Pedro Oliveira e Costa Gomes de Oliveira

Licenciado em Ciências da Engenharia Eletrotécnica e de Computadores

A Knowledge-Based Framework to Facilitate E-training Implementation

Dissertação para obtenção do Grau de Mestre em Engenharia Eletrotécnica e de Computadores

Orientador: Doutor Ricardo Luís Rosa Jardim Gonçalves, Professor Associado com Agregação da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Co-Orientador: Doutor João Filipe dos Santos Sarraipa

Júri:

Presidente: Doutora Anikó Katalin Horváth da Costa

Arguente: Doutor João Francisco Alves Martins

Vogal: Doutor Ricardo Luís Rosa Jardim Gonçalves

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março 2014

A Knowledge-Based Framework to Facilitate E-Training Implementation

To my Parents, Sister and Dog

ACKNOWLEDGEMENTS

The closing of this work marks the most important day of my academic life so far, and no doubt has a lot of meaning for me. However, this is a personal achievement, it was not possible without the unconditional love and support of my parents, Ana Cristina and Rui, that in all those years never stopped believing in me and gave me the chance to be who I am, so this work is dedicated to them in the first place. I also want to thank and dedicate this document to my sister, Marta, for believing in my capabilities and the support she gave me all her life. To my grandfather Arménio for being my inspiration to become an Engineer, and to my grandmother Lídia for never let me study on an empty belly.

After my family, I want to thank my supervisors, João Sarraipa for a great guidance, and for being persistent especially at the final phase, without him this work was even harder, and Professor Ricardo Gonçalves for the opportunity to do this work at GRIS, and for all his advices. I also want to do a special thank to Catarina Lucena. This dissertation is also dedicated to my friends. Bruno Caixinha for being my closest friend and have accompanied me during the course. Raquel Melo for being so special to me and have helped me in various situations. To my godson Baltasar that puts me a smile in the face all the time, Inês Pereira, Jorge Dias, Mafalda Lancinha, Nádia Sousa and Xavier Pereira. Along my years in FCT I met a group of real friends that also mean a lot to me, a special thank to you, David Aleixo, Diogo Pinto, Hugo Pereira, Nuno Barreira, Pedro Bueno. To my dearest friends Ana Castelão, Ana Cruz, Inês Pinheiro, Joana Urbano, Liliana Marinho, Mariana Borges and Telma Barroso. And also to all of you that somehow contributed to this work.

ABSTRACT

Nowadays, there is an evident increase of the custom-made products or solutions demands with the objective to better fits to customer needs and profiles. Aligned with this, research in e-learning domain is focused in developing systems able to dynamically readjust their contents to respond to learners' profiles demands. On the other hand, there is also an increase of e-learning developers which even not being from pedagogical curricula, as research engineers, needs to prepare e-learning programmes about their prototypes or products developed. This thesis presents a knowledge-based framework with the purpose to support the creation of e-learning materials, which would be easily adapted for an effective generation of custom-made e-learning courses or programmes. It embraces solutions for knowledge management, namely extraction from text & formalization and methodologies for collaborative e-learning courses development, where main objective is to enable multiple organizations to actively participate on its production. This also pursues the challenge of promoting the development of competencies, which would result from an efficient knowledge-transfer from research to industry.

KEYWORDS

Knowledge Acquisition, *Ontology*, e-*Training*, *and Learning management approaches*.

RESUMO

Hoje em dia, há um aumento evidente dos produtos ou soluções sob medida com o objetivo de melhor se adaptarem às necessidades dos perfis de clientes. Alinhado a isso, a pesquisa no domínio e-learning está focada em sistemas capazes de reajustar os seus conteúdos de forma dinâmica para responder à procura dos perfis de alunos. Por outro lado, há também um aumento de programadores e-learning que mesmo não sendo de currículos pedagógicos, como engenheiros de investigação, deveram preparar os programas de e-learning sobre os seus protótipos ou produtos desenvolvidos. Esta tese apresenta uma estrutura baseada no conhecimento, com o objetivo de apoiar a criação de materiais de e-learning, que serão facilmente adaptados para uma geração efetiva de cursos ou programas de e-learning feitos sob medida. Abrange soluções para a gestão do conhecimento, ou seja, a extração de texto e formalização, e metodologias de e-learning colaborativo de cursos de desenvolvimento, onde o principal objetivo é permitir que várias organizações participem ativamente na sua produção. Assim também persegue o desafio de promover o desenvolvimento de competências, o que resultaria num conhecimento eficiente de transferência da investigação para a indústria.

PALAVRAS-CHAVE

Aquisição do Conhecimento, *Ontologia*, “e-training“ e *Aproximações de gestão de aprendizagem.*

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACRONYMS

AI	**-**	Artificial Intelligence
HTML	**-**	Hypertext Markup Language
ITS	-	Intelligent Tutoring System
KA	**-**	Knowledge acquisition
KB	**-**	Knowledge Base
KR	**-**	Knowledge Representation
KRE	**-**	Knowledge Representation Element
KS	**-**	Knowledge System
LKOs	**-**	Learning Knowledge Objects
LMS	**-**	Learning Management Systems
LO	**-**	Learning Objects
NG	**-**	Nominal Groups
OBIE	**-**	Ontology-based extraction
OM	**-**	Organizational memory
OWL	**-**	Web Ontology Language
SCORM	-	Sharable Content Object Reference Model
SV	**-**	Semantic Vector
SVKB	**-**	Semantic Vector Keyword-based
SVOB	**-**	Semantic Vector Ontology-based
SVTB	**-**	Semantic Vector Taxonomy-based
SWRL	**-**	Semantic Web Rule Language
TM	**-**	Topic Map
TU	**-**	Training Unit
XML	**-**	Extensible Markup Language
XSTM	**-**	XML Specification for Topic Maps

1. INTRODUCTION

Nowadays with the globalisation phenomenon, companies are pushed to improve their strategies towards deconstruction and in focus on core competencies, giving rise to the concept of distributed virtual enterprises [1]. They are alliances of organizations that come together to share skills or core competencies and resources in order to better respond to business opportunities [2]. This increase demand of collaborative working environments, have required an high support for workers flexibility, which conducted to the necessity of having an efficient training programme implementation, available online (e-training) [3].

E-training is considered as a sub-set of e-learning, but more focused in a more practical training objective, as in for "how to use/implement" technical solutions through online mainly to industry. E-learning or e-training make use of Internet technologies to enhance knowledge transfer to enterprises workers, facilitating intra and inter enterprises knowledge evolution. These learning technologies offer learners or trainees control over content, learning sequence, pace of learning, time, and often media, allowing them to tailor their experiences to meet their personal learning objectives [4]. Other main feature that e-learning or e-training provides is related to the dynamic essence that their learning environments (Learning Management Systems – LMS) offer, namely, the capability for continuously updating of contents [5].

The author intends to contribute to this area with an e-training development approach, which main objective is to facilitate the training implementation in research projects. Such training programmes goal is to facilitate the knowledge transfer to the project stakeholders. Although, author in his recent experience in developing training in such projects, have faced some difficulties in pushing the project participants (engineers and researchers) to produce good training courses or programmes. Thus, the author has proposed a training building framework to serve as a guideline to e-training courses development, covering from information extraction to e-learning course creation. As a result of this framework a methodology was created, which objective is to organize training development in a specific way that could facilitate further adaptive e-training programmes creation functionalities or services.

The proposed methodology, which is presented in this thesis, intends to help in the creation of training materials in a structured organised and easier way, enabling various organisations members to actively participate on its production. Such proposed training organisation is accomplished through the use of ontologies to model its knowledge representation that enable the reasoning over the developed training materials facilitating the creation of web services able to provide adaptable training programmes for specific trainees purposes or profiles.

This thesis starts by presenting a state of the art, from which it was made an analysis to help in definition of the approaches and technologies customizations, to implement the framework, methodology and prototype developed. Then, the proposed training development framework supported by qualitative information collection methods analysis is presented. Such methods were introduced to help handling discussions between the training authors to reach a common and agreed view of the training implementation. Based on the framework defined the author developed a training implementation methodology that is also presented. This methodology was tested and implemented in a prototype application, which used an extension of a WIKI for particular customizations to better meet the thesis objective. Afterwards, its introduced knowledge is exported to an ontology, which objective is to characterize all the training to enable the generation of advanced services as adaptable training programme services. Then the thesis presents training courses validation, execution and analysis, topics. Finally it is described as a demonstration a real training implementation example that followed the proposed methodology in a European research project.

1.1. Motivation

The main motivation of this dissertation work is the fact that companies are being pushed to improve their strategies and products, which require an effective e-training solution able to support an effective knowledge transfer to their workers that would develop their skills and competencies.

1.2. Research Method

The classical research method was taken on consideration for this dissertation. This method has seven steps, starting on a more theoretical view and progressing to a more practical view of the system. It begins by studying the problem and defining the area of research and ends up with the proof-of-concept, i.e. the tests and analysis made to its results. This methodology is iterative, meaning that if the results are not what the researcher was expecting for, it is practical to go back to the first steps and try a new approach.

Figure 1.1 - Phases of the Classical Research Method [6]

The method is based on the following steps:

1. **Research Question / Problem:** This is the main step for a research. It is a period of study that intends to define the area of interest of the research. Even though the research statement is not declarative like a hypothesis, the research question must be clearly defined, making the study feasible and there is possibility of being confirmed or refuted. Generally the research question comes together with several minor questions to complement and narrow the focus of the main idea of the research subject.

2. **Background / Observation:** Is the study of prior work on the subject. This is accomplished by reviewing literature and scientific projects bringing up the ideas of other authors and what was already tested and accomplished taking the readers to the start point of the dissertation. The state of the art is at the same time, an important study to be made, reviewing literature and scientific projects having a big variety of documents for searching information on the area of interest, since some of the literature although very reliable, can be outdated and on the other hand, some documentation can be recent and have very innovative ideas but low reliability.

3. **Formulate Hypothesis:** The researcher formulates the hypotheses in order to make the research problem simpler to understand and to manage the predictions for the results of the research work. The hypothesis serves to bring clarity, specificity and focus to the research problem and define the desirable outcomes.

4. **Design Experiment:** This step works as a detailed planning of the experimental phase, where the solution design is seen as the system architecture. In addition, it is also significant to find a plan for validation which can be replicated by others in a feasible way.

5. **Test Hypothesis:** This step is required for evaluating the hypothesis and comprehends the implementation of the designed prototype and the evaluation of the obtained results. To test the hypothesis it is needed to get the results from system architecture and evaluate them. These outcomes are supposed to be collected for later analysis.

6. **Interpret / Analyse Results:** after all tests were done and the data output is collected it is the time to evaluate and analyse the achieve results. It is at this point that the veracity and confidence in the hypothesis are put to the test. Several outcomes are possible to happen, the results can be satisfactory, proving the author was right, or just fail

completely the initial idea. When positive results are obtained, then it is reasonable to say that a good prevision was made and it is possible to consider the future research and provide recommendations. But even the results are not the expected it should not be taken as a failure, but rather a way to improve the original approach with a new expertise on the subject going back again to the first steps of Phases of the Classical Research Method as shown in Figure 1.1, trying a different approach from the taken before.

7. **Publish Findings:** Results must end up in valuable contribution to the scientific community as scientific papers, either they are in line of the original hypothesis or against it. Accordingly to the type of research, scientific papers should be written to present intermediate results, These papers can be then presented in conferences, where the author has the chance to show in person his ideas for the research, presenting the results and answer questions of other researchers to prove the efficiency of the results and finalized with a dissertation about the hypothesis.

1.3. Research Questions and Problems

Can specific collaborative knowledge acquisition approaches facilitate knowledge transfer from research projects?

1.4. Hypothesis

The knowledge transfer from research can be facilitated if a proper framework with the support of knowledge extraction and modeling features to handle the building of structured courses is used to guide the training implementation.

1.5. Dissertation Outline

The first section of this dissertation, the Introduction, is the place where the main ideas that conducted to the study for this research project. According to what was done prior to this project work some new ideas and solutions are thought in being tested as a way of giving another step in the right direction for solving the research problem. It also manages the expectations on the chosen approach when it comes to analyse the results.

The next section, Knowledge Representation, is the topics that talk about the background observation work. In these section state what was done previously to this dissertation study. The Knowledge Representation section covers several topics with high significance for this dissertation, covering the main ideas of knowledge and knowledge representation, the different knowledge acquisition methods and e-learning creation. Proposed Framework and Training Development Methodology, sections present the proposed framework and the used methodology for assisting in course building. It is where the proposed solution for the research problem is presented, and explains the proceedings from knowledge acquisition to learning course creation.

After, the Prototype Implementation supported by an application using WIKI as a platform is presented. Finally, the section of Conclusions and Future Work topics are presented.

2. KNOWLEDGE REPRESENTATION

In the Oxford English Dictionary Knowledge is defined as facts, information, and skills acquired by a person, through experience or education; the theoretical or practical understanding of a subject. It is the awareness or familiarity gained by experience of a fact or situation [7]. Knowledge is also used to represent the confident understanding of a subject through the ability to use it for a specific purpose if appropriate.

Knowledge is the proper compilation of information, and its intent is to be useful. When someone memorizes information, they generate amassed knowledge. This kind of knowledge has useful meaning to the person itself, but it does not provide an integration such as would infer further knowledge. This is not the meaning of knowledge. Knowledge pursues the gathering of new knowledge in a loop cycle kind. Knowledge acquisition (KA) is a method of learning. KA involves complex cognitive processes: perception, learning, communication, association and reasoning. Its main objective is to transform tacit in explicit knowledge, and effectively to improve the approach to elicit knowledge from domain experts, towards interoperable intelligent systems.

Tacit knowledge is knowledge that people carry in their minds, all those things that they know how to do but perhaps do not know how to explain, which provides context for people, places, ideas, and experiences.

Explicit knowledge is knowledge that has been or can be articulated, codified, shared and stored in certain media. Is the opposite of tacit knowledge.

Knowledge representation studies the formalisation of knowledge and its processing within machines. Techniques of automated reasoning allow a computer system to draw conclusions from knowledge represented in a machine-interpretable form [8].

A Knowledge Representation Element (KRE) facilitates the formal representation of knowledge in a specific domain. Figure 2.1 illustrates the KRE's that should be defined in the path to build a domain's knowledge base. It represents the distinct level of conceptualisation that each one has, showing an increase of its presence from Terminology to the Knowledge Base.

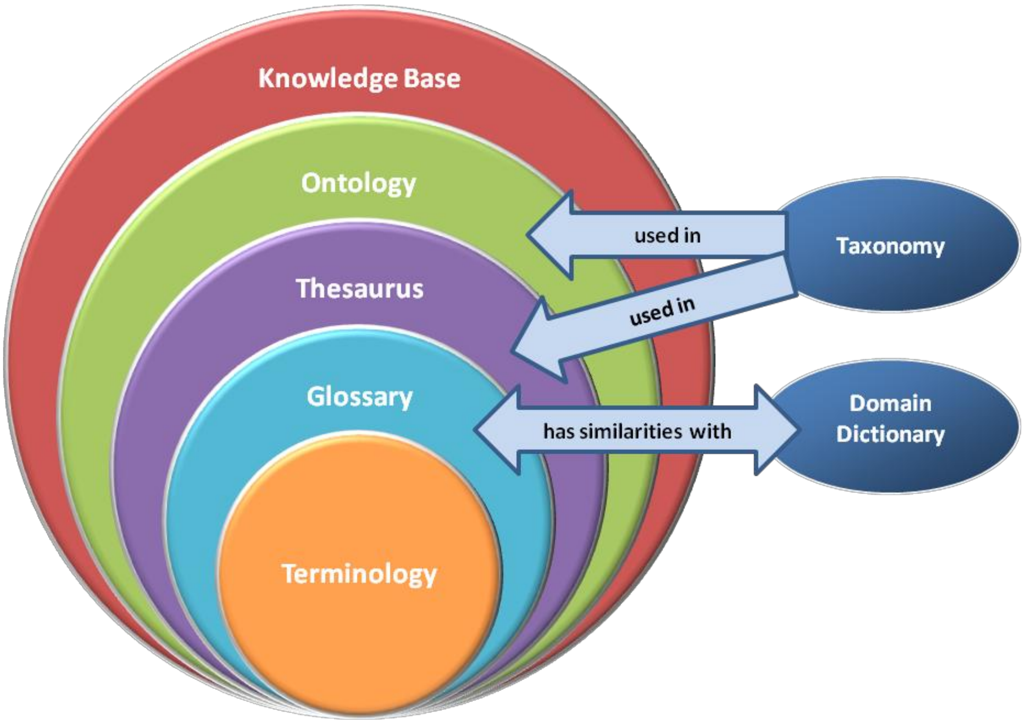

Figure 2.1 - Knowledge Representation Elements [9]

In the following, the knowledge representation elements shown in Figure 2.1, i.e. Terminology, Domain Dictionary, Glossary, Taxonomy, Thesaurus, Ontology, and Knowledge Base, are discussed.

2.1. Knowledge Representation Elements

Terminology is the study of terms and their use. Terms are words and compound words that are used in specific contexts. Terminology therefore denotes a more formal discipline which systematically studies the labelling or designating of concepts particular to one or more subject fields or domains of human activity, through research and analysis of terms in context, for the purpose of documenting and promoting correct usage[10].

Domain dictionary has been found to be one of the most useful tools for a domain analysis. The dictionary lessens a great deal of miscommunication by providing users with information: 1) in a central location to look for terms and abbreviations that are completely new; 2) where definitions of terms are used differently or in a very specific way within the domain [11].

Glossary is a list (normally in alphabetic order) of specialized terms that sometimes are exclusive to a specific subject. Each term is compound by its corresponding description. It includes descriptive comments and explanatory notes, such as synonyms, definitions, references, etc. A glossary can be used when there is reporting, in order to unify knowledge sharing.

