the concepts of the operational context of the specific organization.

The role of models , related to the reference ontologies , and domain ontologies is complementary , the first model the knowledge characterizing the application domain of reference identifying all the elements necessary to describe it, the latter , however, describe how deep those that are the

⁴⁹ Dublin Core Metadata Initiative: <http://dublincore.org/>

⁵⁰ MPEG-7: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>

characteristics of the specific business domains . Referring to the research work, the models are ontologically the industrial domain Automotive immersed in a context Enhanced Enterprise 2.0, while the corporate domain ontologies define the specific characteristics. In the design phase of writing in this document, the most significant corporate domain ontologies are identified as follows:

- **Pillar**: technical pillar on which rests the WCM methodology worked ;
- **Role** : representation of business roles defined in an organization ;
- **Sector/Plant**: structuring corporate asset with particular reference to the multinational nature of the company. This organization , in fact, takes into account not only the various operational areas, but also of the establishment in which they are allocated to these sectors;
- **Processes**: set of business processes that characterize the various sectors of the establishment;
- **Areas of responsibility**: closely related to the classification of roles, such a taxonomy specifies which are the areas to which a role of managerial nature , is responsible for ;
- **Plant**: defines the elements that characterize a given facility deputy to perform the production process of an establishment;

7.2.3 Model Evaluation

The proposed model has been implemented using Protégé⁵¹, an open source tool developed and maintained by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine, thought out for the development of ontologies and advanced editing. As a first step, using the feedback functionality (Figure 46), have been defined classes identified in the definition of the Model of Best Practice, also considering the reports of the sub-class.

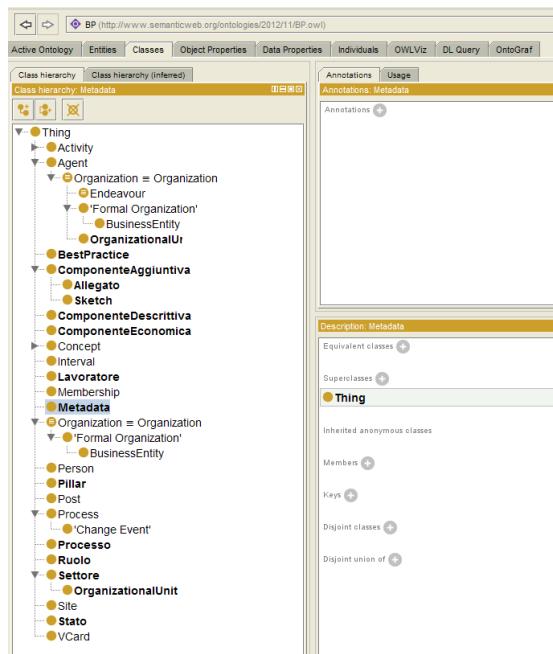


Figure 46 - Creating classes in Protégé

⁵¹ Protégé: <http://protege.stanford.edu/>

Later, have been defined previously identified relationships between classes. The tool used in this operation is achieved by setting the domain, i.e. choosing the class "person", and the range, i.e. the class "object" (Figure 47).

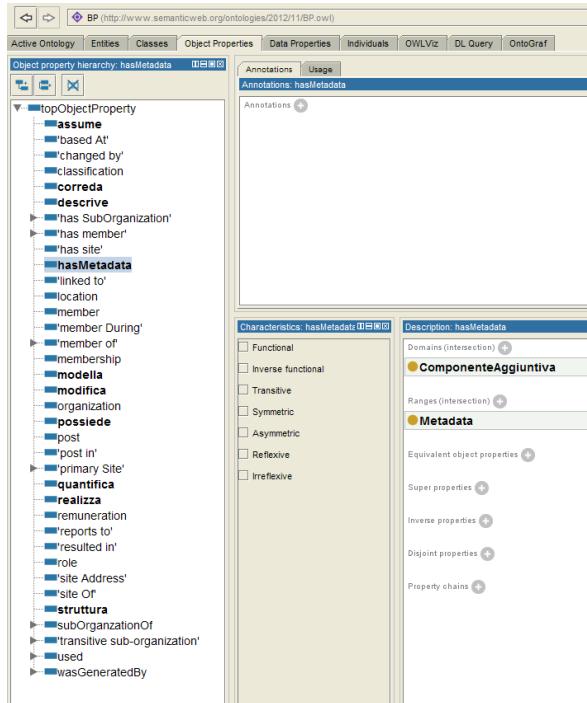


Figure 47 - Relations in Protègè

The resulting ontology is displayed in the appropriate tab (Figure 48).

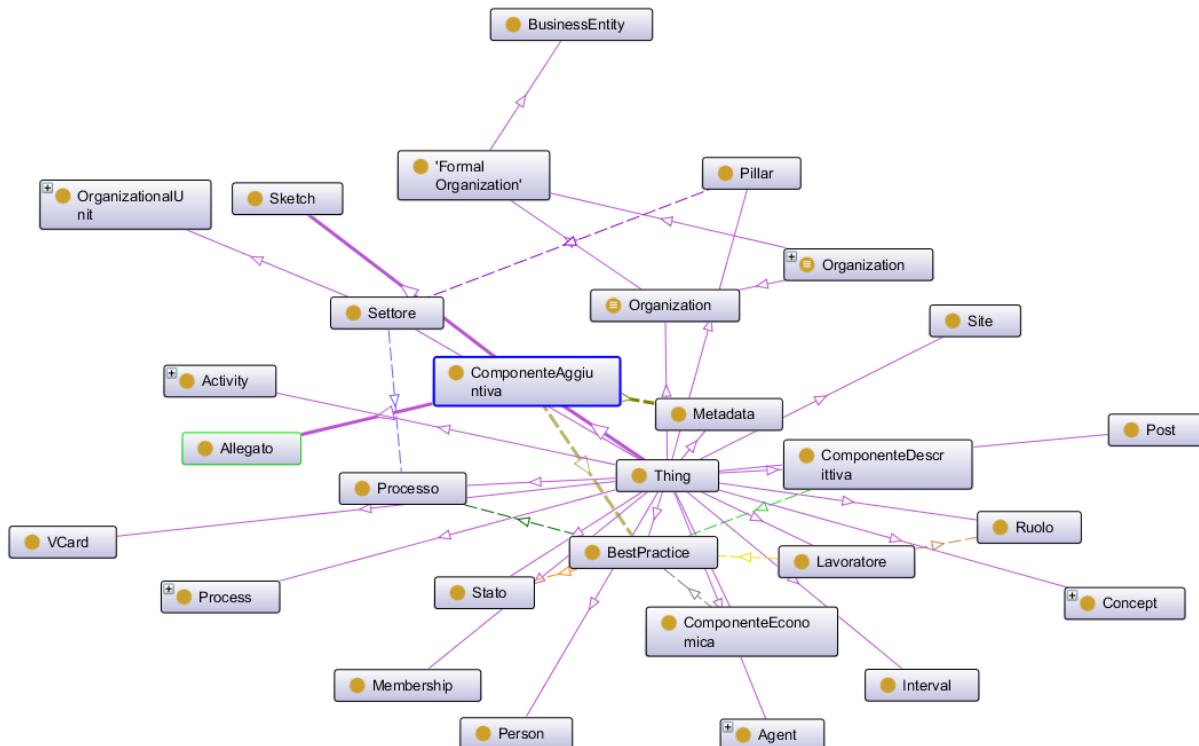
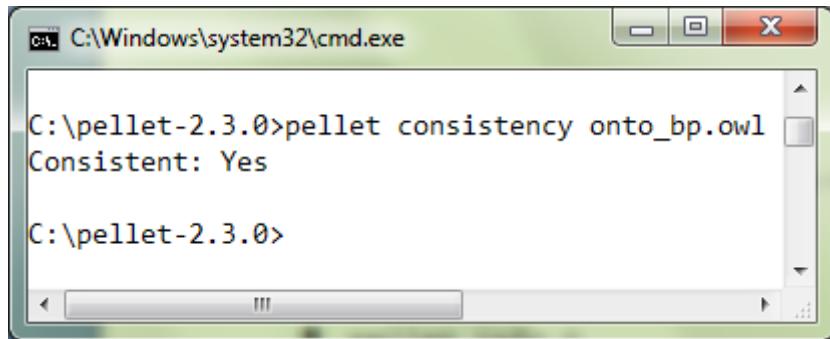