Taxonomy is a classification system to categorize all the information in a class/subclass relationship, representing a simple tree. The root node represents the most general category that the domain is related to. There is a taxonomy for any category and each subcategory. Each child is a subset of the parent. Branch is the path from the root to a leaf. For users to navigate intuitively each category division should be consistent with their expectations.

Thesaurus is a structure that organizes concepts into a classification hierarchy. Establishing a formal lexicon of a specific domain it's is main objective. Taxonomy of domain concepts composed by its reference meanings is like a thesaurus.

An ontology is a formal, explicit specification of a shared conceptualization [12]. 'Conceptualization' is an abstract, simplified view of some phenomenon in the world that we wish to represent for some purpose. 'Explicit' means that the kind of concepts used and the constraints on their use are defined in an explicit way. 'Formal' reflects the notion that the ontology should be machine understandable, and can be translated into some form of logic. The 'specification' takes the form of definitions of representational vocabulary (classes, relations, and onwards), which provide meanings for the vocabulary and formal constraints on its coherent use. And finally, 'Shared' refers to the fact that ontologies capture consensual knowledge that is not restricted to the knowledge view of some individual, but reflects a more general view, shared and accepted by a group. Ontologies can also be composed by definitions of concepts, relationships, and other distinctions for modeling a determined domain.

There is a wide range of uses for ontologies, from people, data bases, and software applications that have the need to share information, where their information domain is related to a particular area of knowledge. Ontologies include definitions of basic concepts used by computers that may contain relationships between them to enable the information organization in several domains..

In Artificial Intelligence (AI), Knowledge Representation (KR) – studies the formalisation of knowledge and its processing within machines. Techniques of automated reasoning allow a computer system to draw conclusions from knowledge represented in a machine-interpretable form [8].

2.2. E-Training Content Creation

This thesis main objective is to develop a framework that besides extracting knowledge from wikis, web, text documents or ontological sources in an automatic way, also can assist in building a training course with the information gathered.

Below, the author explains how this process can be achieved, and discuss with more detail each step, in order to achieve the e-training content.

2.3. Automatic Information/ Knowledge Extraction

Most of the information extracted, are gathered using ontologies or used to build an ontology. The main text information is on web and company documents. In these texts, many words have specific meanings (called semantic classes) such as a person's name, a location's name, an organization's name, and so on. By extracting these words, we it can effectively use the text information on a Q&A system and for text categorization, machine translation, and so on [13].

They have researched an information extracting method that gets good accuracy even if the data contains various kinds of data. Their clustered learning method makes clusters from training data and rules from each clustered training data using machine learning techniques. Their method can divide problem spaces like kinds of data by clustering and creates rules for each problem space. However, the clustered training data by only clustering has bias: the amount of training data in each cluster and the density in each problem space. To limit these effects, their method modifies the data of clusters. Specifically, the data in large and high density clusters are moved to another cluster.

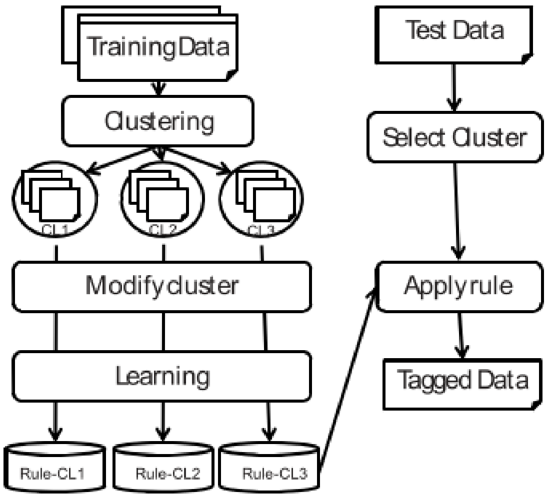

Figure 2.2 - Overview of clustered learning method [13]

In the context of ontology-based information extraction, identity resolution is the process of deciding whether an instance extracted from text refers to a known entity in the target domain (e.g. the ontology) [14]. They present an ontology-based framework for identity resolution which can be customised to different application domains and extraction tasks. The framework is intended to provide a general solution to the identity problem that can be used within different applications regardless of their particular domain or type of entity which need to be resolved. The input is an entity together with its associated properties and values, the output is an integrated representation of the entity which will have new properties and values in the ontology.

Ontology-based extraction (OBIE) is the process of identifying in text or other sources relevant concepts, proper- ties, and relations expressed in an ontology [14].

2.3.1. Knowledge from Web

Extraction of meaningful content from collections of web pages with unknown structure is a challenging task, which can only be successfully accomplished by exploiting multiple heterogeneous resources [15]. In the “Ex” information extraction tool, so-called extraction ontologies are used by human designers to specify the domain semantics, to manually provide extraction evidence, as well as to define extraction subtasks to be carried out via trainable classifiers. Extraction ontologies in “Ex” are designed so as to extract occurrences of attributes, i.e. standalone named entities, and occurrences of whole instances of classes, as groups of attributes that ‘belong together’, from HTML pages or texts in a domain of interest.

2.3.2. Knowledge from Documents

There are several types of documents from where knowledge can be extracted. A common type of one of those documents that is used by researchers and contains great amounts of knowledge and information is a deliverable. As seen in [16] the author as developed a tool that besides other functionalities, can by receiving a keyword based query give back a collection of deliverables that have a similarity in common with the keywords. The core contribution of this work is the enrichment process of knowledge representation. The overall approach comprises 5 stages (Figure 2.3): (i) pre-processing (preparation of the operational environment and input sources); (ii) ontology evolution (Enrich it with new concepts and relations); (iii) semantic enrichment (the enrichment process); (iv) classification (classify input KS into categories); and (v) evaluation (measure accuracy of the overall approach). For the purpose of this thesis the author will focus on the semantic enrichment stage.

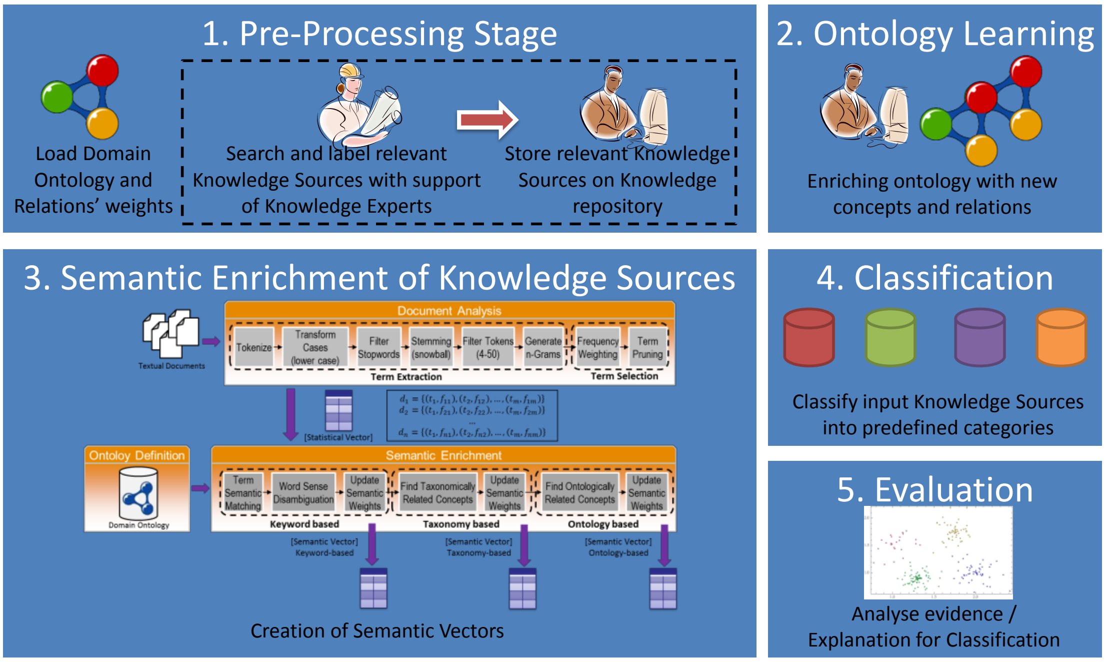

Figure 2.3 - General overview of enrichment process[16]

The semantic enrichment is the main focus of [17] work. Indeed, it focuses the enrichment of knowledge representations (which the author calls semantic vectors), also extends the classical vector space model approach by adding two additional steps to the process: (i) the use of taxonomical relations to improve semantic relevance of neighbours concepts; and (ii) the use of ontological relations for the same reason.

Semantic Vector Creation in the Document Analysis step, uses Rapidminer to calculate the term frequency - inverse document frequency scores for all terms, a stored procedure developed on MySQL reduces the size of the statistical vector according to a certain relevance degree defined by the knowledge expert, and another stored procedure normalize the statistical vector after pruning the terms [16]. Afterwards, the semantic enrichment is performed by three Java services responsible for the generation of the three types of vectors, the keyword, taxonomy and ontology-based, respectively.

The enrichment process is characterized by two steps: (i) *Document Analysis*: extracts terms from knowledge sources, constructs the key term set, and produce a statistical vector based on term occurrence; and (ii) *Semantic Enrichment*: reworks the statistical vector using taxonomical and ontological elements (such as relations, concepts weights) in order to produce a semantically richer KR, called Semantic Vector.

In the document analysis phase there are two processes running, namely *Term Extraction* and *Term Selection*, which reduce the statistical vector dimension, taking out less relevant terms.

In the semantic enrichment phase is where the semantic vectors (SVs) for all documents in corpus D are built. A statistical vector semantically richer that make use of the following ontological elements: concepts, relations, equivalent terms, and weights it is called SV. Each SV is represented by two columns: the first column contains the concepts that populate the knowledge representation of the KS, i.e., the most relevant concepts for contextualizing the information within the KS; the second column keeps the degree of relevance, or weight, that each term has on the knowledge description of the KS [16].

The author approach takes into account three complementary procedures to create SVs, where each procedure successively adds semantic richness to the Knowledge Representation. The first step creates a SV keyword-based (SVKB), the second step creates a SV taxonomy-based (SVTB), and the final step creates a SV Ontology-based (SVOB).

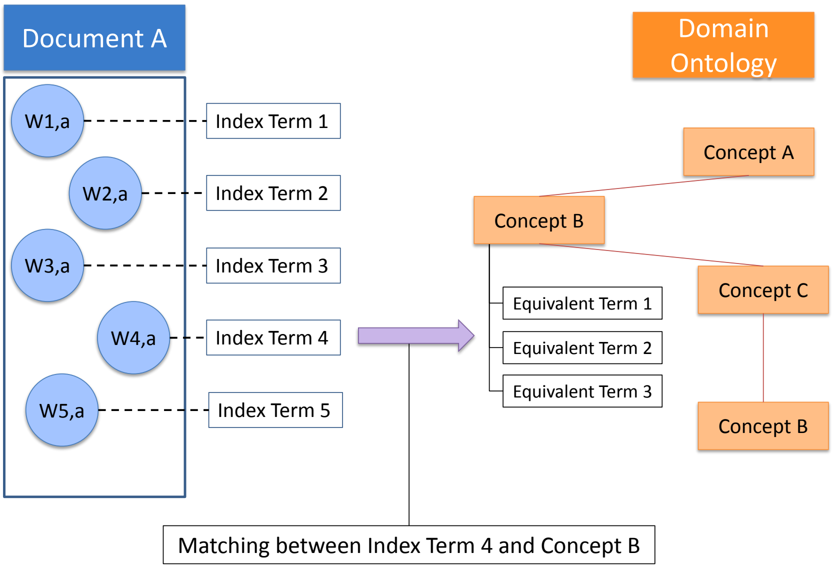

Figure 2.4 - Vector terms mapping against the Ontology concepts [16]

The keyword-based semantic vector takes into consideration only the bond between terms from the statistical vector and the concepts in the domain ontology. This step matches the statistical vector keywords with equivalent terms linked to each ontological concept in the domain Ontology (Figure 2.4). This process starts by identifying the statistical vector keywords associated to a particular document and afterwards finding similarities between each keyword and the equivalent terms within the ontology.

The semantic vector taxonomy-based is the next level in the semantic evolution of KRs. It is created by adjusting the weights of concepts according to the taxonomic relation among them. If the SVKB have two or more concepts taxonomically related, then the existing relation can increase the relevance of the expressions within the KR and consequently enhance weightings. SVTB is created based on kin relations between concepts within the ontological tree. Specifically, the kin relations can be expressed through the notion of homologous/non-homologous concepts (Figure 2.5).

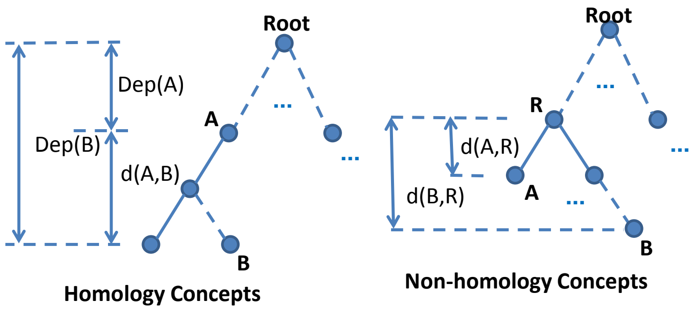

Figure 2.5 - Homologous and non-homologous concepts [18]

Using the keyword-based vector as input the SVTB is calculated, where taxonomical relations are used to enhance the importance of the concepts already present within the vector or to add new concepts. When two concepts found in the keyword-based vector are highly relevant the weight of the concepts is boosted, with the degree of relevance being defined by a given threshold.

Semantic vector ontology-based is the final stage in the semantic evolution of KRs, which is based on the ontological relations. As described in [17] the objective of applying an association rule theory to construct ontological concept relations and evaluate the relevance of such relations for supporting the enrichment process of a domain ontology, is to analyse the co-occurrences of concepts in unstructured sources of information in order to provide interesting relationships for enriching ontological structures.

The input from experts in the building and construction domain to establish the final numerical weights on each ontological relationship complements the ranking of such semantic association. Experts' intervention is an effort to assure that relevancies of relationships reflect a accurate knowledge representation requirement.

The creation of the SVOB is a two-stage process using the taxonomy-based SV as input: the first stage boosts weights of concepts already present in the taxonomy-based vector, depending on the ontological relations among them; the second stage adds new concepts that are not present in the input vector, according to ontological relations they might have with concepts belonging to the taxonomy-based vector [19].

2.3.3. Knowledge from Text

A TM (Topic Map) is a formalism to represent knowledge about the structure of an information resource and to organize it in topics. These topics have occurrences and associations that represent and define relationships between them. Information about the topics can be inferred by examining the associations and occurrences linked to the topic. A collection of these topics and associations is called a topic map. TM Builder, that is a processor that extracts topics and relations from instances of a family of XML documents [20]. A TM-Builder is strongly dependent on the resources structure. So, to extract a topic map for different collections of information resources (sets of documents with different structures) we have to implement several TM-Builders, one for each collection. They present a language to specify topic maps for a class of XML documents, that we call XSTM (XML Specification for Topic Maps).

2.3.4.Knowledge from Wiki

Wiki software is an ideal platform because it is easy for non-technical users to learn and many organizations are already using this technology [21].

In [21] the authors choose an extraction method that uses regular expressions to extract information. This means the required information can be extracted from the page independently of the rest of the structure of the page. One limitation of the Java implementation of Regular Expressions is that it does not support nested or recursive expressions.

It is a straightforward combination of existing Wiki systems and the Semantic Web knowledge representation paradigms. However, we see the following obstacles:

– Usability: The main advantage of Wiki systems is their unbeatable usability. Adding more and more syntactic possibilities counteracts ease of use for editors.

– Redundancy: To allow the answering of real-time queries to the knowledge base statements have to be stored additionally in a triple store. This introduces a redundancy complicating the implementation.

– Scalability: Knowledge base changes which involve statements with different subjects will scale very bad since all corresponding Wiki texts have to be parsed and changed.

The OntoWiki strategy, on the contrary, does not try to mix text editing with knowledge engineering, instead it applies the Wiki paradigm of “making it easy to correct mistakes, rather than making it hard to make them” [22] to collaborative knowledge engineering.

Wikipedia is based on the MediaWiki software. As an open-source project, its entire content is easily obtainable in the form of largeXML files and database dumps that are released sporadically every several days or weeks [23].

2.3.5.Knowledge from ontology sources

OntoLearn is a method and a tool aimed at the extraction of domain ontologies from Web sites, and more generally from documents shared among the members of virtual organizations. OntoLearn first extracts a domain terminology from available documents. Then, complex domain terms are semantically interpreted and arranged in a hierarchical fashion. Finally, a general-purpose ontology, WordNet, is trimmed and enriched with the detected domain concepts [24].

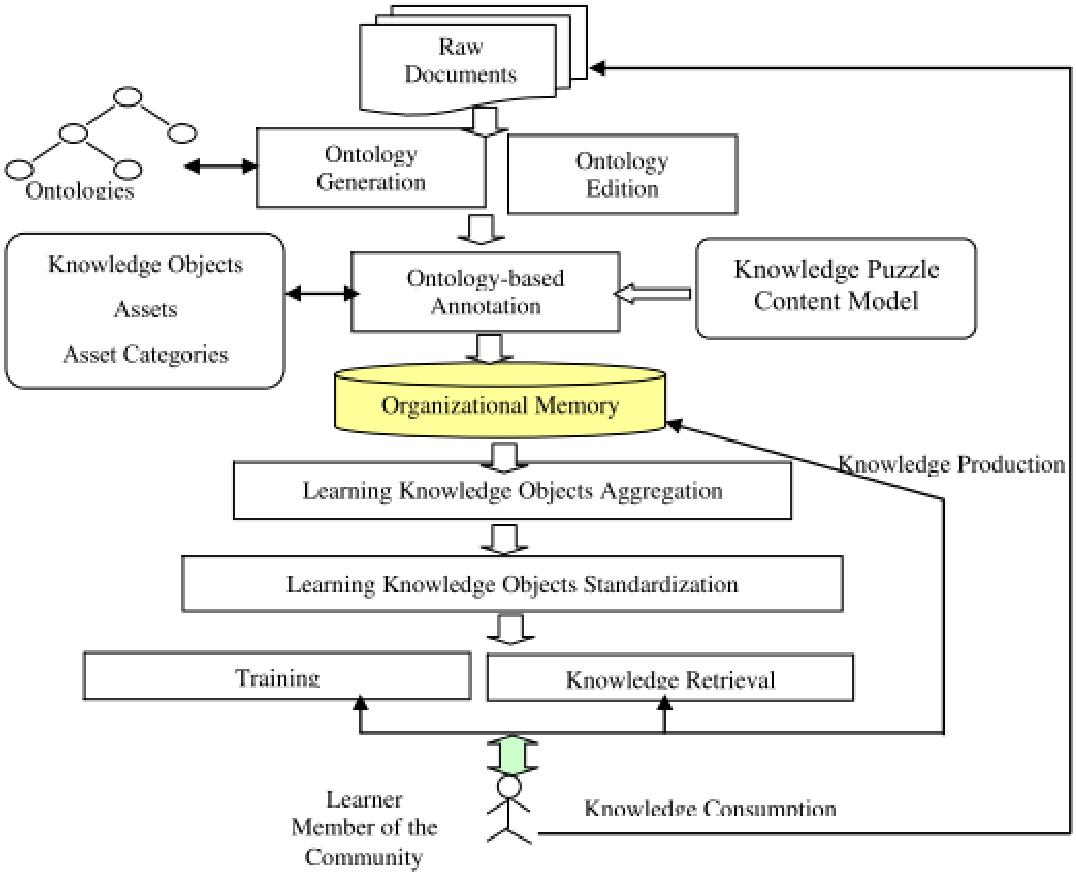

Figure 2.6 - The knowledge Puzzle Platform Functional View [25]

Knowledge Puzzle, seen in Figure 2.6, is an ontology-based platform that was designed to ease domain knowledge acquisition from textual documents for knowledge-based systems. First, the Knowledge Puzzle Platform performs an automatic generation of domain ontology from documents' content through natural language processing and machine learning technologies. Second, it employs a new content model, the Knowledge Puzzle Content Model, which aims to model learning material from annotated content. Annotations are performed semi-automatically based on IBM's Unstructured Information Management Architecture and are stored in an Organizational memory (OM) as knowledge fragments. The organizational memory is used as a knowledge base for a training environment (an Intelligent Tutoring System or an e-Learning environment). The main objective of these annotations is to enable the automatic aggregation of Learning Knowledge Objects (LKOs) guided by instructional strategies, which are provided through SWRL rules. Finally, a methodology is proposed to generate SCORM-compliant learning objects from these LKOs [25].

2.4. Automatic E-Learning Creation

E-learning courses can be created from different sources like, web, text, wiki or by the mean of ontologies. The Figure 2.7 can be seen as a way of e-learning creation.

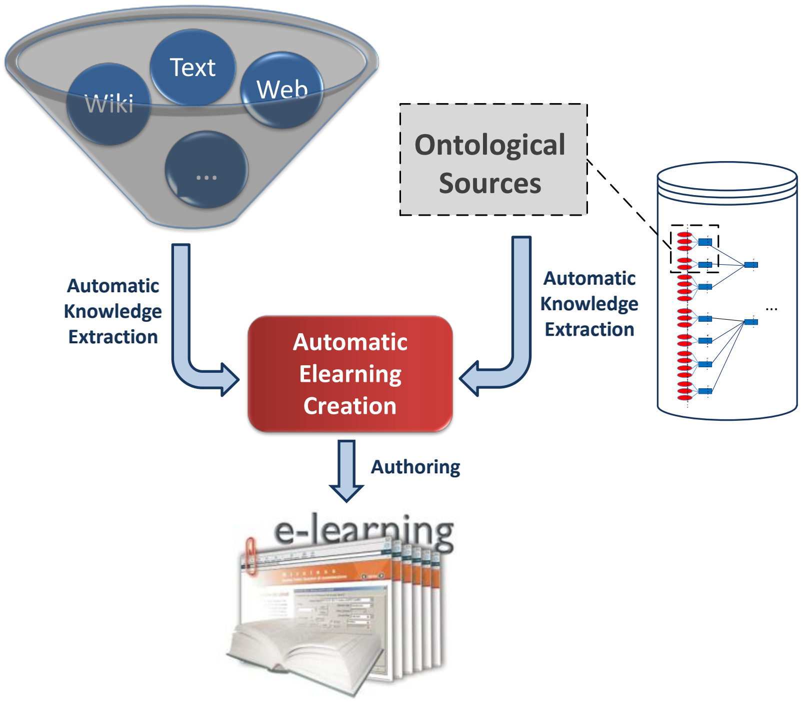

Figure 2.7 - An e-learning creation method

2.4.1.From Web

In all cases, information on the web is often not effectively organized and learners spend considerable time in futile interactions and may not properly integrate information to address their immediate learning need. A more flexible approach is needed that is sensitive to each learner's unique needs and context, but also provides focused and structured learning. A web service in an adaptive learning environment the reusability of the learning content can be ensured. Web services provide a solution to a major problem in the computer world i.e. interoperability. Interoperability is provided by allowing different applications from different sources to communicate with each other without time-consuming customized coding [26].

2.4.2.From Text

In [27], a tool was created that assists case authoring from texts, SmartCAT-T (Smart Case Authoring Tool for Tree structure cases).

The key phrase extraction technique relies on identification of a few key terms that are then used as anchors to obtain other key problem descriptors.[27] When extracting key phrases, when possible, background domain knowledge is used. When working in an unfamiliar domain, it is not possible to use seed words to obtain anchor terms. Instead, the author uses a frequency-based technique and syntactic information of the words in the documents, in order to automate as much as possible, the process of obtaining the anchors terms. Although only a few terms are used as anchors, SmartCAT-T is able to identify more phrases that do not have the anchor words as their substrings or even as part of the sentences in which the phrases appear.

Overall, Smarter performs reasonably well in this unfamiliar domain. Most of the false positives can be attributed to Smarter's inability to distinguish terms that appear similar but occur in different contexts in determining similarity scores as was done in the SmartHouse domain when terms' corresponding disability-terms were considered. The other contributing factor could be the fact that Smarter does not utilise importance scores of key phrases in the query and the activated concepts in the cases to determine similarity scores during problem-solving.

In conclusion, the reliance of statistical approaches on large volumes of data to extract features from text, and the fact that the approaches typically result in knowledge-poor representations, made them unattractive. Approaches based on Information Extraction systems were also ruled out because of the level of human intervention required to tune the information extractors.

The SmartCAT-T project strikes a good balance between the need to automate the system and the desire for a knowledge-rich case representation that comprises various levels of knowledge abstraction. This is achieved by using background domain knowledge to mitigate some of the effects of data sparseness in order to automatically extract key phrases. Key phrases are used to construct a conceptual model for the domain which in turn, is used to create knowledge-rich structured cases from the text and for query interpretation during problem-solving. A language model is also developed to enable the case authoring process and retrieval mechanism to meaningfully interpret text by allowing the recognition of word synonyms in text. The result is a case authoring tool that identifies important phrases in text and uses them to create hierarchically structured cases, and a case-based problem-solving tool for the corresponding domain. The major contributions are a conceptual model in which concepts are interpretable by a human, automatic acquisition of cases, a case based reasoning tool, unsupervised approach to Word Sense Disambiguation and generic techniques that assist case authoring from text.

Since the World Wide Web became popular and penetrated every ELT (English language teaching) classroom in the late 1990s, ELT teachers have turned to add value to their teaching by creating web sites or web pages for different purposes. When faced with the challenges of the eLearning era to build more interactive, mediarich pedagogically sound materials, ELT teachers are tempted to search for and learn to use different tools like HTML, XML, Flash ActionScript. Unfortunately, most ELT teachers, who are usually nontechnically minded, find these tools difficult to master and soon give up [28].

2.4.3.From Wiki

Wiki is a website which allows its users to add, modify, or delete its content via a web browser usually using simplified markup language or a rich-text editor. Wikis serve many different purposes, such as knowledge management and note taking.

Wikis seem very helpful if there is a need for a platform that enables online collaboration and the creation of a knowledge base that can be accessed and amended by any- one on the World Wide Web [29]. In fact, one of the first thoughts that probably come to mind in association with wikis is the free online encyclopedia Wikipedia, which has developed into a well-known online knowledge base based on wiki technology. Even though wikis were already in use long before the appearance of this free online encyclopedia, Wikipedia increased their popularity and made media pay attention to them.

In general, one could say that a wiki enables creating a website where every user can get involved [29].

We can define wiki as web-based software that allows all viewers of a page to change the content by editing the page online in a browser." This means that every user can obtain author and editor privileges, which makes the wiki truly democratic [30].

A wiki consists of several interlinked web pages that can incorporate text, sound, images and videos. Apart from that, each wiki can look different, because a wiki has no predefined structure. Wikis therefore represent a multi-facetted repository, which may at the same time serve as a space for asynchronous as well as synchronous collaboration or collaborative activities. The changes that are being made are shown instantly. Since every user can post, edit or delete contents a version control is usually implemented so that malicious or incorrect changes can be made reversible.

The number one reason why wikis are so popular in the first place is also the number one reason for others to react critically to the concept of wikis: openness [29]. The freedom of author and editor privileges, which also includes the right to delete or edit someone else's content, seems to be a major issue. For newcomers, the ingrained norm of authorship creates objections to wikis.

Despite the challenges, wikis are convenient tools for collaboration, collection and reflection and therefore are applied in numerous different ways at educational institutions [31].

They are considered a tool that "[…] facilitate[s] collaborative finding, shaping, and sharing of knowledge, all of which are essential properties in an educational context" [32].

One of the reasons why wikis are applied more frequently in educational institutions is the ease of use and the low cost. It is fairly easy to learn how to use a wiki, it is generally free to use for it is an open source application and does not take a lot of time to create new contents. Therefore the focus remains on the contents and not on the software itself.

The wiki can be used as a repository for learning resources, which would also enable the interdisciplinary addition, exchange and commenting of learning resources across institutions in a university context. This way the sustainability is ensured and the wiki can be used in different contexts. Thus, the sustainability of the wiki's content can serve as a source of information and knowledge.

A general objective of all E-Learning modules created is to improve the quality of teaching and to enhance the individual study as an additional support to the present teaching. In this respect, special attention is paid to the content development of the E-Learning modules in consideration of adding value facilitated by the integration of interactive multimedia. The E-Learning modules shall provide students with access to the learning resources independent of time and place and facilitate the process of learning from each other through knowledge sharing and knowledge creation.

2.5. Authoring Creation

The development of good quality educational software is expensive and time-consuming. Besides the usual challenges, developers face interdisciplinary and specific issues in personnel organization, instructional design, project and implementation, use and reuse by teachers, code maintenance, among others [33]. In order to develop an application framework tailored to build interactive learning modules (iLM), besides several specific characteristics, providing authoring tools for teachers is a requirement.

The definition of authoring tool is a program that helps you write using hypertext or multimedia applications and enable you to create a final application merely by linking together objects, such as a paragraph of text, an illustration, or a song. By defining the objects relationships to each other, and by sequencing them in an appropriate order, authors (those who use authoring tools) can produce attractive and useful graphics applications [34]. In most cases, there are numerous authoring tools that supports a scripting language for develop more sophisticated applications. Storing, modification, reusability and sharing information are the important features that make working with E-content easier that paper-content.

Authoring tools classification can be done based on different aspects such as complexity, fee, and purpose. Concerning complexity, the tools can be classified in range from simple to advance. Simpler tools are those which can support features as drag and drop facilities, wizard ...etc. Advance tools require programming capabilities and technical competency to build a course material. About the fee, tools can be free or commercial. Finally the purpose of some tools concern on creating courses, although there are some multipurpose tools that are being used while they aren't specialized for creating online courses.

The majority of authoring tools fall into two broad categories: the pedagogy-oriented systems which "focus on how to sequence and teach relatively canned content" and the performance-oriented systems which "focus on providing rich learning environments in which students can learn skills by practicing them and receiving feedback." [35]. There seven categories of ITS (intelligent tutoring system) authoring systems according to the type of ITSs they produce. These are: (i)Curriculum Sequencing and Planning; (ii)Tutoring Strategies; (iii)Device Simulation and Equipment Training, (iv)Expert Systems and Cognitive Tutors; (v)Multiple Knowledge Types; (vi)Special Purpose;(vii)Intelligent/Adaptive Hypermedia.

An ITS "shell" is a generalized framework for building ITSs, while an ITS "authoring system" (or authoring tool) is an ITS shell along with a user interface that allows non-programmers to formalize and visualize their knowledge [36].

CATEGORY	STRENGTHS	LIMITS	VARIATIONS
Curriculum Sequencing and Planning	Rules, constraints, or strategies for sequencing courses, modules, presentations	Low fidelity from student's perspective; shallow skill representation	Whether sequencing rules are fixed or authorable; scaffolding of the authoring process
Tutoring Strategies	Micro-level tutoring strategies; sophisticated set of instructional primitives; multiple tutoring strategies	(same as above)	Strategy representation method; source of instructional expertise
Device Simulation and Equipment Training	Authoring and tutoring matched to device component identification, operation, and troubleshooting	Limited instructional strategies; limited student modeling; mostly for procedural skills	Fidelity of the simulation; ease of authoring
Domain Expert System	Runnable (deeper) model of domain expertise; fine grained student diagnosis and modeling; buggy and novice rules included	Building the expert system is difficult; limited to procedural and problem solving expertise; limited instructional strategies	Cognitive vs. performance models of expertise
Multiple Knowledge Types	Clear representation and pre-defined instructional methods for facts, concepts, and procedures	Limited to relatively simple fact, concepts, and procedures; pre-defined tutoring strategies	Inclusion of intelligent curriculum sequencing; types of knowledge/tasks supported
Special Purpose	Template-based systems provide strong authoring guidance; particular design or pedagogical principles can be enforced	Each tool limited to a specific type of tutor; inflexibility of representation and pedagogy	Degree of inflexibility
Intelligent/ Adaptive Hypermedia	WWW has accessibility & UI uniformity; adaptive selection and annotation of hyperlinks	Limited interactivity; limited student model bandwidth	Macro vs. micro level focus; degree of interactivity

Table 1 - ITS Authoring Tool Strengths and Limitations by Category [37]

The appropriated choice of the content authoring tool depends on the requirements and where the content would be implemented. The common categories of authoring tools which produce content complying with the E-learning standards use some base software application. The most common ones are those that use PowerPoint presentations, Web based content and Flash based presentations [37].

The ITS Authoring Tool Strengths and Limitations by Category are shown in Table 1.

On the classification topic authoring tools can be Unspecialized, such as MS PowerPoint, Front Page, Dreamweaver and Flash where those who use these tools need to have a good command over programming to make an interactive course. Specialized authoring software such as Articulate, Adobe Captivate, ObjectJ and 4GLO concentrates on creating courses and Learning Objects (LO) without need for programming level. Some explanations about the software will be presented next.

MS PowerPoint is presentation software used to create slideshow. The slides are usually linear and can include hyperlinks to jump to other sections. This allows the user to get to more detail on a subject of interest. In addition, it allow user to add audio and video clips in a presentation. Even though this tool is mainly for creating presentation, it is considered to be one of the most popular course authoring tools [35].

Flash has become a popular method for adding animation and interactivity to web pages. Flash is commonly used to create animation, advertisements, and various web page Flash components, to integrate video into web pages, and more recently, to develop rich Internet applications. Therefore, some course authors prefer using flash, because it allows many features and programming using Action Script [35].

Front Page and Dreamweaver tools are used to create WebPages. They provide an easy way to create links and images without need to know about HTML. However, these tools alone cannot provide the interactive course. It needs the help of other tools and graphics designers [35].

Adobe presenter / Captivate Adobe Presenter help to easily create professional Flash presentations and self-paced courses complete with narration. In addition, it allows adding animations, quizzes, and software simulations to E-learning courses. Similar to articulate, this tool is also based on MS PowerPoint and can rapidly author professional E-learning content with interactivity, simulations, quizzes, and other experiences no programming skills required. However, Captivate does not need MS PowerPoint because it is independent [35].

The E-learning authoring tool is a simple, easy to use tool for creating online course content. Its design allows you to create rich media courses containing text plus images, Flash animations, audio and video created with your favourite media editors. Every course you create will fully conform to the SCORM standard. It is possible to quickly create courses for employees, customers, partners, or even the general public. The learners will enjoy an easy to navigate course with a clean, professional design. The trainer will not need any programming skills to create courses with the AT. You author your course with easy to use forms and a built-in HTML editor. When you deliver your course or test with a SCORM-based LMS, the questions are automatically graded and the results sent back to the LMS. The Authoring Tool reports the overall score (raw, minimum and maximum), the course status (incomplete, completed, passed or failed) and the results of individual questions[35].

The Advanced E-learning Builder authoring tool is designed for creating e-learning materials such as e-tests, tutorials, quizzes, etc. The system enables the creation of E-tests, just with several mouse clicks (visual design), which contain several types of exercises. The Advanced E-learning Builder authoring tool creates Standalone EXE files. By planning exercises in visual mode, exercises can be designed, corrected, examined, complemented, and updated in a simple and survivable way. Various elements such as gap-fill tasks, multiple choices, and alternative questions are used to set tests questions. Additionally, many other elements such as various types of texts, graphics, multimedia, Object Linking and Embedding - buttons ensure that your test has an attractive look and professional feel. And, all these elements can be easily moved and resized using a mouse. Elements offer various customization possibilities including changing colours, fonts, and contents; thus, creating and modifying all kinds of e-tests is quick and easy. Visual design mode allows the building of E-learning materials in a simple and comprehensive way. Therefore, you can always see exactly how the final test to be deployed to students will look. Additionally, built-in dynamic resolution assures you that a test will automatically adjust to all possible screen resolutions that end-users may have so the look won't be deformed in any way. Advanced E-learning Builder also contains many pre-defined test templates that allow you to make professional looking tests in no time at all. This way, creating e-tests is as easy as editing text in a word processor, which allows beginners to start using the application immediately. An additional feature of advanced E-learning Builder is test flow control, which allows you to write teaching sequences that can adjust to the students' knowledge and supply students with additional information when needed. This way, individualization is granted. All tests created with advanced E-learning Builder are finally graded and supplied with a customizable commentary. The package also contains a real world example that can be used in the class with, or without, customization [35].