Figure 48 - Graphical representation of the ontology created in Protègè

The ontological model developed was then evaluated by the reasoner Pellet 2.3.0. In particular havebeen used commands:

- Pellet consistency
- Pellet info

The first command is used to verify the consistency of the model, or if there are axioms that are in contradiction. As is apparent from Figure 49, the product model is consistent result.

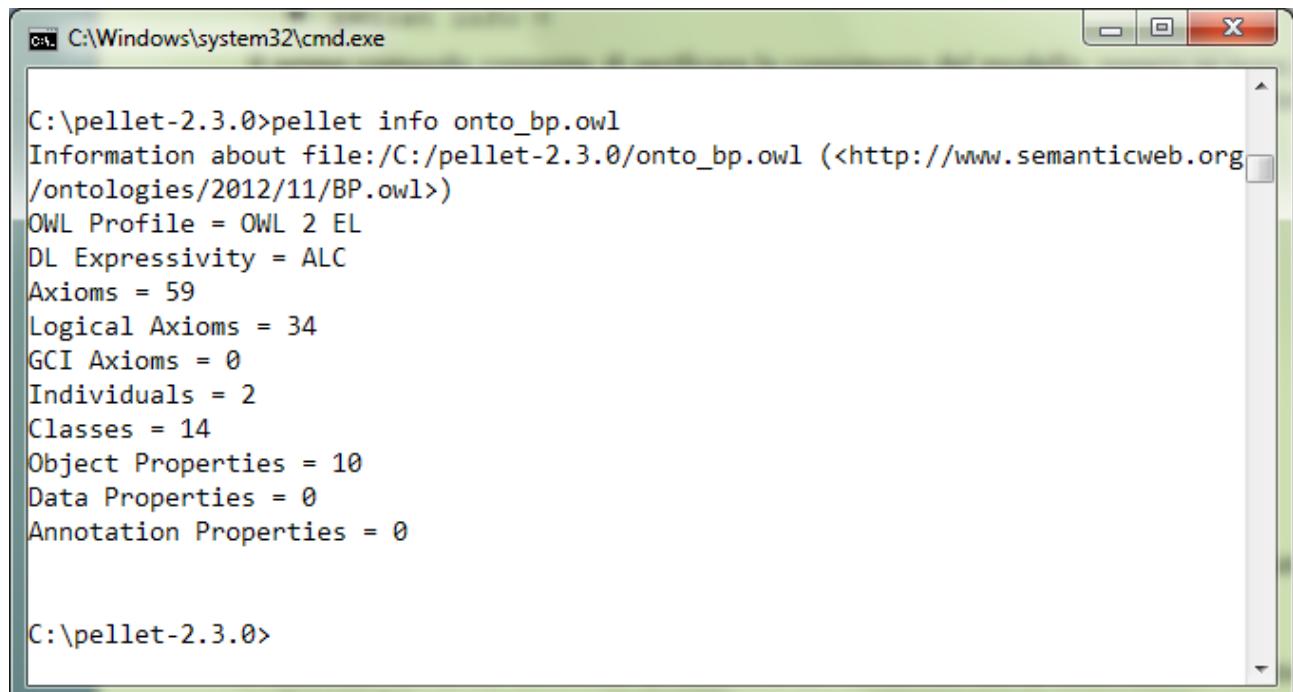


```
C:\Windows\system32\cmd.exe
C:\pellet-2.3.0>pellet consistency onto_bp.owl
Consistent: Yes

C:\pellet-2.3.0>
```

Figure 49 - Assessment of the consistency of the model made by Pellets 2.3.0

The second command enables you to obtain a large set of information about the model analyzed by the reasoner.



```
C:\Windows\system32\cmd.exe
C:\pellet-2.3.0>pellet info onto_bp.owl
Information about file:/C:/pellet-2.3.0/onto_bp.owl (<http://www.semanticweb.org
/ontologies/2012/11/BP.owl>)
OWL Profile = OWL 2 EL
DL Expressivity = ALC
Axioms = 59
Logical Axioms = 34
GCI Axioms = 0
Individuals = 2
Classes = 14
Object Properties = 10
Data Properties = 0
Annotation Properties = 0

C:\pellet-2.3.0>
```

Figure 50 - Evaluation of the model achieved by Pellet 2.3.0.

Figure 50 shows the results obtained by subjecting the reasoner Pellet 2.3.0 the ontological model realized. The ontology was classified as responsive to the OWL 2 EL profile. This profile is based on the tractable fragment of Description Logics called EL ++ ontology that allows for conjunction and existential restrictions and has significant algorithmic properties.

Another aspect of the analysis conducted by Pellet 2.3.0 concerns the expressive capacity of the ontological model. The expressive DL ALC is found and then the ontology model is made of the domain of interest in the following description logics:

- AL: Indicates the logic of attributes and introduces atomic negation operators , intersection of concepts, universal and existential quantification restriction limited ;
- C: describes the possibility of using the negation operator for complex concepts (for example, denial of concepts that are included within other concepts).

8. Enterprise 2.0 Knowledge Management: Best Practices

The role of knowledge began to be an important resource since the 80's because it is able to create value for the enterprise. It may be defined as the most significant resource of our time, a kind of hallmark of modern society. The increased importance of knowledge has marked a crucial step in modern theories of economics and business [62]. This does not mean that up to this time there was no knowledge in enterprises, but simply that his administration was done unconsciously and without considering the importance of the relationship between knowledge and value to the company [62].

The possession of knowledge does not ensure the enterprise to obtain the benefits that could derive from it, but it requires interaction between people who possess such knowledge and are able to create new. So, she made space the need to capitalize on the business knowledge fostering the acquisition, re-use, dissemination and creation. This has resulted in a growing interest in the concepts of knowledge engineering and from the role of organizational memory and processes of accumulation of organizational knowledge. The creation of social enterprise is therefore attributable to an 'action aimed at managing knowledge.