Web-based collaborative authoring environments are intended as repositories of encyclopedia knowledge. Two examples of this latter trend are Wikipedia and Everything2. Wikipedia is a wiki authoring environment designed for the purpose of creating a user-written encyclopedia containing information on all subjects. Everything2 is a web-based community bulletin board designed to create, organize and store information about "everything." [38].

2.6. E-learning Execution

The e-learning courses must be applied in a way that the users can find it easy and accessible.

2.6.1.Standards

In [39] there are some organizations working to develop standards like IMS, ADL, ARIADNE, IEEE, ISO, with the objective to provide frame work for e- Learning architectures, to facilitate interoperability, content packaging, content management, Learning Object Meta data, course sequencing and many more.

PROMETEUS - Telematics, knowledge content, and multimedia- based tools are widely considered central ingredients for evolving new ways to provide learning and training. PROMETEUS an acronym for PROmoting Multimedia access to Education and Training in EUropean Society established with a clear underlying ideal to promote access to knowledge, education and training for all European citizens, regardless of their age, work situation, geographical location or social status.

PROMETEUS reach for: optimal strategies for multicultural, multilingual learning solutions; new instructional and training approaches and new learning environments; affordable solutions and platforms based on open standards and best practices; publicly accessible and interoperable knowledge repositories.

2.6.2.Tools

MediaWiki is a free and extremely powerful, scalable software and a feature-rich wiki implementation, that uses PHP to process and display data stored in a database such as MYSQL, for running web-based wikis originally intended for Wikipedia and others like Wiktionary. It's designed to be run on a large server farm for a website that gets millions of hits per day.

Pages use MediaWiki's wikitext format, so that users without knowledge of XHTML or CSS can edit them easily. When a user submits an edit to a page, MediaWiki writes it to the database, but without deleting the previous versions of the page, thus allowing easy reverts in case of vandalism or spamming. MediaWiki can manage image and multimedia files, too, which are stored in the filesystem. For large wikis with lots of users, MediaWiki supports caching and can be easily coupled with Squid proxy server software.

Semantic MediaWiki (SMW) is a free open-source extension to MediaWiki, which is able to store and query data within the wiki’s pages and helps to search, organise, tag, browse, evaluate, and share the wiki's content. While traditional wikis contain only text which computers can neither understand nor evaluate, SMW adds semantic annotations that allow a wiki to function as a collaborative database.

Although wikis are very good for storing and retrieving individual facts, they are less useful for getting queried or aggregated information.

SMW have some benefits like automatically-generated lists, visual display of information, improved data structure, searching information, external reuse and integrate and mash-up data.

Semantic MediaWiki is being used in hundreds of sites, in many languages, around the world, including Fortune 500 companies, biomedical projects, government agencies and consumer directories. There are a growing number of consulting companies that implement SMW as part of their solutions, with some stating their use of MediaWiki and SMW explicitly and others keeping it as a hidden implementation detail.

2.7. Ontology Construction and Learning

Many interesting papers exist regarding ontology learning from text and explain the general process of such a generation such as [40] and [41] . In [40] the authors describe the ontology generation layers as consisting of six extraction layers of growing complexity: terms, synonyms, concepts, taxonomy, relations and rules. Other systems implement specific algorithms and present case studies and interesting results such as ASIUM [42], TextToOnto [41], Ontolearn [43] and OntoLT

[44]. Those papers explain the general process regarding ontology learning. For example, in [42] the authors implemented a system called ASIUM that uses an unsupervised method based on syntactic parsing to acquire sub-categorization frames of verbs and ontologies.

In TextToOnt, a collection of domain documents is annotated with NLP tools to extract a number of occurring terms. An association rules algorithm then finds correlations in the co-occurrence of classes of terms, and the system identifies possible relations between these terms. Finally, the system represents these terms and relations as classes in the ontology [41].

OntoLT is an interesting project because it provides a plug-in for the Protégé ontology development environment. It defines a number of linguistic patterns to map Protégé classes and slots to annotated texts [44].

Finally, in [45] the authors present Text2Onto, a framework to generate consistent OWL ontologies from learned ontology models by representing the uncertainty of the knowledge in the form of annotations. These annotations capture the confidence about the correctness of the ontology elements. They generate ontologies based on a Learned Ontology Model (LOM), which is then transformed into a standard logic-based ontology language.

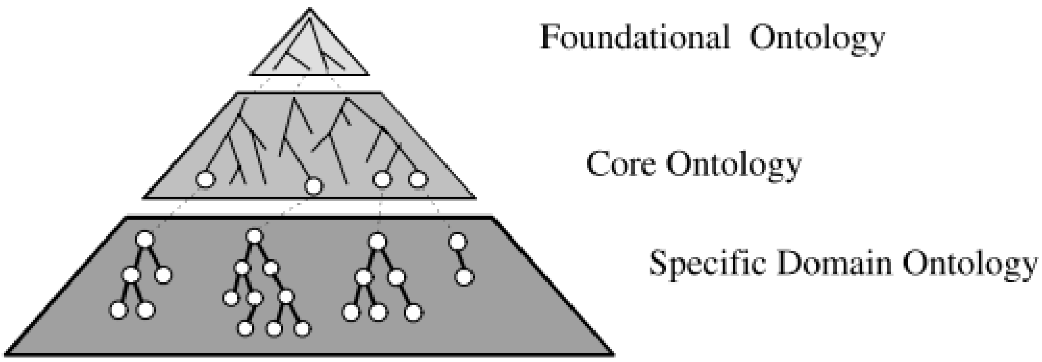

Figure 2.8 - The three levels of generality of a domain ontology [24]

The most needed features to build usable ontologies are:

• Coverage: The domain concepts must be there; the SDO must be sufficiently (for the application purposes) populated. Tools are needed to extensively support the task of identifying the relevant concepts and the relations among them.

• Consensus: Decision making is a difficult activity for one person, and it gets even harder when a group of people must reach consensus on a given issue and, in addition, the group is geographically dispersed. When a group of enterprises decide to cooperate in a given domain, they have first to agree on many basic issues; that is, they must reach a consensus of the business domain. Such a common view must be reflected by the domain ontology.

• Accessibility: The ontology must be easily accessible: tools are needed to easily integrate the ontology within an application that may clearly show its decisive contribution, e.g., improving the ability to share and exchange information through the web.

Figure 2.9 reports the proposed ontology-engineering method [24], that is, the sequence of steps and the intermediate outputs that are produced in building a domain ontology.

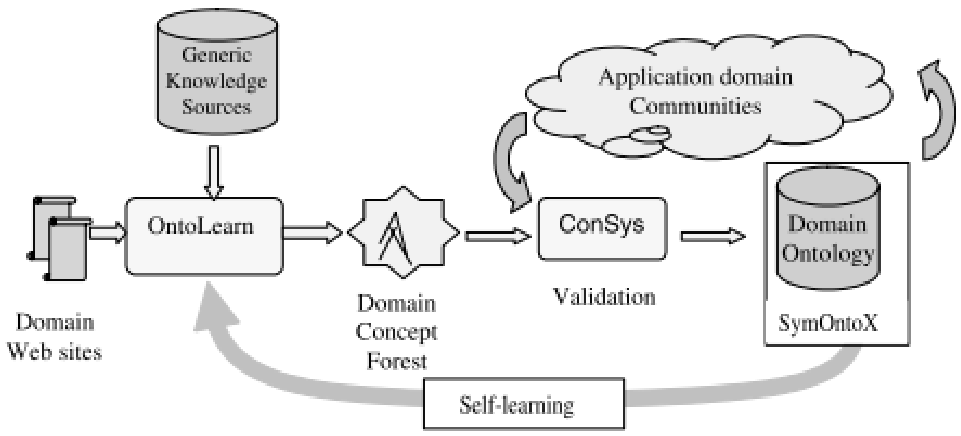

Figure 2.9 - The ontology-engineering chain [24].

Figure 2.10 shows the architecture of the OntoLearn system. There are three main phases: First, a domain terminology is extracted from available texts in the application domain (specialized Web sites and warehouses, or documents exchanged among members of a virtual community), and filtered using natural language processing and statistical techniques. Second, terms are semantically interpreted and ordered according to taxonomic relations, generating a domain concept forest (DCF). Third, the DCF is used to update the existing ontology (WordNet or any available domain ontology).

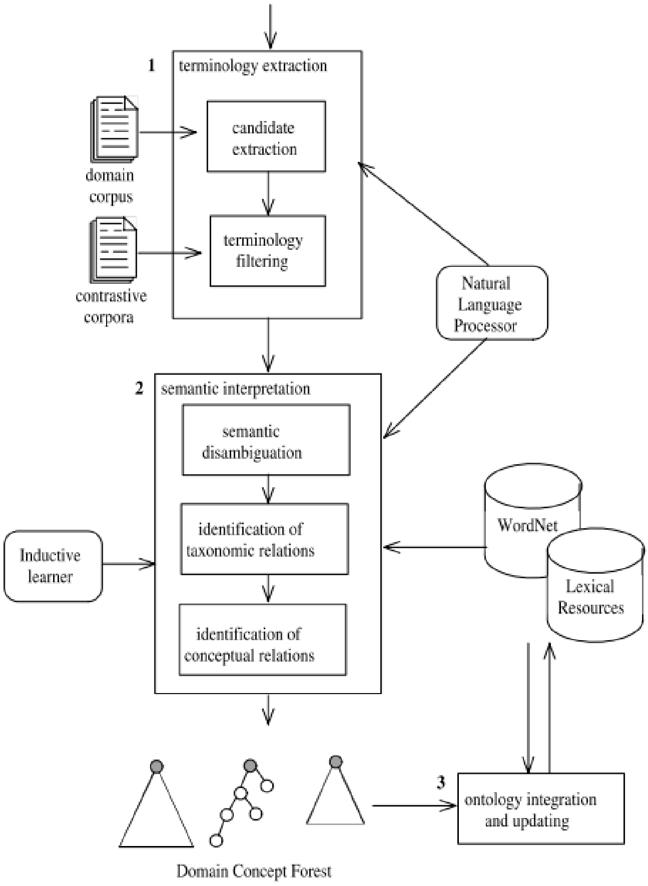

Figure 2.10 - The architecture of OntoLearn [24]

Ontology generation from text is becoming an important area of ontology engineering [25]. As manual generation of ontologies is a very time-intensive and error-prone process, the automatic way is very appealing. Automatic ontology generation requires the use of natural language processing (NLP) technologies and text mining strategies.

The Knowledge Puzzle Approach for Domain Ontology Generation [25] is able to reuse document content for information retrieval and training purposes. The extraction of document content can be obtained through the automatic construction of a concept map for each document identifying the important concepts and relations between them.

In this chapter, the knowledge-based framework proposed by the author is presented along with an extensive description of its purpose and guidelines.

In general, a framework is a real or conceptual structure intended to serve as a support or guide for the building of something that expands the structure into something useful [46].

In this sequence, a knowledge-based framework was developed with the purpose to support the creation of e-learning materials, which would be easily adapted for an effective generation of custom-made e-learning courses or programmes. It embraces solutions for knowledge management, namely extraction from text & formalization and methodologies for collaborative e-learning courses development, where main objective is to enable multiple organizations to actively participate on its production.

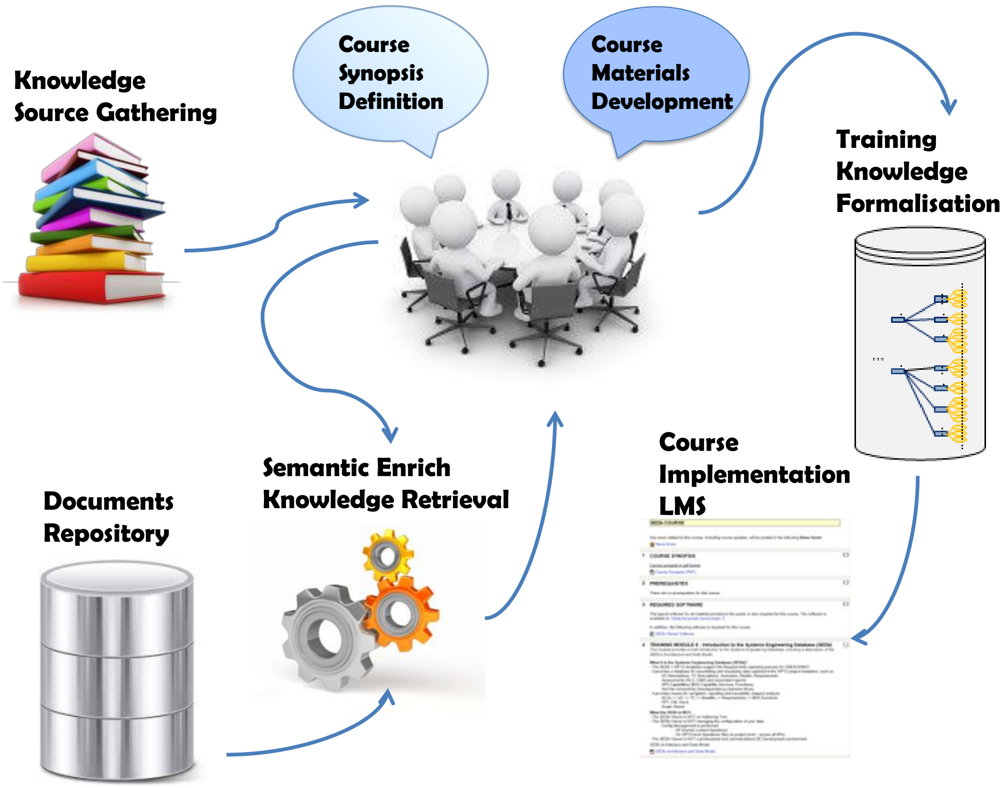

Figure 3.1 - Developed Framework

3.1. Knowledge Source Gathering

With the amount of knowledge sources, it is important to decide the information that must be gathered in order to build the course. So knowledge source gathering is the beginning of the creation process. In this phase the training unit members, which should be composed by domain experts and knowledge engineers, start by gathering source material and information that may be used on the course creation.

The acquisition of knowledge, skills, and competencies as a result of the teaching of skills and knowledge that relate to specific valuable competencies it's called training. The objective of training is improving individual capability, capacity, productivity and performance.

To implement a training programme as a result of research projects it is needed to have a good knowledge acquisition approach able to formalise the developed knowledge in training, so this could effectively to perform an efficient knowledge transfer between the various project stakeholders. And this is aligned with Levine and Gilbert in [47], which stated that knowledge transfer can be accomplished through the support of incentives for generation of new ideas, specific structures to facilitate information discussion and flow within organisations, and also through appropriate communication technology and effective training. They also stated "Knowledge transfer is only valuable when it is integrated into a set of policies for knowledge generation and capture" [47]. This is completely reached by the proposed training implementation approach, because in the various research projects to which it is focused on, they normally have its knowledge transfer process fully addressed and controlled by a set of identified directives, as training policies, supported by organized knowledge acquisition and maintenance.

The knowledge acquisition is the process of knowledge identification and its capture to an explicit format to enable further digital use. Nowadays, the research community has developed automated tools to knowledge identification and capture, and they have identified that how much informal the knowledge is represented on its source more its results decrease on efficiency. Consequently, a full automatic machine learning knowledge acquisition was not yet reached, but we are walking in that direction.

There are several technologies and methodologies, which intend to facilitate knowledge acquisition, to then allow the creation of courses and training materials. However, these approaches are mostly academic-focused solutions where the process of building courses and materials it's more pedagogically exhaustive then what the industry typically requires. This process it's bound to be subject of inefficiencies brought by the using of academic-focused technology into industrial training. It is on the idea of having a more "industry based" approach to build training courses, using results from research projects that the proposed methodology is focused on.

Usually, research projects establish an inter-organizational Training Unit (TU) for e-training development that intends to manage the creation of the training policies and the curricula. Due to its heterogeneous composition members, it is always need some meeting accomplished with discussions to define such training elements (e.g. training courses). Thus, author present in the following two qualitative information collection methods that were analysed, with potential of being part of such discussions in the proposed methodology.

3.1.1.Nominal Groups

A Nominal Groups (NG) session requires several participants to discuss about a topic to generate a list of ideas. This approach is an alternative to a regular brain storming session, because it is a more structured discussing method. In this method are given time to participants to think and write down their ideas before telling them to the group [48]. The process prevents the domination of discussion by a single person, encourages the more passive group members to participate, and results in a set of prioritized solutions or recommendations. The steps to follow in this technique are [49]: 1) silent generation of ideas in writing; 2) round robin recording of ideas; 3) serial discussion for clarification (for each idea); 4) voting on the priority strategies; 5) discussion of preliminary voting; and 6) final voting. The advantages and benefits of the NG method include: 1) Motivates all participants to get involved; 2) Generates many ideas in a short period of time; 3) Obtain ideas from people of different backgrounds and experiences; 4) Stimulates creative thinking and effective dialogue; 5) Allows clarification of ideas.

The relevant disadvantages can be related to: 1) the necessity of having a skilled leader; 2) it can take too much time if the group is not properly controlled and is allowed to run for too long; and 3) assertive personalities may dominate unless leadership skills are exercised.

3.1.2.Metaplan

The Metaplan technique is a learning method especially for groups. It is a collaborative and moderated technique with focus on solving group decisions or problems [50]. It combines individual and collective contributions and is used to organize concrete ideas into more general conclusions leading to recommendations. It is mostly useful to explore an issue and dig out what is key in it.

The Metaplan process can be described in five steps: 1) Agreement on principles; 2) Exploratory discussions; 3) Development of the “dramaturgy”; 4) Meeting / Workshop; 5) Post meeting follow-up.

The advantages of this method are related to the difficulties to: 1) reach common points of view and take actions to support participants’ convictions; 2) lead the group into joint actions; and 3) discuss and clarify the objectives within the allotted time.

The disadvantages are related to: 1) the necessity of having a skilled leader; 2) the possible existence of some overlap of ideas due to unclear wording or inadequate group discussion; and 3) "Knowledgeable" individuals selected to participate may not represent all community subgroups.

From the presented methods, NG was the chosen to support in training definition, because it actively promotes the participation of all the members and generates a good amount of ideas in a short period. However, some of the disadvantages mentioned could be reduced by its introduction into a LMS, where its timers or schedule planning, could efficient replace the need of a leader to manage the sessions.

3.2. Course Definition

Following the knowledge gathering the training unit have the prime material for the course development. A learning course is an ordered procedure or sequence of a number of lectures about a subject. It is conceived as a method that meets the specific requirements and expectations of a determined target audience. A learning course is separated into several modules, according to the topics that are addressed. These modules are called learning modules, and are a small piece of a learning course, basically a lecture, with a very clear purpose. Several modules of similar topic area can be grouped together to form a learning course as illustrated in Figure 3.2 (a).

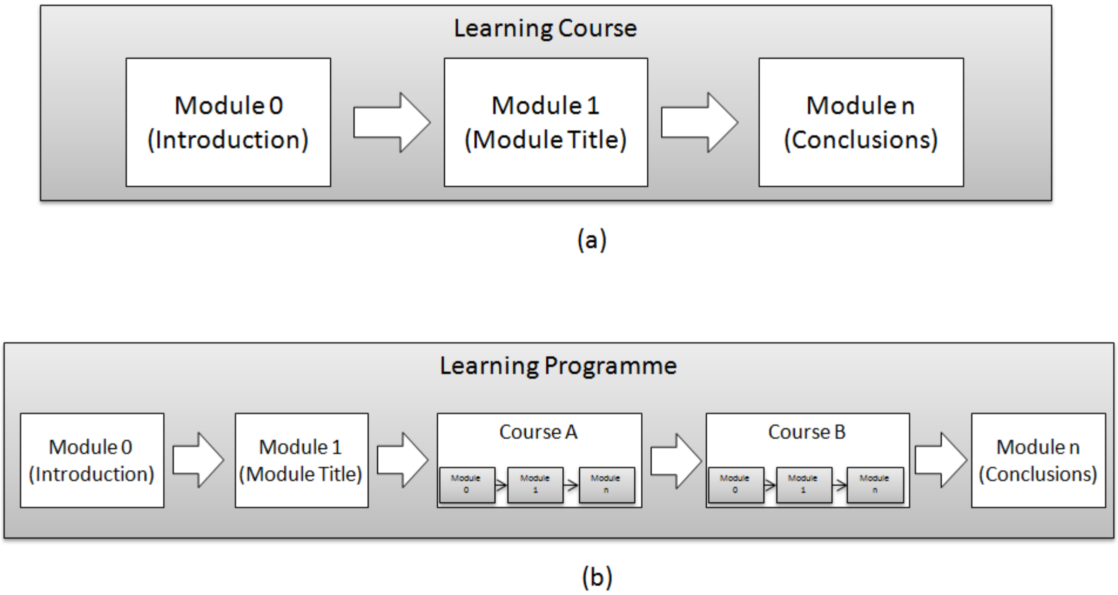

Figure 3.2 – (a) Learning course and (b) Learning programme

A Learning Programme is a significant long-term learning activity which comprises a set of learning courses and/or learning modules Figure 3.2 (b) [51]. It's a construct conceived for learning in specific skills focusing on a given target audience and using a selected delivery approach. Reference learning programmes are those that are designed for reference target audiences (especially relevant within a given learning environment) and that serve as orientation for targeted learning execution [51].

To proceed with the training course development, that intends to be aligned with the general training objectives defined by a training policy, a course synopsis definition must be made. To start this course synopsis definition process it is needed to know what are the training target audience profiles intended to reach.

Afterwards the course materials must be developed in a way that can be delivered to the subject. Prior to this stage, the synopsis must be exported the defined to the ontology.

3.3. Semantic Enrich Knowledge Retrieval

The main purpose of semantic enrich knowledge retrieval in this framework is to complement the training course with bibliographical elements, such as deliverables and other scientific documents, that have a certain value of similarity with the desired keywords. This can be seen in Figure 3.3.

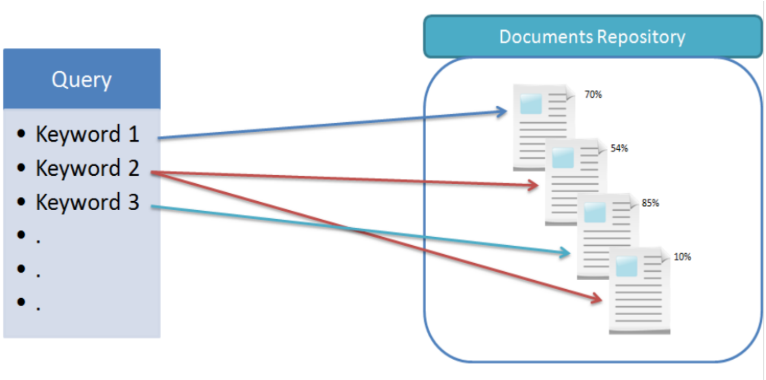

Figure 3.3 - Input query and output documents

To achieve this, first terms must be extracted from the documents. These terms are decided by a query with keywords as an input. The extraction process starts by splitting each document into sentences. Then, apply tokenization where terms in each sentence are extracted as tokens. All tokens must be transformed to lower case font. Then terms belonging to a predefined stop word list are removed, to avoid word such as *the*, *is*, *at*, *which*, and *on* etc, so that only the meaningful words can be processed. Using the snowball method, the remaining terms are converted to their base forms. Related terms are then combined for frequency counting.

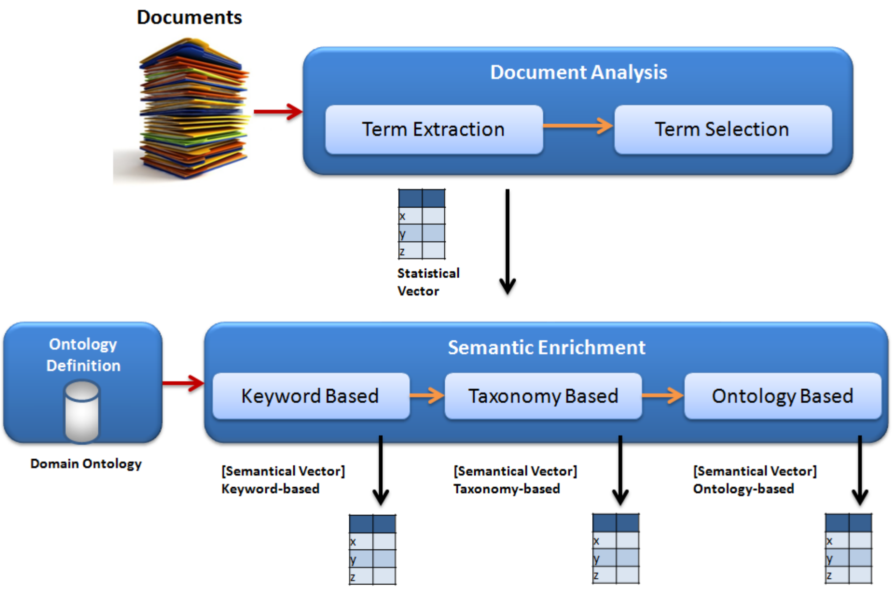

Figure 3.4 - Semantic Vector creation process

After the term selection process, the semantic enrichment phase starts. Starting with a statistic semantic vector that is a vector that only takes in count the frequency of each term, more rich vectors are created as seen in Figure 3.4.

The keyword-based semantic vector takes into consideration only the relationship between terms from the statistical vector and the concepts in the domain ontology. Then matching the statistical vector keywords with equivalent terms linked to each ontological concept, by finding similarities between each keyword and the equivalent terms within the ontology.

The taxonomy-based semantic vector is the next stage in the semantic evolution of KRs. This vector is made by adjusting he weights of concepts according to the taxonomic relation among them. If two or more concepts taxonomically related appear in a keyword-based semantic vector, then the existing relation can improve the relevance of the expressions within the KR and therefore enhance weightings. The taxonomy-based semantic vector is created based on the relations between concepts within the ontological tree.

Ontology-based semantic vector is the last level in the semantic evolution of KRs, which is based on the ontological relations. The purpose is to study the co-occurrences of concepts in unstructured sources of information in order to provide remarkable relationships for improving ontological structures. The creation of the ontology-based semantic vector is a two-stage process using the taxonomy-based semantic vector as input. The first stage increase weights of concepts previously presented in the taxonomy-based vector, depending on the ontological relations among them; the second stage append new concepts that are not currently in the input vector, according to ontological relations they might have with concepts belonging to the taxonomy-based vector. This process is an approach of the work done by [17]on semantic enrichment.

3.4. Training Knowledge Formalisation

Formalisation is the act of modelling knowledge into an ontology that further on would facilitate simple specific querying and advanced reasoning.

Ontologies allow key concepts and terms relevant to a given domain to be identified and defined in a structure able to express the knowledge of a domain or a segment of the reality/world [52]. Its standard capacity to characterize knowledge, to assist reasoning, use and exchange knowledge among systems or users contributes to increase the computational intelligence of its system. Thus, ontologies can be used to support knowledge management and to provide some intelligence to eLearning systems.

Since an ontology can be used to represent a learning knowledge base, it facilitates the classification of its elements and subsequently reasoning over it. To reach this point it is required to understand how to organise learning related knowledge and convert it in appropriate and appellative learning objects. Adding up, such related knowledge should be planned to assure learning objects management. Thus, the build of an ontology to represent learning is an suitable objective.

Ontologies for e-learning are different in the following aspects: content what the learning material is about, context in which form a topic is presented, and structure as learning material does not appear in isolation[53].

With the course synopsis created, in order to transfer the course to the LMS, the solution thought was transfer the synopsis to an ontology where it is formalised and become more versatile to be exported to a LMS. To do this the author used a technology called OWL (Web Ontology Language) that is one of the languages for knowledge representation. OWL uses formal semantics and RDF/XML-based serializations, like many other languages of the Semantic Web [54].

To achieve the implementation of the course synopsis on the LMS, the ontology developed by[55] seen in Figure 3.5 was used as base, and altered to match the problem requirements.

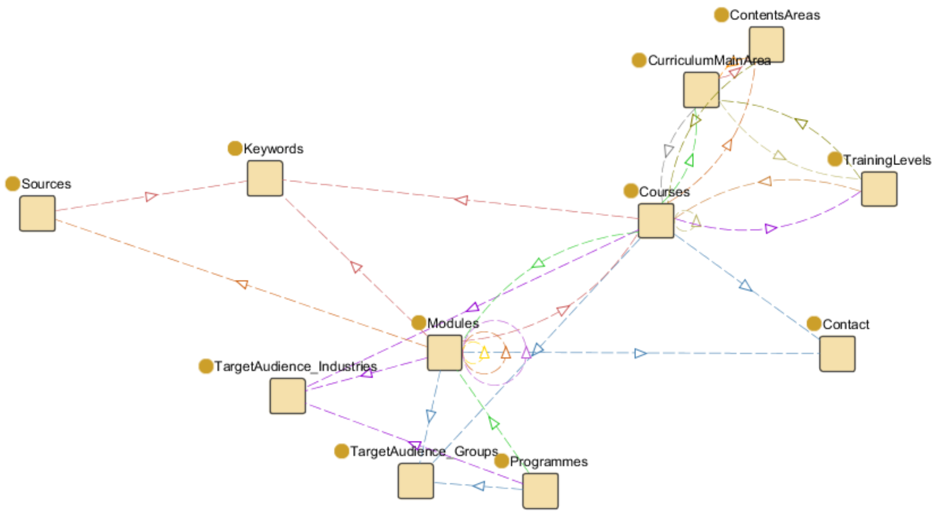

Figure 3.5 - Relationship structure of the Ontology [55]

3.5. Course Implementation in a LMS

As a Learning Management System (LMS) is a software application for the administration, documentation, tracking, and reporting of training programs, classroom and online events, e-Learning programs, and training content. Some LMSs are web-based to facilitate access to learning content and administration[56].

Since time and efficiency are becoming more important, it is needed a training delivery that is both independent of location and time, such as web-based courseware that the learner can access universally, at anytime and anywhere, i.e. e-training. Learning Management Systems (LMS) are trying to answer the challenge, integrate more and more functionalities in management and improvement of training. Although, such grow in functionalities presented to build e-training materials involve an augment of the complexity to produce them.

Among many others, MOODLE is well-known open source web-based learning management system. For being free MOODLE it's the most popular among the academic community, with many universities using it as its LMS and with a strong community supporting, which leads to its rapid evolution. Moodle stands for Modular Object Oriented Developmental Learning Environment, and is more than a Content Management System, since it provides support for learning activities. Moodle is able to handle the standard features of a course and its associated settings, such as users, roles, resources, common site navigation, etc. It allows the creation of courses from scratch, but it also allows other information packages to be imported and used (e.g. SCORM, Content)[55].

The development of training materials is associated to all the actions needed to produce materials to training implementation. A training course structure with assessments included is a great approach for constructing an interactive course. Thus each training course module should have at least one assessment question so the trainees can test their self-progress in learning.

In addition, each course should also have other type of questionnaires to obtain feedback, first from experts and then from trainees to accomplish the course validation and evaluation, and thus to improve the quality of the materials.

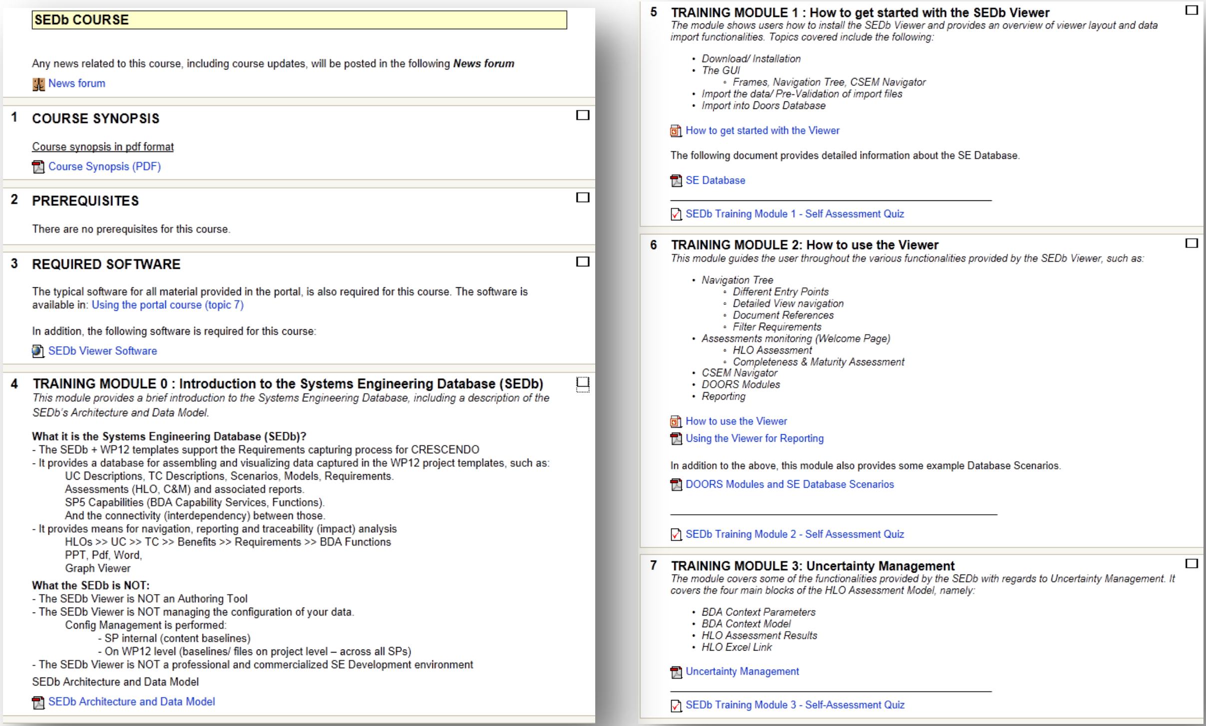

Figure 3.6 - LMS course implementation example [58]

4. TRAINING DEVELOPMENT METHODOLOGY

The proposed training development methodology translates design specifications into training materials. It starts by identifying the training objectives and the target audience including desired roles & competences. Then it uses an appropriate instructional approach to perform the training courses' materials development, complemented with a set of different quality reviews.

The overall process of developing training follows a specific process, composed by three different task tracks (training development, overall training validation and training execution) that complement each other (Figure 4.1).

The training development track starts by defining the course's synopsis according to the directives obtained from a training overall objective. Then the TU performs a quality review of the synopsis. This is done in the training validation track. Thereafter, the course material is developed, followed by two more quality review cycles, similar to the first one. The first is made internally in the TU. Experts who are familiar with the training contents and able to analyse and validate the courses perform the second cycle. The training authors improve the courses based on the feedback from these two validation activities. After this, in the next phase, another quality review is performed through a pilot course execution. Here trainees are the ones who give comments about the courses and the training authors once again improve the courses for the final release. All this process is accomplished by the support of a LMS, following a predetermined steps status shown on the left side of Figure 4.1 and detailed in Table 2.

The courses after reaching the final stage of this methodology should be able to train a specific audience, about some specific topics, delivering trainees the appropriate skills to help them gain specific competences. Thus, it was identified the need of defining all these objectives in a courses synopsis.