This chapter will introduce the theory behind the enterprise knowledge management and summarize the results in terms of the application of this theory to the application context reference research work in the Automotive Enterprise 2.0 i.e. the management of Best Practice. They play a major role in the context of corporate knowledge and are enablers for optimal operation of processes. The proper management of the flow of knowledge involves the maintenance and continuous updating of Best Practice, avoiding obsolescence and making sure that they are the basic elements of competitiveness. Specifically, this chapter defines the life cycle of Best Practice to promote the continuous improvement of business processes, whether production or coordination.

8.1 Knowledge management flow

The knowledge is manifested in different forms in the organization [63]:

- **Tacit:** represented by what people know but do not know how to express in a formal way through the normal channels of communication, typically based on writing. Tacit knowledge is closely linked to the person, depends heavily on the context and it's hardly made explicit and formalized typically being much less of a practical nature.
- **Implicit:** is that component of the knowledge that you cannot or do not want to express, but of which you are aware and that you would be able to explain, formalize, communicate.
- **Explicit:** available in documentary form in the format: Structured or when stored in enterprise databases, management systems, processes, systems for the representation of knowledge that make use of ontologies.
 - **Semi- structured:** or when stored in the web pages of corporate intranets and the Internet (based on HTML and XML).

- **Non- structured:** when stored in textual documents of any kind used in the organization.

In particular, according to Nonaka and Takeuchi, Japanese creators of the theory of knowledge, the creation of knowledge is to be understood as a diffusion process in which knowledge is created by individuals within the network of systematized knowledge of the organization. From here Nonaka proposes the model called Organizational Knowledge Conversion which presents the process of knowledge management as a spiral in which new knowledge is created always [64].

Figure 51 shows the spiral on the dynamics of knowledge creation based on conversions tacit / explicit through the processes of socialization, externalization, combination, and internalization. This spiral develops along two dimensions: the first, called "epistemological", concerns the interactions between tacit and explicit knowledge, the second dimension, the "ontological", concerns the individual and the organization. According to this model an organization is able to create knowledge only through individuals working in it, in the hope, therefore, a valuation and an incurrence of the most creative by inserting them in a collaborative environment in which knowledge is created.

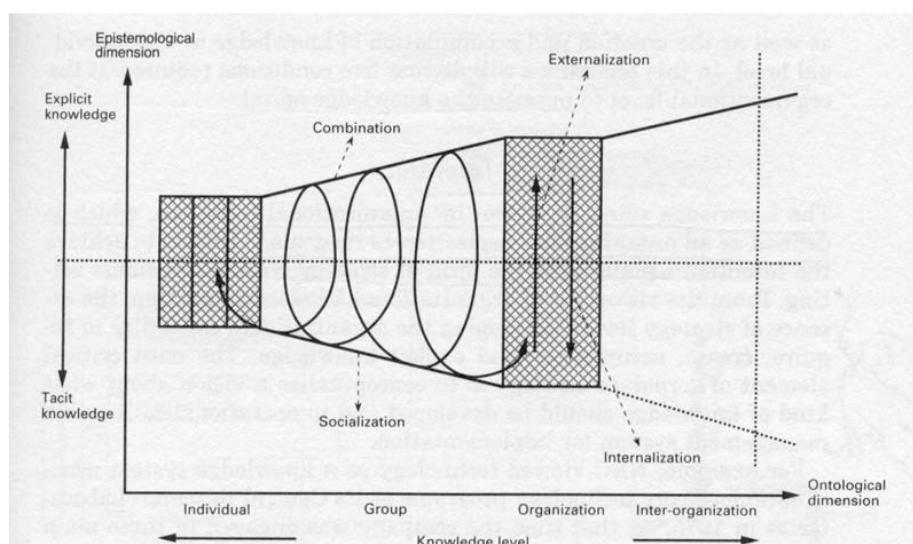


Figure 51 - Spiral of organizational knowledge creation

This process is governed by the SECI model, proposed by the same Nonaka (Figure 52).

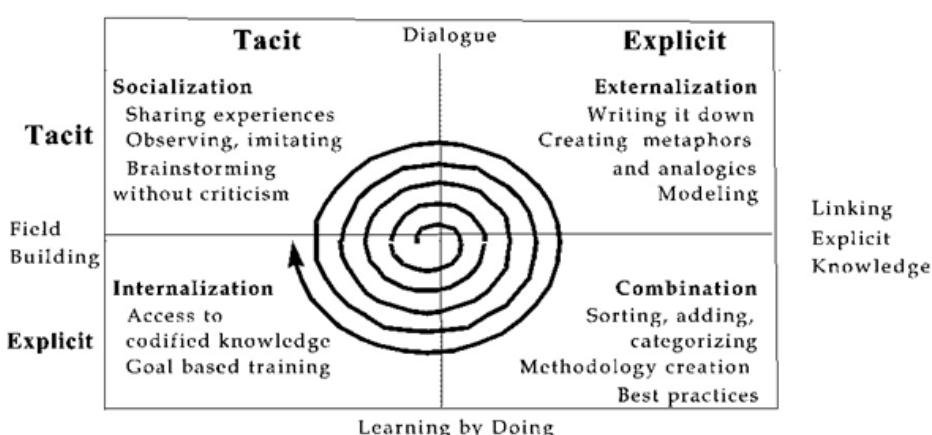


Figure 52 - Seci Model

The SECI model consists of four modes of knowledge conversion: socialization (tacit to tacit), externalization (tacit to explicit), combination (explicit to explicit), and internalization (explicit to tacit).

- **Socialization** is the process of sharing tacit knowledge through observation, imitation, practice, and participation in formal and informal communities. The socialization process is usually preempted by the creation of a physical or virtual space where a given community can interact on a social level.
- **Externalization** is the process of articulating tacit knowledge into explicit concepts (Yeh et al., 2011). Since tacit knowledge is highly internalized, this process is the key to knowledge sharing and creation.
- **Combination** is the process of integrating concepts into a knowledge system.
- **Internalization** is the process of embodying explicit knowledge into tacit knowledge.

The SECI model shows that the process of knowledge creation is cyclical, was born at the individual and develops at a group level, ending at the organizational level. In order for this process to take place it is necessary to implement models of social interaction that support the realization of the SECI process, among them we can mention, in particular, the Ba and communities of practice. The model of social interaction called Ba (Japanese term that means " place , a place , an arena for creative exchange") , which was introduced by the same Nonaka defines the ways in which people communicate in order to achieve the above process of converting knowledge from tacit to explicit . The interactions can take place physically or virtually through special software tools. Nonaka identifies different types of Ba each specialized to support a different stage of the process SECI: Originating Ba for the phase socialization, Dialoguing Ba for the ' externalization, Systemizing Ba to implement the combination and Exercising Ba for internalization [64]. The interplay between the four categories of ba is illustrated in Figure 53.