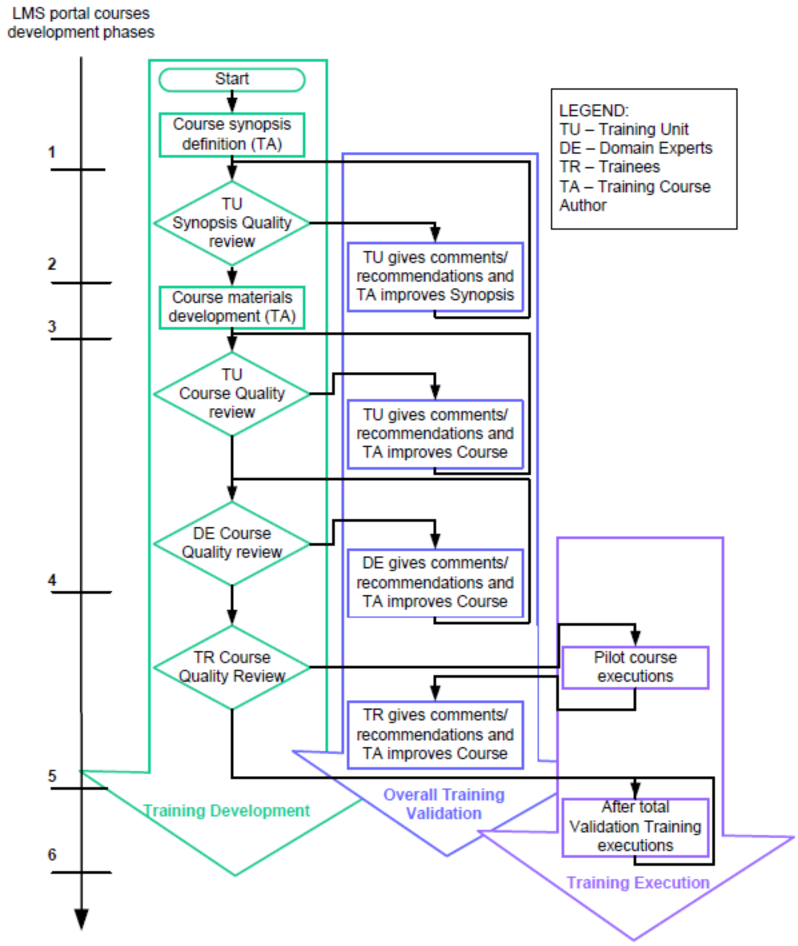

Figure 4.1 - Detailed LMS Training Implementation Process

A course synopsis is an official description of the course as stated in an institution's catalogue of courses. It should indicate the overall goal of the course, briefly characterize the main topics covered, point out why the course is important to students, identifying any special instructional methods to be used, and comment on what background students should have in order to best appreciate the course content [57]. The courses synopsis also acts as course development guidelines to the training course's authors.

4.1. Course Synopsis Definition

The training development intends to be aligned with the general training objectives defined by a training policy. To start a course synopsis definition process it is needed to know what the training target audience is (profiles) and industry roles & competences intended to reach. One example of it was developed in CRESCENDO project [58]. It should contribute to the skills and competences development of the trainees as required for specific understanding and exploitation. This reflects the need to develop, organise and run courses, for example, to train "future users", in how to use, implement and support for instance, a specific software.

Step	Activity	Artefacts accomplished on the LMS Portal
1	Training course/topic identified	A line in the list of courses
2	Course owner accepted training course proposal	Synopsis + Early row material uploaded
3	Defined contents and agreed course structure	Early material uploaded and organised according the agreed structure
4	Training course in development	1st full version uploaded
5	Training execution	Feedback forms
6	Improve course based on feedback	Updated final version uploaded

Table 2 - Steps of the course development

The author proposed a 5-step procedure to assist the definition of the courses synopses, which was defined based on experience resulted from organising training implementations in several European research projects. The following course synopsis definition is described with additional explanation of its implementation using a WIKI. Due to its required ability to manage knowledge the implemented system is supported by adapted qualitative information collection methods.

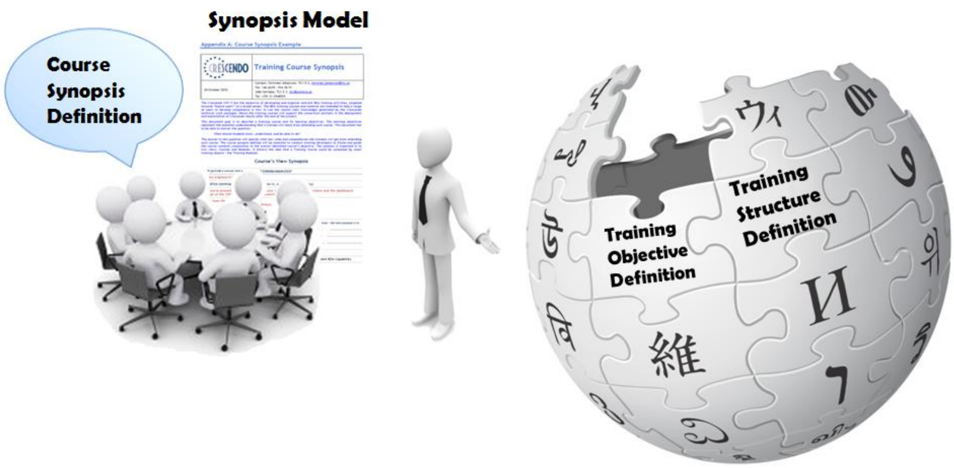

Figure 4.2 - Proposed Course synopsis definition solution

The Figure 4.2 presents the proposed solution made by the author to develop the course synopsis definition and hereafter create the synopsis model on a wiki based platform.

Step 1 – Training Topics Definition

This step helps people identify topics that should be addressed in the training courses. The topics are defined by answering the following questions:

• **Question 1.1** - What contents/topics (discussed under this research Integrated Team/group) should/could be used for training purposes?

• **Question 1.2** - What do you envisage will be the new knowledge (or know-how) from application / deployment of results from this research group work?

• **Question 1.3** – In your own words, which will be the new 'updated' competencies from this work?

In a WIKI based implementation the question 1.1 topics are defined following 5 steps as described in Figure 4.3. Such steps were defined based in the NG sessions. The questions 1.2 and 1.3 follow a similar structure as shown in Figure 4.4. The main difference is that there are various reformulations until reach a full agreement in what could be the new knowledge and competences identified from the conducted research work. This is made in a special configured WIKI; the questions have textboxes like an essay to ensure that answers can be used on further steps. Thus, an advanced management of the WIKI databases are required to accomplish knowledge transfer from step to step.

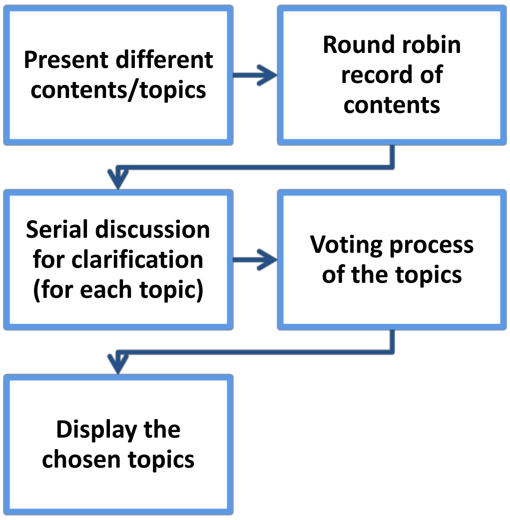

Figure 4.3 - Topic choice process based on NG

Step 2 – Target Audience and Training Course Types Definition

This step is related to the identification of the target audience and competences, for each of the training topics identified in step 1. Thus, the questions to answer in this step are:

• **Question 2.1** - What do you envisage is the target audience for each of the training topics identified in step 1?

• **Question 2.2** – For the target audience identified in 2.1, what do you envisage to be the associated competences and/or roles that are enabled by those training topics?

Since the target audience was chosen before the process start, the model to follow in question 2.1 will be the one seen in Figure 4.3. In the other hand question 2.2, must have some discussion to reach a common view about the competences defined, thus the model adopted is the one in Figure 4.4, because it is prepared to handle cycles to accomplish discussions until reach their agreement.

Step 3 – Training Objectives Definition

Step 3 is related to the training objectives definition for each of the training topic identified. The training objectives are represented by the soft and hard skills that a trainee should learn from the training topic that will help him/her to reach the related competence.

• **Question 3.1** - For a training topic without associated competences:

o What should trainees know, understand, and be able to do on this training topic?

• **Question 3.2** - For a training topic with associated competence(s):

o What should trainees know, understand, and be able to do on this topic in order to reach the identified competence?

In this step it is promoted open discussions in order to achieve the goal proposed. Consequently, Figure 4.4 describes the process followed in the WIKI implementation to perform these step activities.

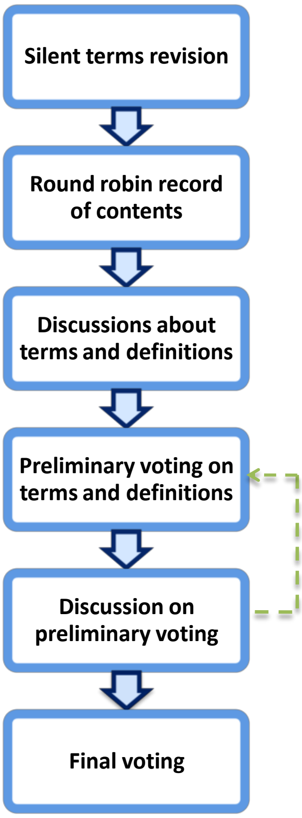

Figure 4.4 - NG approach

Step 4 – Training Themes and Structure Definition

A training course is an ordered process of a number of lectures dealing with a subject. It is conceived in a way that meets the specific desires and expectations of a determined target audience. A training course is divided into several modules, according to the topics that are addressed. A training module is a small piece of a training course, essentially a lecture, with a very clear objective. Several modules of the same topic area can be grouped together to form a training course as illustrated in Figure 4.5.

Three sub-steps compose step 4:

• **Sub-step 4.1** - Organise a list of groups of **{topic ideas; statements; competences; and training objectives}** related to each other (identified in the previous steps). Each of these groups is at the first instance, a rough, and potential, training module.

• **Sub-step 4.2** - Try to **identify a more generic theme** that could **categorise and consequently aggregate one or more of the rough training modules identified**. Note that each theme identified could represent a potential course. Themes that aggregate a higher number of rough training modules would be the most appropriate to become a course.

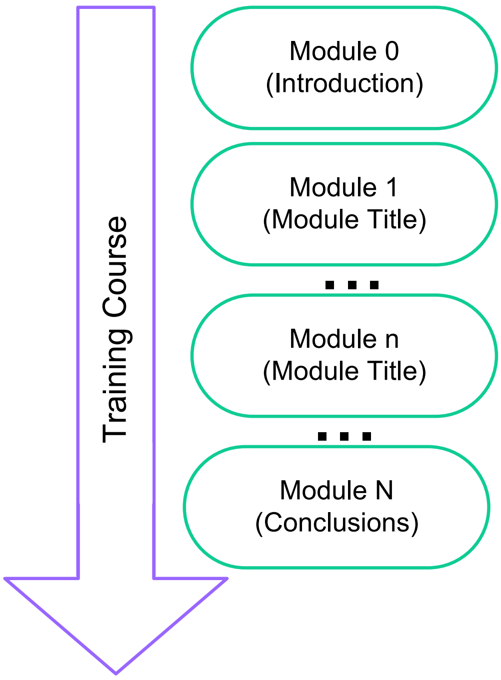

Figure 4.5 - Training Course Structure

• **Sub-step 4.3** - Try to organise in a specific order each of the rough training module that were categorised around a specific theme. This establishes a course with an organised set of training modules around a theme. Thus, each rough training module could become a final training module of the training course structure. A course structure should typically be initiated by a *"Module 0"* that is the introduction to the topic. *"Module 1"* and the following ones (except for the last) are related to the content itself. The last module should be the conclusions module (see Figure 4.5), which summarises the main points and concluding remarks of the course.

All these sub-steps are implemented in a WIKI following the NG approach steps as presented in Figure 4.4. These sub-steps are accomplished with various discussions that requires a full agreement, thus, they would need various voting cycles to accomplish a decision.

Step 5 –Fill out the Courses Synopsis Form

This step defines the course using the previous identified resources/ideas by filling out a "course synopsis" form composed by the following metadata:

• The **course title** provides a concise title of the proposed training course.

• The **narrative summary** is a summary of the training course, with its objective and highlights.

• The **Keywords** represent the topics addressed by the courses.

• **Roles and Competences** are those Competences to which a course is related with.

• **Recommended Precedence** is for example, a course that the trainee should follow before attending this one.

• **Student requirements** can be previous students' knowledge, which is required for a good understanding of this course.

• **Technical requirements** could be specific software needed to the trainee follow in a specific training course.

• **Estimated Time** is the duration of a training course in minutes.

• **Input Content** is the input content sources used as relevant and essential for the development of the training contents.

• **Training Modules** represents a section of a course (Figure 4.5).

• **Skills** can be of three different types: **-Know** (Skills to be acquired related to the knowing and understanding (Theoretical knowledge of a field; the capacity to know and understand)); **-Do** (Skills to be acquired related to the knowing how to act (Practical and operational application of knowledge to certain situations; be able to accomplish)); **- Be** (Skills to be acquired related to the knowing how to be (Values such as an integral element of the way of perceiving and living with others and in a social context)).

• The **Contact Person** is the contact information of the main responsible/author of this training course.

4.2. Course Materials Development

Before to start the course materials development it is exported the training metadata (synopsis) defined to the ontology. Then a set of authoring sessions are available with the objective of providing some directives and explanations to the authors in how they should prepare their training materials ensuring the same structure and approach.

As an example, for a training author it is easier to start by developing small pieces of training than an entire course. Thus, in first the author starts by developing first modules and only then the course itself. With this, the reference training courses contents can be also divided into manageable modules to facilitate the organization in producing courses but also characterize them through the prepared ontology. Then, authors can develop the course modules contents using materials developed inside the project and, when needed, complemented with external materials.

In additional, through an ontology-based system, any training module can be searched and retrieved, to be used separately in an atomic way by other authors or trainers (e.g. academics in their classes as additional educational materials).

4.2.1. Overall Training Validation

The training validation and evaluation ensures that training-under-development stays on track, safeguarding achievement of training goals and analysing system performance. A quality review process based on decisions and revisions for future course iterations can be made after evaluating the strengths and weaknesses in a completed training programme, thus ensuring achievement of desired goals.

Training validation ensures that courses follow their own synopsis and are in accordance to general quality aspects defined for the purpose. These validations are accomplished in LMS through specific feedback forms. Such forms can be developed for trainees, TU and domain experts. Each of the groups provides different feedbacks concerning their expertise on the course domain.

4.2.2. Training Execution

The training execution is the act of delivering training about a specific theme/domain with a comprehensive purpose to enrich the knowledge and/or to provide precise competences to a pre-determined target audience.

The training execution phase focuses on setting up the training environment and delivery or distributes the instructional materials, ensuring the delivery of a training session capable to capture the trainees' interest. There are three different training types:

Bound by time and place - It is the traditional training delivery method that requires the physical presence of students at the location at which the training takes place;

Bound by time only - e.g. instructor-led synchronous virtual classroom sessions (**webinars**), where a tutor guides the learning activities over the Internet;

Both independent of time and location – it is a **webbased** courseware that the learner can access ubiquitously, i.e. anytime and anywhere.

Due to the existence of different tools or interfaces available in a LMS, it is able to provide training in the various types identified.

4.2.3. Training Implementation Analysis

The Training Implementation is in a broad sense related to all the activities needed to perform training delivery/execution. Thus, its analysis is related to the work conducted since the training courses definition, until the delivery or distribution of such instructional materials ensuring training sessions able to captures trainees' interest.

Training implementation analysis is a special research area because its intention is to use the various outputs of all the training system to conclude results to use as further improvements of the training approach. To be implemented in a LMS, it requires additional advanced tools for analytical studies. This is a future research area to be addressed by the author.

As demonstration it is shown a training program that followed the proposed training development methodology. It was implemented in the European research project CRESCENDO. CRESCENDO it was focused in Collaborative and Robust Engineering using Simulation Capability Enabling Next Design Optimisation in the aeronautics domain. Such training implementation resulted in the development of ten courses: 1) BDA Introductory Course; 2) Virtual Testing Training; 3) Value Generation Training; 4) Behaviour Architecture Course; 5) Simulation Integration Course; 6) Simulation Quality Course; 7) Enterprise Collaboration; 8) Architecture Executive Guide; 9) Architecture Implementers Guide; 10) Systems Engineering Database (SEDb) course. Figure 4.6 presents one of its courses developed, using MOODLE as LMS system.

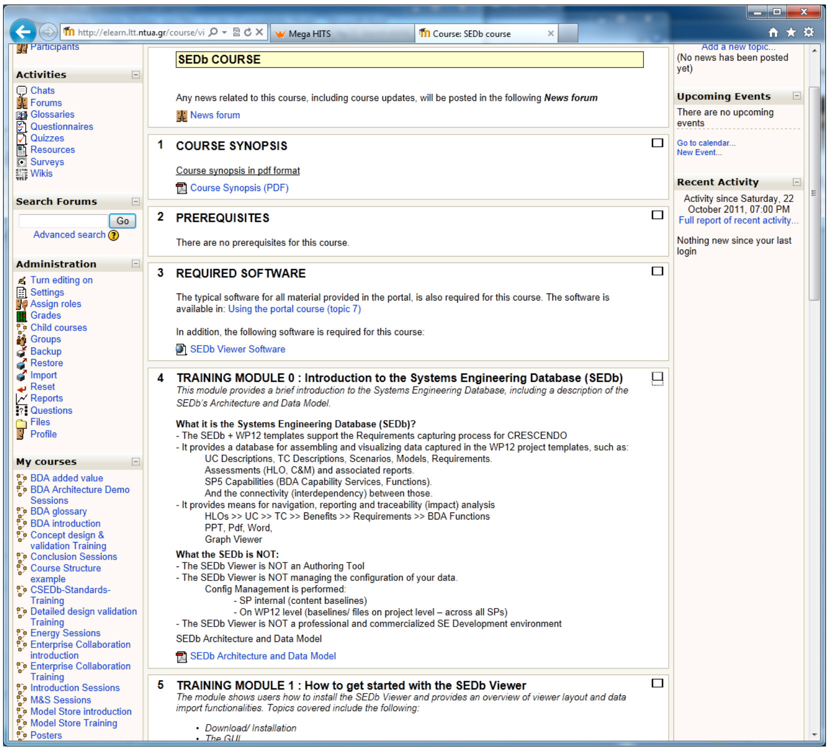

Figure 4.6 - LMS Implementation

In the training to authors course of this training programme it was stated, from its various messages, as an example, that each training course structure should include at least one self assessment question at the end of each module, to make the courses a little more interactive. Following the presented methodology, all the courses were reviewed by the TU members, who provided comments, such as the following one: "…the 'sota template' training module has to have a small sentence to introduce the module to clearly introduce its contents …"

After TU made their approval/validation of the courses, these were conducted to the next step, the CRESCENDO/Domain Experts validation. On this step some discussions took place. The main contributions were related to the correctness, reliability and clearness of the contents. This conducted to the formalisation of some guidelines as a kind of checklist to improve the quality of the materials. Afterwards a first pilot training execution was performed, where people that attended the courses provided their feedback through the mentioned feedback questionnaire.

Following the course synopsis definition, a process was created in order to achieve the goal of creating an E-training course. The following Figure 5.1 presents the author solution to the framework presented on chapter 3. From this architecture solution a prototype platform was developed using a WIKI extension as a support to particular customizations.

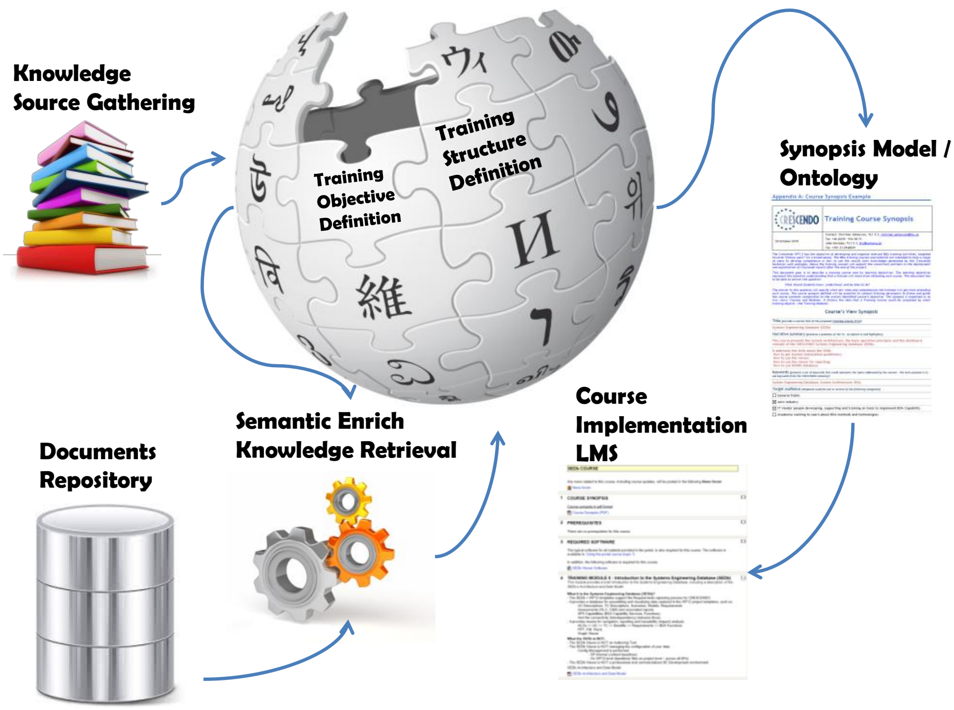

Figure 5.1 – Developed Architecture

The wiki was used due to the fact that is a versatile platform, it is very intuitive to use and can be accessed anywhere the only requirement is internet connection. An approach using moodle was tried but with no success since altering moodle pages automatically was a very difficult process. Due to the way that all the page information is uploaded in one text line and so altering anything was not practical.

5.1. Implementation

In this section is presented an application developed in JAVA and using Wiki as a benchmark to user interaction and discussion to support the course creation based on the presented methodology.

Java is a versatile technology that allows developers to create software that is autonomous of the platform where it will run. It also allows multi-threading programs that run in Web browsers and Web Services, and combine Java objects between applications. Other than the fact that Java is a very remarkable technology, the adoption of Java on this project was also predisposed by the lack of support for other languages by Protégé (that only provides an API for Java).

The application has a moderator interface window, which allows a specialized individual within the specific domain, control the course creation process.

The following diagrams show the course creation procedure as a voting and discussion method. Each of the steps represents a step of the methodology presented on the previous chapter. The process starts on step 0 that is not displayed on the diagram since it is the target audience introduction, and this is made by the moderator by inserting it in the wiki. This step does not have any discussion involved or any sort of voting, is predefined to the process.

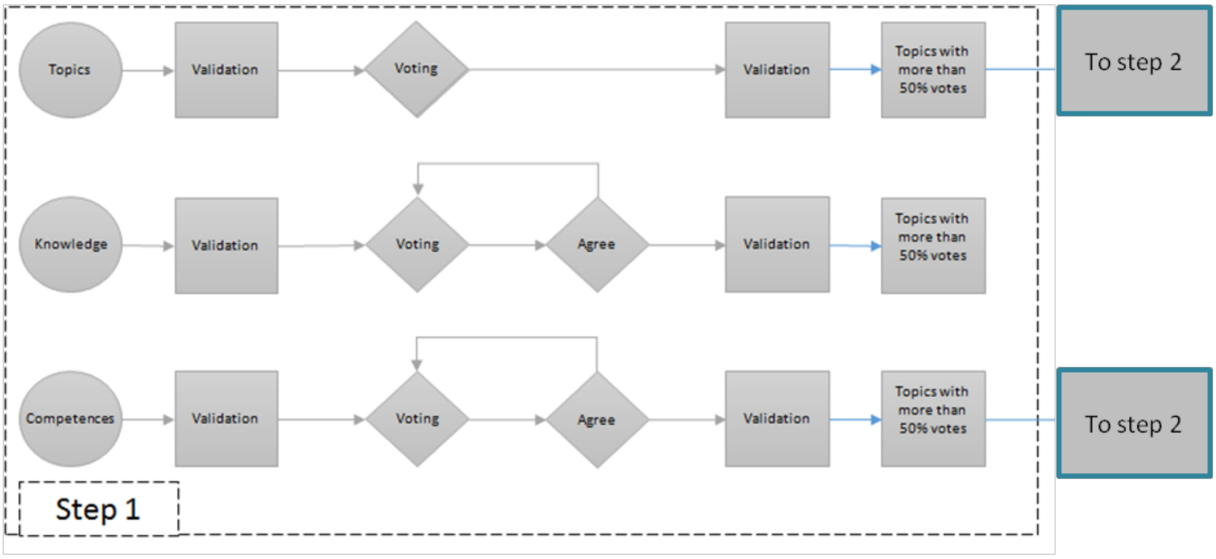

Figure 5.2 - Training Course Types Definition process diagram Step 1

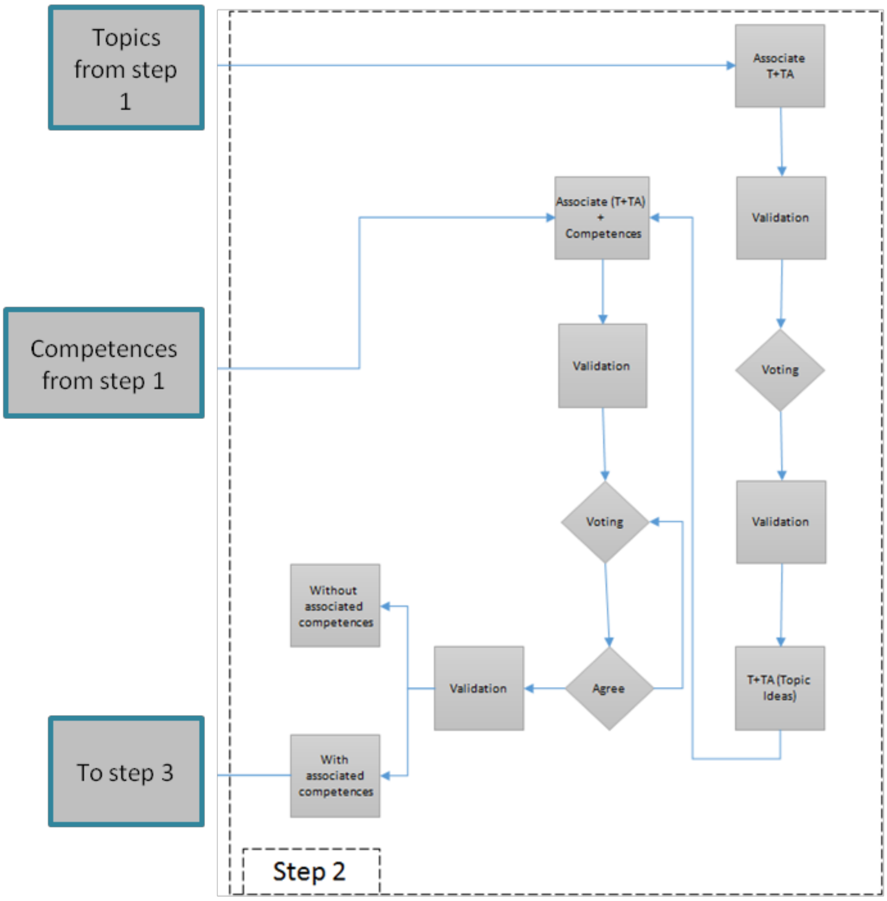

Figure 5.3 - Training Course Types Definition process diagram Step 2

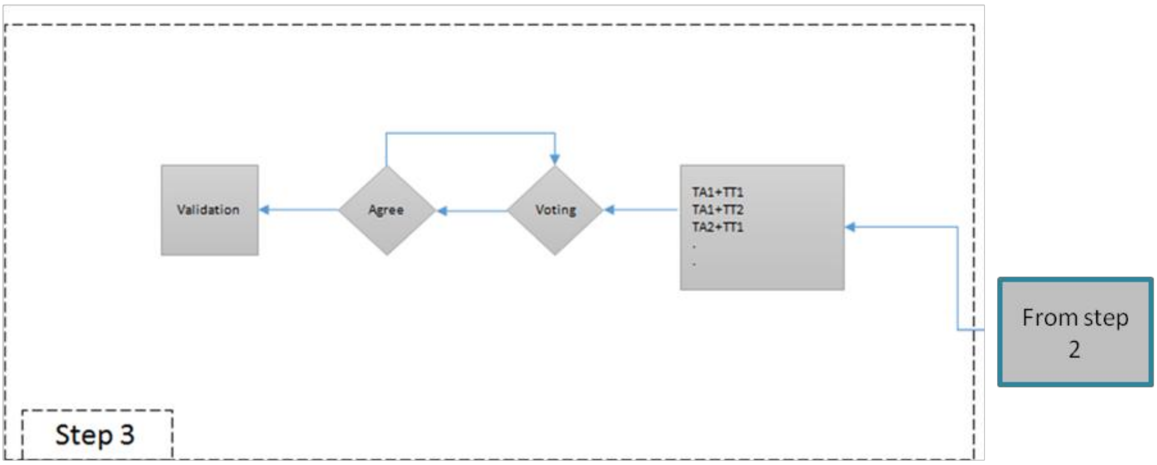

Figure 5.4 - Training Course Types Definition process diagram Step 3

The info is then gathered and used as input to Step 4.

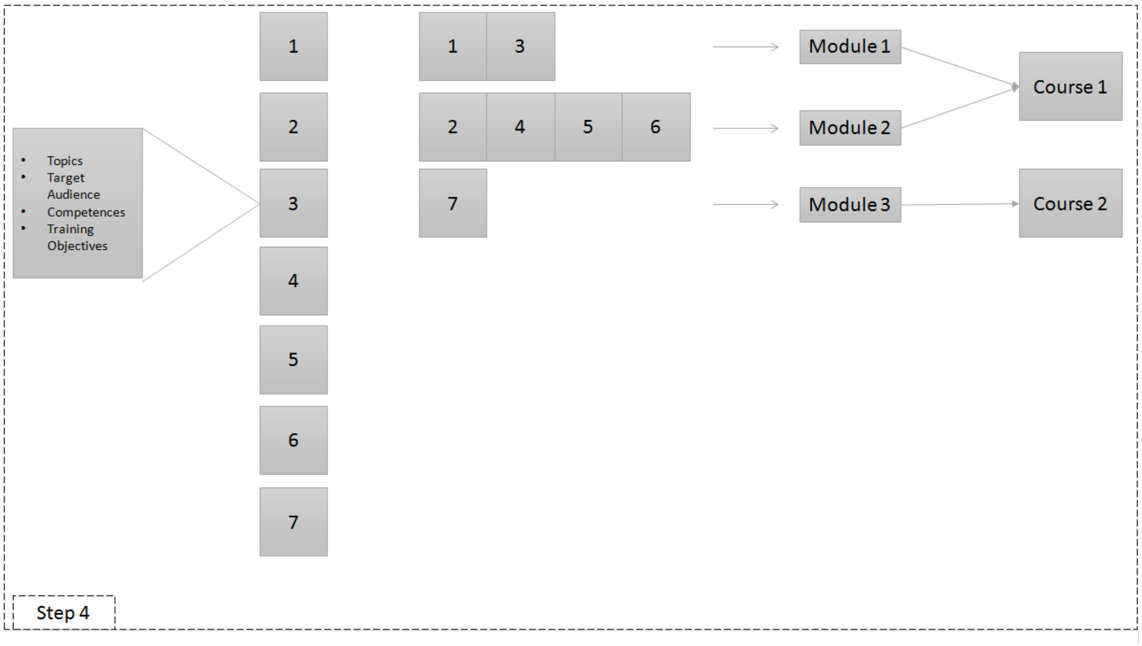

Figure 5.5 - Training Themes and Structure Definition

Then, the author explains the implementation of each step of the course creation described on the methodology.

From step 1 to step 3 the procedure is very similar, some information is inserted then the training unit discuss it, and a first voting process is made. In the cases that exist a second voting process, if all the members agree on those results, the creation process continues, if not the voting process must be done again, so new information can be added completing the step with more information. In step 1 it is visible three different selection branches, each have a different input (topics, knowledge and competences). What differs between each these steps is the type of information being introduced and the amount of information already gathered in the structure. In step 2 the inputs are the topics chosen and the target audience. The output this step is the association of the target audience with each topic chosen. Lately on the same step, the competences selected are going to be associated to the topics and target audience group (author called this groups Topic Ideas). Lastly the third step, the input are this topic ideas and their associated competences, and the objective is to define what are the training objectives to each group of topic ideas and associated competences.

Step 4 is the stage where the course is being assembled since all the information is already gathered. Each of those blocs numbered from 1 to 7 are information blocks, they may be possible training modules. Those are then grouped in modules each of these modules are now a potential training course. Afterwards, these modules are then assembled in courses.

The wiki is composed by five category pages that represent each step of the Course Synopsis Definition. Each step category page contains other different pages that display information for the contributors, with the objective that after following the process a course can be created.

The process is regulated by a moderator that must be a person with knowledge of the matter in order to control the entire procedure and assist the contributors in any way possible. The moderator has a JAVA interface window to control each step of the process. Figure 5.6 show the moderator interface window and all the options the moderator have to control the course creation process.

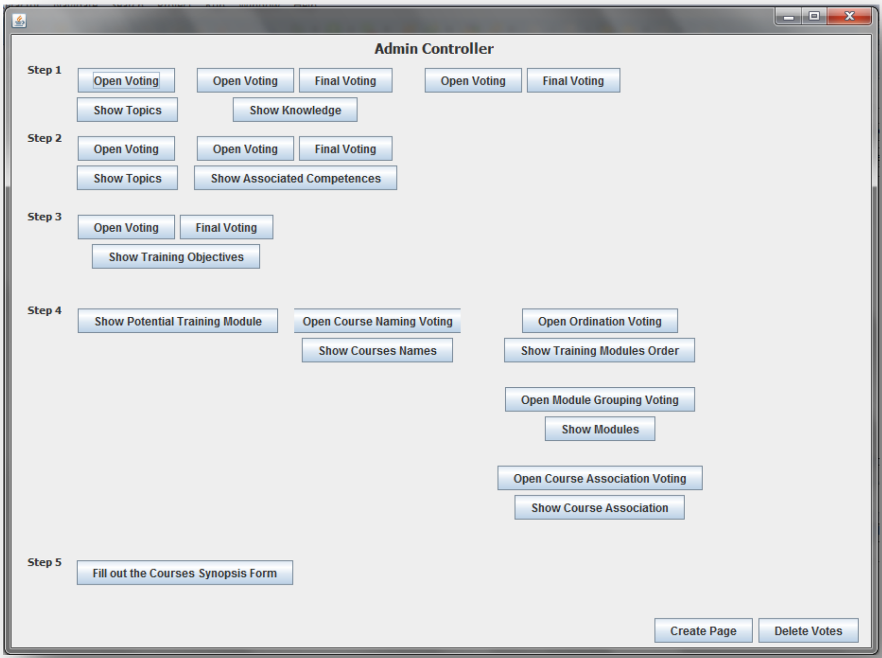

Figure 5.6 - Moderator control window

On the main page (Figure 5.7) there are the links to each step page. An order must be followed, or else an error message appears saying that the previous step must be completed first. Since each step page represents a step of the course synopsis definition of methodology, it has all the information necessary for the contributors fulfil the page objective in order to continue the process.