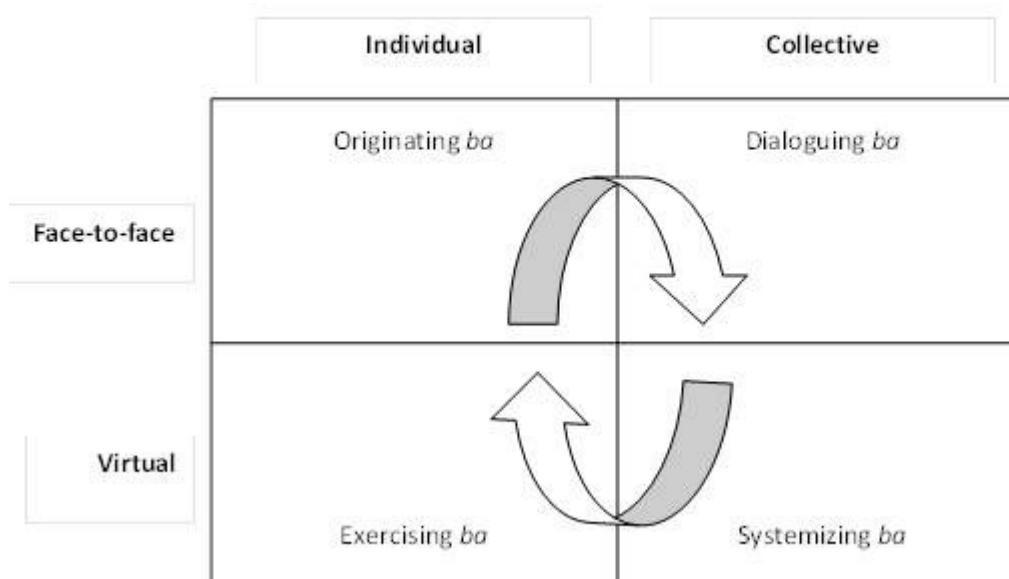


Figure 53 - Four categories of ba

Another model of representation of social interactions aimed at the generation and development of corporate knowledge is that of communities of practice, groups of employees who share a common heritage of knowledge and interact informally, exchanging knowledge on issues of mutual interest; information exchanges result in the generation and sharing of new knowledge. Communities of practice are easily implemented through collaborative tools of the social network business, plus the ability to capture and capture a portion of tacit knowledge, making it explicit.

The ICT knowledge management are the real enabler of any strategy for knowledge management. Understanding the potential offered by the different technological solutions available is crucial for their correct application: the attempt to create a management culture of shared knowledge, based solely on an organizational approach and on the active collaboration between people, without the presence of an appropriate computer system, leads to results necessarily partial and potentially disastrous.

8.2 Benchmarking and Best Practice

The analysis of business processes to maximize the performance is certainly one of the elements that characterize the management of an organization. One possible approach to this problem is realized by comparing the processes characterizing their business model with those of other organizations or with new models, assessing the characteristics and ensuring any improvements. This approach is called benchmarking and has been formally defined as "the search for industry best practices that lead to have superior performance" [65]. In other words, benchmarking is an activity which consists in "learn, share best practices and adapt to a business reality [66].

This operating mode essentially comprises three steps:

- understand the need to improve a process ;
- find a best practice that improves the process;
- Assess extent to which this process has been improved.

The first two steps refer to a process of continuous research typical of benchmarking, which therefore differs from traditional analysis of competitiveness as it is a constantly evolving , which compares business processes with all the " better " that can be researched or proposed , both inside and outside the organization. This research has potential limitations, since any process can be challenged by another at any time, the more it stimulates the creativity of the business community and it is open to all possible instances of itself, the more it is likely to find a best practice that take the place of a process currently in place.

The last phase involves, essentially, the verification of the performance of the best practices applied to the process that you are trying to improve, whereas the concept of performance is very variable and dependent on the particular organization you are analyzing. It seems clear that the concept of best practices is central in the whole organization and how much goes to impact on all business processes. A best practice can be defined as "the best way currently

known to implement a particular business process ", the adjective "best" is often questioned, preferring the term good practice, highlighting the fact that this solution is subject to a continuous cycle of monitoring and review, which often leads to finding another best practice for the same process. It can, therefore, affirm that a best practice constitutes the operating model currently used to implement a series of similar processes, but which must be continually questioned, with the aim of finding another that further improve the performance of the process.

Many best practices are inherent in the company's tacit knowledge, that certain processes are carried out under conditions resulting from the practice, but not formally defined. Therefore in addition to managing the best practices defined, it is appropriate to prepare strategies to try to extract the tacit, perhaps preparing communities of practice, communicate and collaborate through social networks. The two approaches are complementary in the sense you cannot think of extracting best practices from the corporate tacit knowledge without good organization of best practices already formalized.

The essence of best practices is therefore to find a better solution to a given problem and to share in the company, with immediate advantages of different types: [67]

- identify and replace practices "poor" and obsolete;
- minimize costs through improved productivity and efficiency;
- facilitate the resolution of problems through the search for solutions to similar problems;
- ensure consistency, having the certainty that a given business process is always done in a certain way;
- decrease the time of learning, because of the greater simplicity of creating educational materials for standardized processes;
- improve the overall quality of the products or services offered.

In the literature point to six fundamental points for the identification and sharing of best practices [68]:

- **Identify the real needs:** never lose sight of the main purpose of a best practice, or improve a business process in order to understand where we can provide added value to the organization. In this sense, in search of best practices should involve primarily corporate sectors in difficulty or performance deficit.
- **Discover best practices:** there are potentially many ways to find a best practice. A first approach consists in observing the operating procedures that have produced good results in other organizations and assess whether they can become best practices for their own. A second approach is embodied in formalizing a best practice, starting from the above company tacit knowledge, which is going to shape a practice applied in an informal way thanks to the skills of workers. A final approach is to harness the creativity of the workers, who should have the opportunity to promote their own ideas to improve a particular process, the best insights could be explored and eventually formalized as a best practice.

- **Template:** the description of the good practice is generally placed in a repository according to a standard format. A typical template should include the following sections:
 - Title - short descriptive title (may be accompanied by a short abstract) ;
 - Profile - short sections that outline processes, function, author, keywords, etc.
 - Context - where it is applicable? What problem does it solve?
 - Resources - what resources and skills are needed to carry on the good practice?
 - Description - what are the processes and steps involved?
 - Margins for improvement - there are good performances associated with this practice?
 - Tools and techniques used
 - Validate best practices. Best practice is cyclical in nature, in the sense that its actual dowry to be "better" must be constantly challenged through precise operations evaluation of existing best practices and validation of new ones. A usual approach is to have a team of auditors including external experts, both internal and external to the organization. Equally important are the input and feedback (i.e. the last beneficiaries) of best practices.
 - Dissemination and application. A database of good practices is a useful starting point, but many organizations are essential to accompany him with a face-to -face sharing of the knowledge of good practice. Usual ways of knowledge sharing of good practice include: communities of practice, improvement groups, learning events organized, etc.
 - Develop and support infrastructure. It is need to make sure that you have the necessary infrastructure to have a proper management best practices. This infrastructure is generally included in the more general business knowledge, so much so that he has often mentioned that the best practices involving both tacit and explicit knowledge. A modern social network, an advanced DMS and a suite for semantic knowledge management constitute the minimum equipment to implement a modern management best practices.
 - There are different methods of business process management using best practices as a driving force. The following is a list of those considered most relevant, with no presumption of completeness:

There are different methods of business process management using best practices as a driving force. The following is a list of those considered most relevant, with no presumption of completeness:

- **Six Sigma:** This methodology, which was introduced in the 80s by Motorola, is its assumptions in the approach of Deming's PDCA cycle, the model developed for the improvement of quality on a long haul through the continuous refinement of processes and to optimal use of resources. Six Sigma seeks to combine the European trend to improvement through systematic changes (breakthrough) with the Japanese approach to

continuous improvement pursued through "baby steps". All business processes are analyzed, with the assertion that each process can be measured, and that therefore it is possible to intervene with measures to improve only after you have performed the measurements of the characteristic parameters or indicators more representative and having analyzed the data thus obtained. [69]

- **TOGAF**: created in 1995, is based on the Technical Architecture Framework for Information Management (TAFIM) of the Department of Defense of the United States of America. It ' a standard infrastructure to handle the corporate structure, with the ultimate goal of maximizing productivity and process performance. TOGAF consists of three basic parts: the Architecture Development Method (ADM), which describes how to get enterprise architecture specific to a particular organization that meets certain requirements, the Enterprise Continuum, a sort of repository of all possible assets of the company that can be taken as an example to develop its architecture, the TOGAF Resource Base, a set guidelines , templates and various information useful to support architects in the use of ADM. [70]
- **eTOM** : The Business Process Framework (eTOM) , formalized in 2003, is a multi- level , hierarchical view of business processes deemed necessary to achieve an efficient and profitable . At a conceptual level, the framework has three main process areas: Strategy / Infrastructure / Product, Operations and Management. The strong point for this approach is to encourage reusing Process from different organizations with a consequent lowering costs while improving Performance Process. . [71]
- **ITIL**: providing a broad set of best practices related to IT processes, ITIL is one of the most widely used framework for the management of the companies or IT departments. Its key element is the continuous evaluation and possible improvement of services offered, both from the company point of view and from that of the client. Published for the first time in England between 1989 and 1995 by Her Majesty's Stationery Office (HMSO) under the aegis of the Central Communications and Telecommunications Agency (CCTA), was initially used almost exclusively in the United Kingdom and the Netherland, from its second version, developed since 2000, its use has spread on a global scale and is probably one of the most well-known frameworks. . [72]
- **Kaizen**: Japanese management strategy which means "continuous improvement." Its definition is derived , in fact , from the Japanese words " kai " which means " continuous " or " change" and "zen " which means " improvement ", " better." This method encourages and supports small improvements to be done day after day, in a continuous manner. The kaizen, initially presented by Toyota in the 80s and applied more and more in the world, is based on the principle of the emergence of corporate knowledge from below, i.e. on the understanding that the performance of business processes can be increased only through the analysis of the proposed improvements by workers in the field of collaborative. [73]

8.3 Best practice and business competencies

The study of various business skills, particularly in relation to the fact that a worker who is in possession or not, is often approached with an assessment resulting from studies carried out in 1980 by brothers Stuart and Hubert Dreyfus faculty at the University of California [74]. This scale describes, dividing it into five levels (Figure 54), the phase of acquisition of skills by the worker, through which he became adept at grooming a given process [75]:

Level	Stage	Characteristics	How knowledge etc is treated	Recognition of relevance	How context is assessed	Decision-making
1	Novice	Rigid adherence to taught rules or plans Little situational perception No discretionary judgement	Without reference to context			
2	Advanced beginner	Guidelines for action based on attributes or aspects (aspects are global characteristics of situations recognisable only after some prior experience) Situational perception still limited All attributes and aspects are treated separately and given equal importance		None	Analytically	
3	Competent	Coping with crowdedness Now sees actions at least partially in terms of longer-term goals Conscious, deliberate planning Standardised and routinised procedures				Rational
4	Proficient	Sees situations holistically rather than in terms of aspects Sees what is most important in a situation Perceives deviations from the normal pattern Decision-making less laboured Uses maxims for guidance, whose meanings vary according to the situation	In context	Present	Holistically	
5	Expert	No longer relies on rules, guidelines or maxims Intuitive grasp of situations based on deep tacit understanding Analytic approaches used only in novel situations or when problems occur Vision of what is possible				Intuitive

Figure 54 - Novice-to-Expert scale

Table 5 describes for each of these levels, the manner in which an employee who has that level of competence for a given process, it relates to a best practice that describes it:

STAGE	BEST PRACTICE
Novice	<ul style="list-style-type: none"> Requires a best practice to perform an operation. It is able to judge the performance resulting from its application to a process. It needs training to learn the elements described by the best practices.
Advanced beginner	<ul style="list-style-type: none"> Ability to independently perform many steps of the operation sequence on the process. It is able to judge the performance of the process. It should be followed carefully as the increasing confidence in their own ability could lead to the belief that they can change the process, yet without having the necessary expertise.
Competent	<ul style="list-style-type: none"> Knows and executes the process without the need for training. It is capable of judging the process in its entirety.

	<ul style="list-style-type: none"> • Can think of changes to the process, that is, new best practices, as well as mindful of mistakes when it belonged to the previous levels.
Proficient	<ul style="list-style-type: none"> • It's able to propose improvements to the process, or new best practices. • It's able to establish similarity with other processes. • Must be discouraged from changing the process without following a best practice.
Expert	<ul style="list-style-type: none"> • Can both propose best practices that assess those proposed by others. • It should be encouraged to think more and new improvements to the process. • Must be discouraged from changing the process without following a best practice.

Table 20 - Dreyfus Model and Best Practice

It is evident that more experienced it becomes a worker in a particular process, the more you run the risk that he may turn away from best practices. Minimize this risk involves not only bind workers to follow best practices, but also entice workers with high levels of competence in the provision of new best practices

8.4 Knowledge Enterprise Model : Best practice

From the analysis made in the previous paragraphs shows the need to address two key issues for modern organizations: knowledge and process performance. The latter is achieved by managing the most of a particular area of business knowledge, those best practices introduced earlier. The strength of the model is proposed, shown in Figure 55, is to leverage corporate knowledge to better manage the life cycle of best practices and consequently improve the performance of business processes.

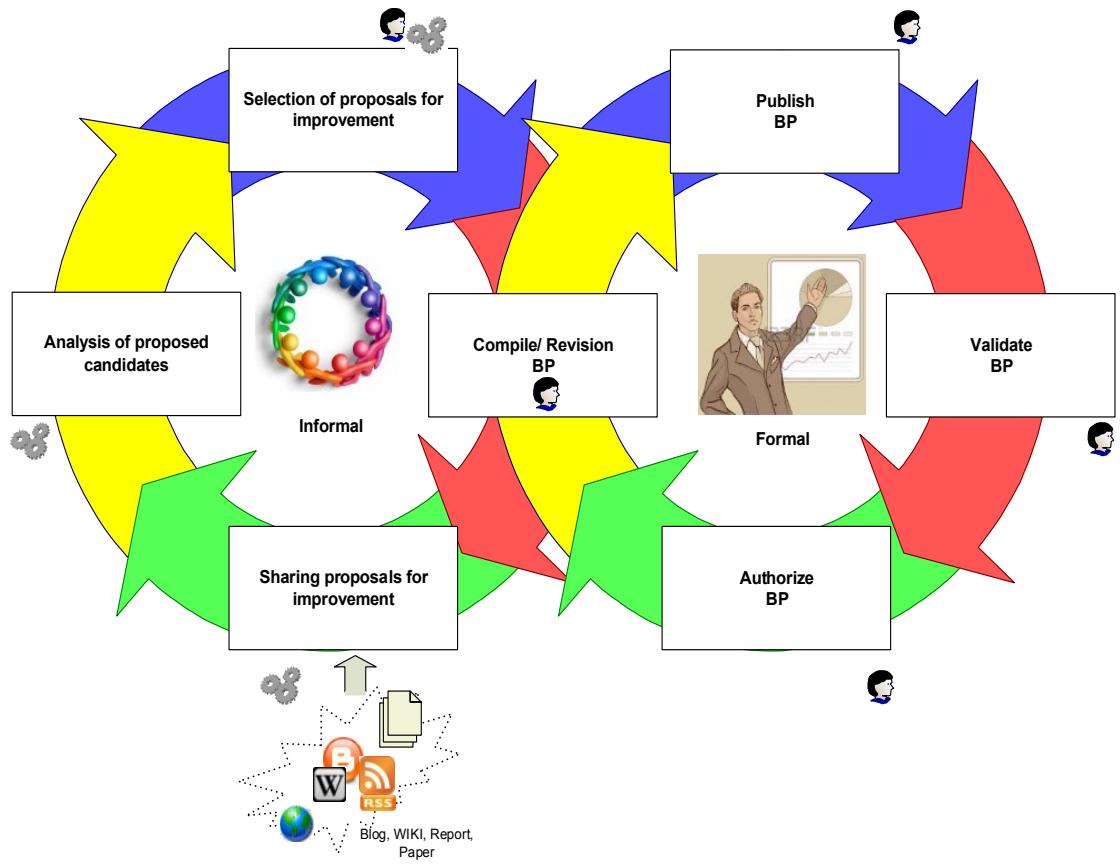


Figure 55 - Best Practice Life Cycle

The cycle is divided into two main phases: a formal and one informal. Formal refers to the management of best practices with regard to the nature of their improvement to a process, in other words, this phase sees the best practices as a formal proposal for the improvement of a business process, typically made explicit by means of special template document. A proposal, whether it is a compilation of a new or a revision to an existing best practice, follows a precise business process before becoming possibly process standards. This path depends strictly on the organization, but the steps that compose it can be traced to the following steps:

- **Compile / Revision:** is the creation of new best practices or bring changes to existing BP. During this phase, an issue resolved, such resolution shall be made available to everyone in the company, entering all the information into a template set up and submitting to being published such best practices.
- **Publish:** completed the drafting process, the best practice passes being published to undergo the procedure for official authorization and validation to become operational.
- **Validate:** is the time in which the best practice is subjected to a control both formal content, before being sent to final approval.
- **Authorize:** is when the best practices , as well as approved in its form and content, is emitted and thus made available to other users of the system, and in general, all users who have the necessary access rights. Typically is evaluated, empirically, in some cases, the improvement

in performance of which has described the process through the implementation of best practices. This phase, like the three previous ones, by one or more persons empowered, which can operate independently or in collaboration with specialized tools (wiki, forum, etc.).

The phase described above only covers the steps of combining and subsequent internalization of the SECI model that deals exclusively with the explicit knowledge, for more in a strictly structured. To lower management best practices in the entirety of enterprise knowledge management, it is necessary to introduce a phase called informal, that manages to bring out instances arising from tacit knowledge and implicit. This phase can be defined by the following steps:

- **Sharing proposals for improvement:** after the phase of socialization, made possible through the collaborative tools provided by the social network business, you can enjoy the considerations expressed by workers in reference to a particular business process. Their tacit or implicit knowledge can lead to deduce proposed improvements resulting from the practice in carrying out a specific process or even from simple intuitions.
- **Analysis of proposed candidates.** The proposals that emerged in the previous step are discussed and refined by the user community, thanks to the outsourcing process. At this point the knowledge of individual workers has been clarified and is potentially available to the entire business community.
- **Selection of proposals for improvement.** The knowledge has emerged, at this point, needs to be conveyed to the processes on which the company has actually need or interest to invest.

Thanks to the work of a committee evaluation are selected among those that emerged, only those proposals that could improve processes relevant to the business asset. These proposals come in the formal stage to be estrinsecate form of potential best practices.

8.5 Conclusions

Best practices are a powerful way to leverage corporate knowledge, increasing competitiveness through continuous improvement and constant asset of the organization. Knowledge, whether express, implied or constructive, of the workers is used to bring continuous improvements to business processes. If explicit knowledge is easily managed through a workflow similar to a document management, tacit or implicit needs to be extracted and fed. An ideal approach is to examine the activities of the workers on the social network business, in particular the interactions and communications made through collaborative tools. In this way the activities are possible externalization and internalization to make explicit the implicit knowledge and make manifest the tacit. For these purposes collaborative tools typical of Web 2.0 and Enterprise 2.0 are not characteristic enough, we need more sophisticated tools that are based on emerging semantic technologies in the area of the Semantic Web

9. Conclusion and Future work

This chapter closes the thesis work by describing a short summary. Furthermore, Section 9.2 describes future challenges.

9.1 Summary

This research work addresses the methodology of automatic extraction of knowledge taking into account the text corpus (e.g., a file txt, pdf, etc.). In particular, the extracted knowledge has been translated into an artifact ontology using semantic web formalism, such as: RDF and OWL. This is achieved through the definition of a methodology for Ontology extraction, in order to map Lattice Fuzzy in the structure ontology. In addition, this methodology has been extended and applied to different research objectives: semantic annotation and Information Retrieval.

These methodologies have been applied to several case study:

- **Ontology / Taxonomy Extraction:** conceptualizing the content.
- **Information Retrieval :** semantic search engines that , on the one hand , extend the traditional system of Information Retrieval (IR) from a simple system of Document Retrieval to a system of the Entity and Knowledge Retrieval , on the other hand , improve the conventional methods of IR , from different points of view : the meanings of words can be formalized and represented in a machine-processable format using ontology languages such as RDF and OWL (i.e., a resource can be described as an ontological class, with a set of attributes, relations with other entities, constraints, etc.).
- **Semantic annotation:** applied to normal text annotation, but the framework is applicable to other types of media resources.

9.2 Future Work

To achieve interoperability and the development of knowledge models extracted, the challenges of the future in the following directions:

- **Ontology alignment.** Study, definition and development of the approaches for extracting approximate matching in order to harmonize heterogeneous ontology conceptualization. The result of a matching operation is the evaluation of relation between two ontologies. Concepts matching enable us to augment knowledge discovery performances.
- **Ontology Merging.** Study, definition and development of the approach to create ontology from two or more source ontologies. The new ontology will unify and in general replace the original source ontologies. By the merging of the concepts ontology, we intend to support knowledge Extraction applications through reduction of redundancy.

Bibliography

- | |
|---|
| [1] D. Milne and I. H. Witten, "Learning to Link with Wikipedia," New York, NY, USA, 2008. |
| [2] P. Cimiano, A. Hotho and S. Staab, " Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis," <i>Journal of Artificial Intelligence Research</i> , vol. 24, no. 1, pp. 305-339, 2005. |
| [3] B. Zhou, S. Hui and C. Alvis, <i>An Effective Approach for Periodic Web Personalization</i> , Washington, DC, USA: IEEE Computer Society, 2006, pp. 284-292. |
| [4] C. De Maio, G. Fenza, V. Loia and S. Senatore, "Hierarchical web resources retrieval by exploiting Fuzzy Formal Concept Analysis," <i>Information Processing & Management</i> , #may# 2011. |
| [5] C. De Maio, G. Fenza, M. Gaeta, V. Loia, F. Orciuoli and S. Senatore, "RSS-based e-learning recommendations exploiting fuzzy FCA for Knowledge Modeling.,," <i>Appl. Soft Comput.</i> , vol. 12, no. 1, pp. 113-124, 2012. |
| [6] C. De Maio, G. Fenza, V. Loia and S. Senatore, "Knowledge structuring to support facet-based ontology visualization.,," <i>Int. J. Intell. Syst.</i> , vol. 25, no. 12, pp. 1249-1264, 2010. |
| [7] C. Aswanikumar and S. Srinivas, "Concept lattice reduction using fuzzy K-Means clustering.,," <i>Expert Syst. Appl.</i> , vol. 37, no. 3, pp. 2696-2704, 2010. |
| [8] S. Dominich, <i>The Modern Algebra of Information Retrieval</i> , vol. 24, Berlin: Springer, 2008. |
| [9] K. S. Cheung K. and D. R. Vogel, "Complexity Reduction in Lattice-Based Information Retrieval.,," <i>Inf. Retr.</i> , vol. 8, no. 2, pp. 285-299, 2005. |
| [10] J. Dobsa and B. D. Basic, "Comparison of information retrieval techniques: {Latent} semantic indexing and concept indexing," <i>Journal of Information and Organizational Sciences</i> , vol. 28, 2004. |
| [11] C. Budayan, I. Dikmen and M. T. Birgonul, "Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping," <i>Expert Syst. Appl.</i> , vol. 36, pp. 11772-11781, November 2009. |
| [12] G. Fenza, V. Loia and S. Senatore, "Concept mining of semantic web services by means of extended Fuzzy Formal Concept Analysis (FFCA)," 2008. |
| [13] B. Ganter and R. Wille, <i>Formal Concept Analysis: Mathematical Foundations</i> , Berlin/Heidelberg: Springer, 1999. |
| [14] G. Birkhoff, <i>Lattice Theory</i> , 3rd ed., AMS, 1993. |
| [15] D. M. Carmen, <i>Fuzzy Concept Analysis for Semantic Knowledge Extraction</i> , 2011. |
| [16] A. Burusco and R. Fuentes-González, "Construction of the L-fuzzy concept lattice," <i>Fuzzy Sets and Systems</i> , vol. 97, no. 1, pp. 109-114, 1998. |
| [17] R. Belohlávek, <i>Fuzzy Relational Systems: Foundations and Principles</i> , Norwell, MA, USA: Kluwer Academic Publishers, 2002. |
| [18] R. Belohlávek, "Lattices of Fixed Points of Fuzzy Galois Connections.,," <i>Math. Log. Q.</i> , vol. 47, no. 1, pp. 111-116, 2001. |
| [19] B. Zhou, S. Hui and K. Chang, "A formal concept analysis approach for Web usage mining," <i>Intelligent information processing</i> , p. 437-441, 2005. |
| [20] "Enterprise 2.0 sotto i Riflettori nelle Aziende Italiane," The Social Enterprise: Value co-creation putting people at the center, 15 Marzo 2008. [Online]. Available: http://www.socialenterprise.it/index.php/2008/03/15/enterprise-20-sotto-i-riflettori-nelle-aziende-italiane/ . |
| [21] A. P. McAfee, "Enterprise 2.0: The dawn of Emergent Collaboration," <i>MIT Sloan Management Review</i> , vol. 74, no. 3, Aprile 2006. |
| [22] S. Epifani, "Dispense del corso di Tecnologie della Comunicazione applicate all'impresa - Facoltà di Scienze della Comunicazione - Università "La Sapienza" di Roma," [Online]. Available: http://tci09.wordpress.com . [Accessed 27 Giugno 2012]. |
| [23] "Cos'è l'Enterprise 2.0," 2010. [Online]. Available: http://www.enterprise2forum.it/it/enterprise-2.0 . |
| [24] D. Del Monte and P. P. Veroni, "Tecnologie ICT per l'Enterprise 2.0: tra Project Management e Collaborazione," 2007-2008. [Online]. Available: http://www.diplod.it/wp-content/uploads/2009/10/tesi-final-pubblica.pdf . |
| [25] C. Mariano, M. Antonella, L. Pellegrini and P. Andrea, <i>Emerging Approach to E2.0: The Case of Social</i> , 2012. |
| [26] S. Epifani, "Tecnologie della comunicazione legate all'impresa- L'Enterprise 2.0," Roma, 2009. |
| [27] D. Hinchcliffe, "The State of Enterprise 2.0," 22 ottobre 2007. [Online]. Available: http://blogs.zdnet.com/Hinchcliffe/?p=143.. |
| [28] P. d. Milano, "Osservatorio Enterprise 2.0," [Online]. Available: http://www.osservatori.net/enterprise . |

- [29] C. Pierani, «Nuovi modelli collaborativi di impresa: nascita e trasformazione dell'Enterprise 2.0 - Tesi di Dottorato in Ph.D in Business History and Management - Dipartimento di Scienze Aziendali Economiche e Statistiche - Università degli Studi di Milano,» 25 Marzo 2011. [Online]. Available: <http://air.unimi.it/handle/2434/155489>. [Consultato il giorno Giugno 28 2012].
- [30] S. Epifani, "L'ENTERPRISE 2.0," 2009. [Online]. Available: <http://www.slideshare.net/stefanoepifani/dispensa-4-enterprise-20>.
- [31] K. Kirchner, L. Razmerita and S. Frantisek, "New forms of interaction & knowledge sharing on Web 2.0".
- [32] D. L. a. R. V. Guha, "Representation and Inference in the Cyc Project.,," *Building Large Knowledge-Based Systems*, Vols. Addison-Wesley., no. ISBN 0-201-51752-3, 1990.
- [33] A. Passant, P. Laublet, J. G. Breslin and S. Decker, "Enhancing Enterprise 2.0 Ecosystems Using SemanticWeb and Linked Data Technologies: The SemSLATES Approach, contenuto," in *Linking Enterprise Data*, Springer, 2010.
- [34] J. G. Breslin, A. Passant and S. Decker, *The Social Semantic Web*, Verlag Berlin Heidelberg: Springer, 2009.
- [35] A. Passant and P. Laublet, "Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data.,," in *WWW 2008 Workshop Linked Data on the Web (LDOW2008)*, Pechino, Cina, Apr 2008.
- [36] T. M. Cover and P. E. Hart, "Nearest neighbour pattern classification," *{IEEE} Transactions on Medical Imaging*, vol. 12, no. 21--27, 1967.
- [37] N. R. and V. P., "Ontology learning and its application to automated terminology translation," 2003.
- [38] I. Nonaka and H. Takeuchi, "The knowledge-creating company: how Japanese companies create the dynamics of innovation," 1995.
- [39] H. Trevor, T. Robert and F. Jerome, *The Elements of Statistical Learning*, Springer, 2009.
- [40] L. Thomas K. and D. Susan T., "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211-240, 1997.
- [41] G. Fenza, V. Loia and S. Senatore, "Concept mining of semantic web services by means of extended Fuzzy Formal Concept Analysis (FFCA)," 2008.
- [42] J. C. Bezdek, R. Ehrlich and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers \& Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.
- [43] N. Chinchor, "MUC-4 Evaluation Metrics," in *Proceedings of the Fourth Message Understanding Conference*, 1992, pp. 22-29.
- [44] N. Chinchor and B. Sundheim, "MUC-5 evaluation metrics.,," 1993.
- [45] D. C. Blair, "Information Retrieval," *Journal of the American Society for Information Science*, vol. 30, no. 6, pp. 374-375, 1979.
- [46] S. Castano, A. Ferrara, D. Lorusso and S. Montanelli, "On the Ontology and Instance Matching Problem," in *19th International Conference on Database and Expert Systems Application*, 2008.
- [47] S. Castano, A. Ferrara, S. Montanelli and G. Varese, "Ontology and Instance Matching.,," 2011.
- [48] S. Araujo, J. Hidders, D. Schwabe and A. de Vries, "SERIMI – Resource Description Similarity, RDF Instance Matching and Interlinking," 2011.
- [49] T. Jie, L. Juanzi, L. Bangyong, H. Xiaotong, L. Yi and W. Kehong, "Using Bayesian Decision for Ontology Mapping," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. Volume 4, pp. 243-262, 2006.
- [50] Y. R. Jean-Mary, E. P. Shironoshita and M. R. Kabuka, "Ontology Matching with Semantic Verification," *Web Semantics*, vol. 7, no. 3, pp. 235-251, September 2009.
- [51] J. Volz, C. Bizer, M. Gaedke and G. Kobilarov, "Silk – A Link Discovery Framework for the Web of Data," in *Workshop about Linked Data on the Web (LDOW2009)*, Madrid, Spain., 2009.
- [52] S. A. Axel-Cyrille Ngonga Ngomo, "LIMES — A Time-Efficient Approach for Large-Scale Link Discovery on theWeb of Data," Germany, 2011.
- [53] W. W. Cohen, P. Ravikumar and S. E. Fienberg, "A Comparison of String Distance Metrics for Name-Matching Tasks.,," Acapulco, Mexico, 2003.
- [54] M. R. Cássia and J. Euzenat, "Ontology Matching Benchmarks: Generation and Evaluation," 2011.
- [55] M. Ehrig, R. G. Castro and J. Euzenat, "Towards a methodology for evaluating alignment and matching algorithms".
- [56] C. Xie, "Semantic Similarity-Based Ontology Alignment for Enterprise Ontologies.,," 2009.

- [57] M. Mao, "Ontology Mapping: An Information Retrieval and Interactive Activation Network Based Approach," Berlin, Heidelberg, 2007.
- [58] P. Avesani, F. Giunchiglia and M. Yatskevich, "A Large Scale Taxonomy Mapping Evaluation," in *International Semantic Web Conference*, 2005.
- [59] Y. Sure, S. Staab and R. Studer, "Ontology Engineering Methodology," in *Handbook on Ontologies*, Springer-Verlag, Berlino, 2009, pp. 135-152.
- [60] S. K. Semy, M. K. Pulvermacher and L. J. Obrst, "Toward the use of an upper ontology for U.S. government and U.S. military domains: An evaluation," MITRE, Bedford, 2004.
- [61] IEEE, "Standard Upper Ontology Working Group (SUO WG)," IEEE, [Online]. Available: <http://suo.ieee.org/>. [Accessed 13 Novembre 2012].
- [62] IT Consult, "LA CONOSCENZA NELLE ORGANIZZAZIONI," 2005. [Online]. Available: http://www.itconsult.it/knowledge/white_papers/pdf/itc_WP_Conoscenza_Aziendale_Cap1.pdf.
- [63] M. Verlengia, "Knowledge Management in ambito sanitario con tecnologie di tipo Semantic Web," 2006. [Online]. Available: <http://airtlab.dii.univpm.it/sites/default/files/theses/verlengia-marco-thesis.pdf>.
- [64] I. Nonaka and H. Takeuchi, The Knowledge Creating Company - how Japanese Companies Create the Dynamics of Innovation, Oxford: Oxford University Press, 1995.
- [65] S. Codling, Best Practice Benchmarking, Gower, 1992.
- [66] R. C. Camp, "Best Practice Benchmarking: the Path to Excellence," vol. 04, 04 2003.
- [67] SDC Knowledge and Learning Processes Division in cooperation with Agridea Lindau, "Identifying and sharing Good Practices," 2009. [Online]. Available: http://www.fsnnetwork.org/sites/default/files/knowledge_management_toolkit.pdf.
- [68] D. J. Skyrme, "Are Your Best Practices Really the Best?," 2002. [Online]. Available: http://www.skyrme.com/updates/u54_f1.htm.
- [69] Accademia Italiana del Sei Sigma, "La metodologia Sei Sigma," [Online]. Available: <http://www.aiss.it/>.
- [70] The Open Group, "About TOGAF," [Online]. Available: <http://www.opengroup.org/togaf/>.
- [71] TM FORUM, "Business Process Framework," [Online]. Available: <http://www.tmforum.org/BusinessProcessFramework/1647/home.html>.
- [72] Best Management Practice Partnership, "An Introductory Overview of ITIL® V3," [Online]. Available: http://www.best-management-practice.com/gempdf/itSMF_AnIntroductoryOverview_of_ITIL_V3.pdf.
- [73] QualitiAmo, "Kaizen: Applicazione in Toyota," [Online]. Available: <http://qualitiamo.com/miglioramento/kaizen/toyota.html>.
- [74] S. Dreyfus and H. Dreyfus, "A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition," 2 Febbraio 1980. [Online]. Available: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA084551&Location=U2&doc=GetTRDoc.pdf>. [Accessed 25 07 2012].
- [75] D. North, "Better Best Practices," 11 Marzo 2008. [Online]. Available: <http://www.infoq.com/articles/better-best-practices>.
- [76] P. Valtchev and R. Missaoui, "Building Concept (Galois) Lattices from Parts: Generalizing the Incremental Methods.,," 2001.
- [77] D. O. S. K. D. Van der Merwe, "AddIntent: A New Incremental Algorithm for Constructing Concept Lattices.,," 2004.
- [78] D. C. Blair, "Information Retrieval, 2nd ed. C.J. Van Rijsbergen. London: Butterworths; 1979: 208 pp. Price: \$32.50," *Journal of the American Society for Information Science*, vol. 30, no. 6, pp. 374-375, 1979.
- [79] J. H. D. S. A. P. d. V. Samur Araujo, "SERIMI – Resource Description Similarity, RDF Instance Matching and Interlinking".
- [80] G. Fenza and S. Senatore, "Friendly web services selection exploiting fuzzy formal concept analysis.," *Soft Comput.*, vol. 14, no. 8, pp. 811-819, 2010.