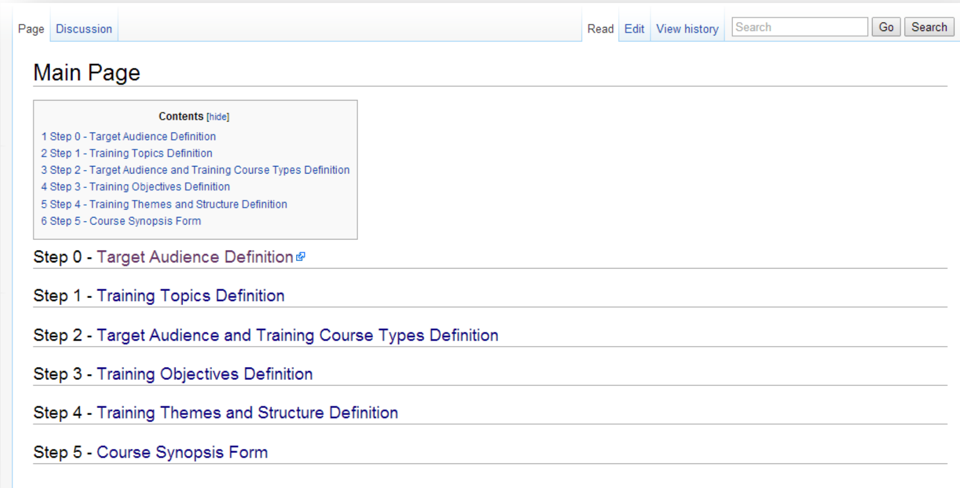

Figure 5.7 - Wiki main page

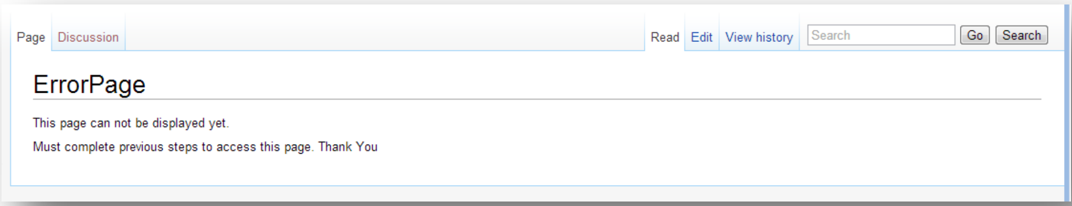

Figure 5.8 - Error page

5.1.1. STEP 0

The objective of this step as said before is to define the target audience. This is made only by the moderator, without any type of discussion or voting process. Using the textbox on the right side (add a subject to this category) the moderator inserts the Target Audience. A page will be automatically created which will be associated to the Target Audience category page, as can be seen in Figure 5.9 (Computer engineering and Electronic engineering).

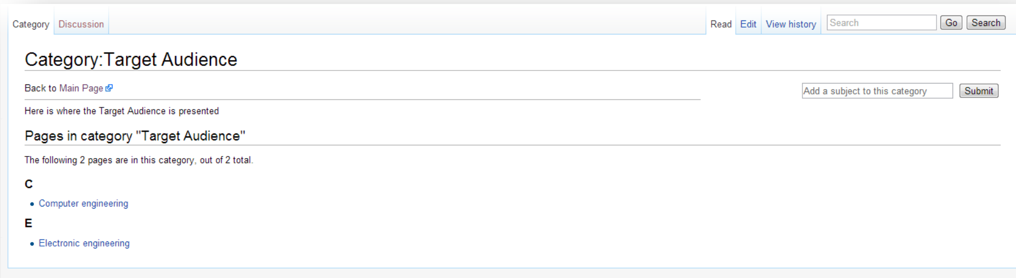

Figure 5.9 - Target Audience page

Inside each category page, the moderator will fill the information related to each target audience with as much as detail possible so the training unit can explore this information later in the selection phase.

5.1.2. STEP 1

Inside the “Training Topics Definition” the three questions related to step 1 are displayed.

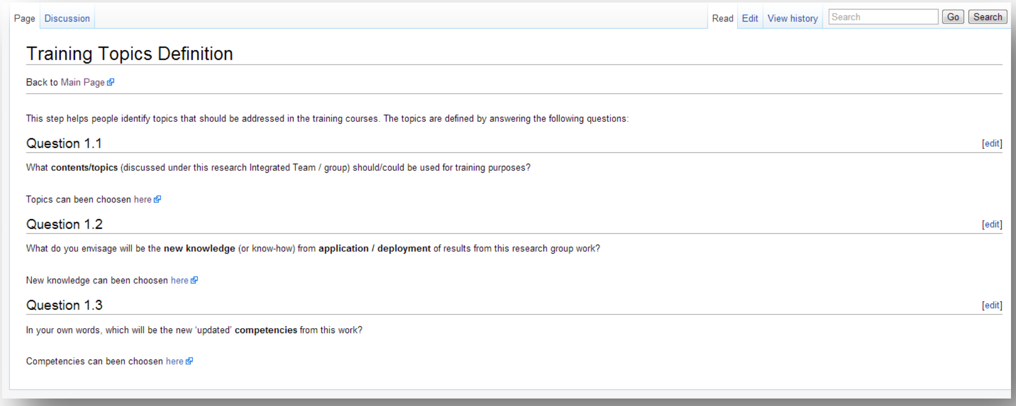

Figure 5.10 - Training Topics Definition page

The moderator controls all the process, has the ability to open voting process, closing it, in some cases open the second voting stage and also show the voting results.

5.1.2.1. Question 1.1

The topics insertion is made through the textbox, for each topic introduced, a page automatically appears.

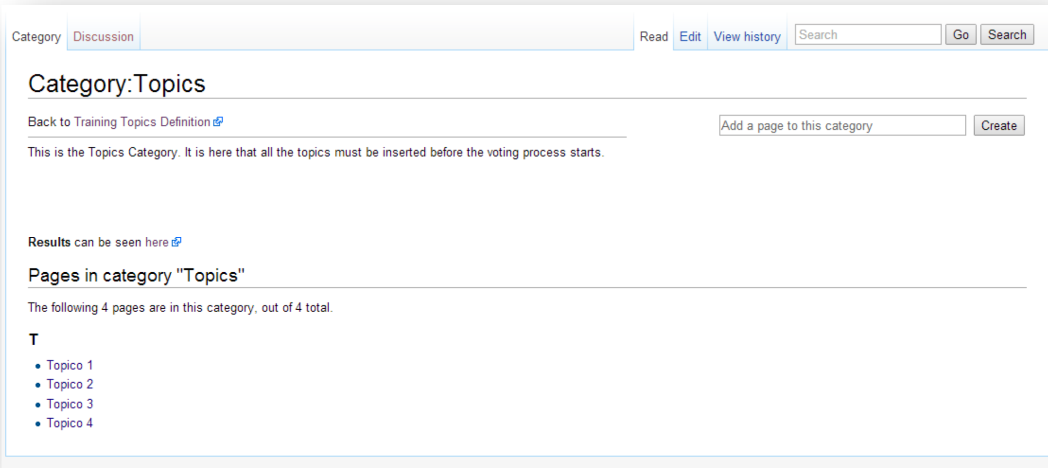

Figure 5.11 - Topics page

Beyond each topic having a description Figure 5.12, it is possible to provide feedback so that a conversation between the collaborators is established. The objective of these conversations is that other training unit members can ask more information about the topic or suggest some alteration they think that may be appropriate.

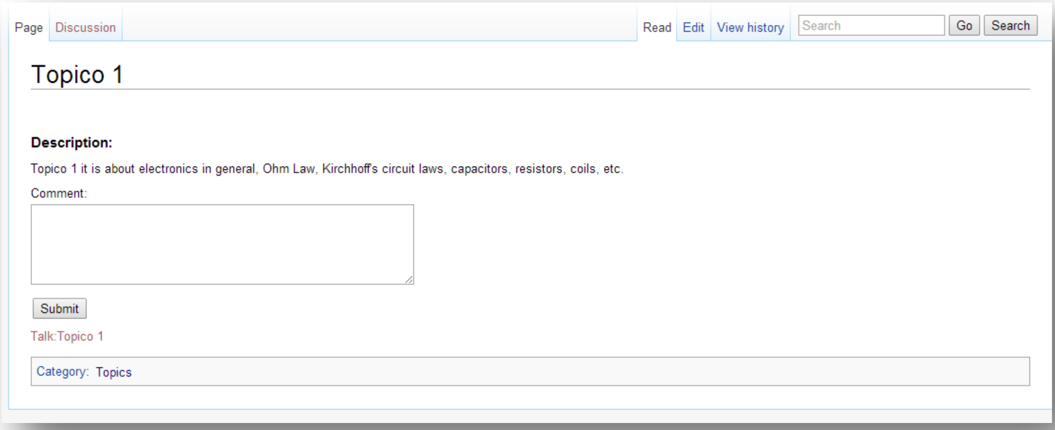

Figure 5.12 - Topic example page

After the moderator opens the voting process through the “Open Voting” button in step 1 on the admin control interface, a vote block appears for each topic page as seen in Figure 5.13. Each collaborator can vote on the topics they think that are appropriate, it is a simple yes or not voting method. The topics that have more than 50% of the votes are the ones that are shown on the result page. This value was chosen by the author and can be changed as needed.

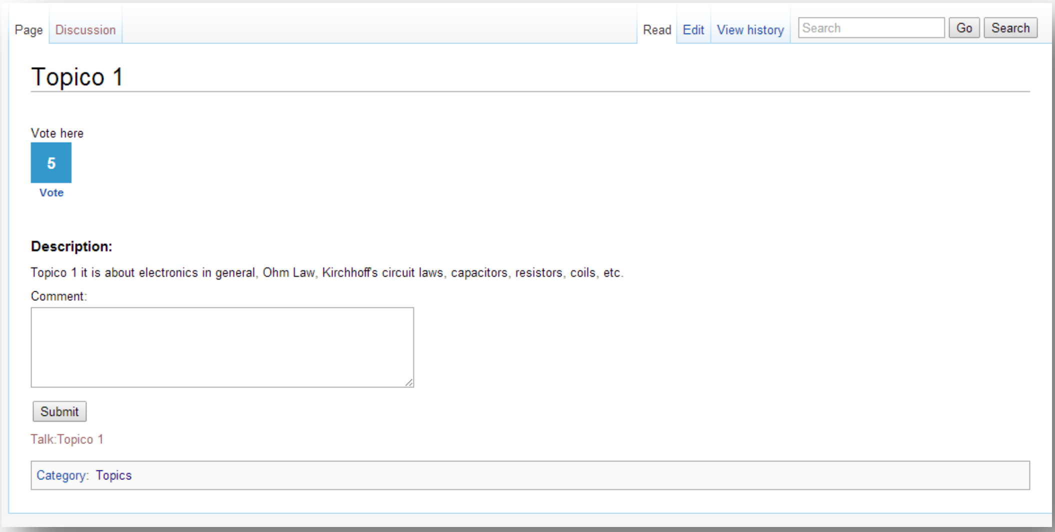

Figure 5.13 - Page during voting stage

5.1.2.2. Question 1.2

The first part of the voting process is equal to the previous question, it only differ on the fact that a second voting process is required.

So after the voting to choose what the “New Knowledge’s”, in the results window are beyond appear those which were the most voted, also appears the chance to vote whether the collaborators agree with the results. If all participants vote yes, the results are saved and the process continues, if not a new voting is opened to choose new solutions for the problem.

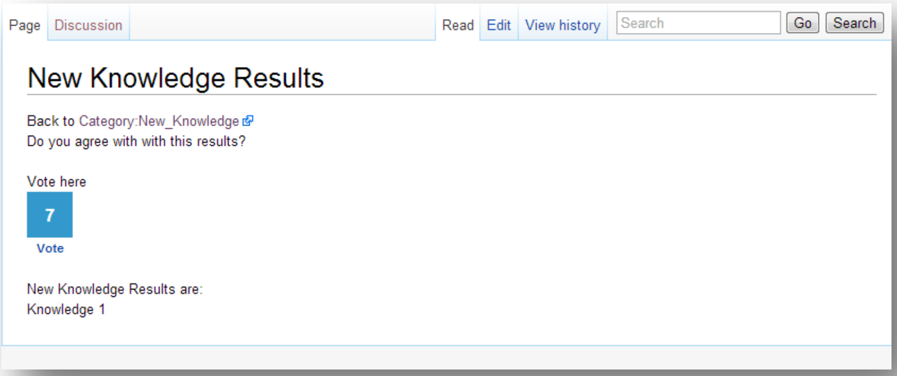

Figure 5.14 - New Knowledge’s results

5.1.3. STEP 2

This is the page relative to Step 2. The process is similar to the previous step, moderator keeps controlling the pace of the process and same voting methods are applied.

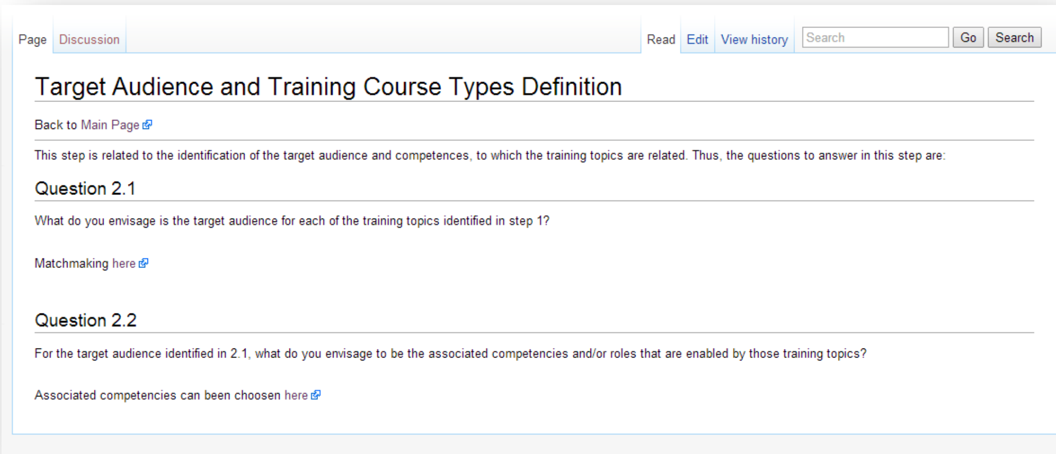

Figure 5.15 - Step 2 question page

In question 2.1, the collaborator is forwarded to the respective page, where it already contains the information resulting from step 1 questions. This is all made in an automatic way; the moderator does not have any role on inserting information.

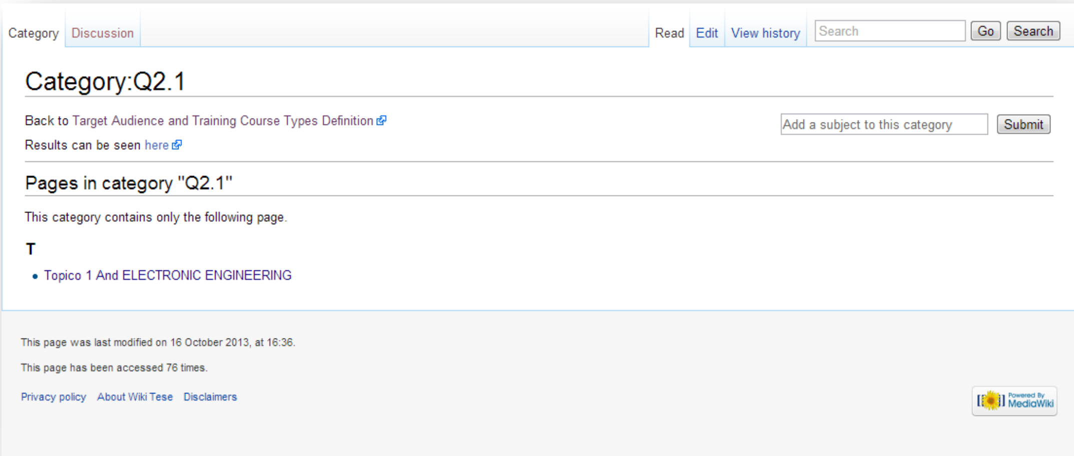

Figure 5.16 - Question 2.1 page

The process repeats for question 2.2, but the main difference is that a competency must be associated to the Topic and Target Audience set (Topic Idea).

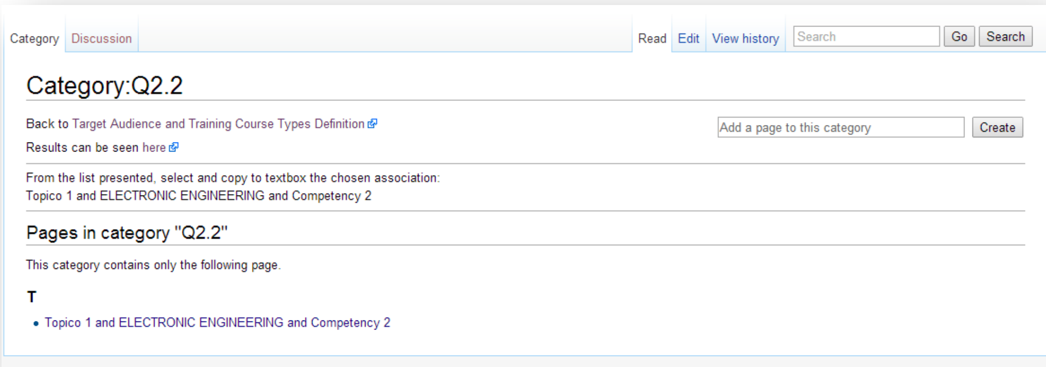

Figure 5.17 - Question 2.2 page

5.1.4.STEP 3

It is on this step that will be chosen the Training objectives, due to each set of topic ideas (Topics + Target Audience) with its associated skills together forming the Associated Competences.

The three fields to fulfil are: 1) Know; 2) Do; 3) Be.

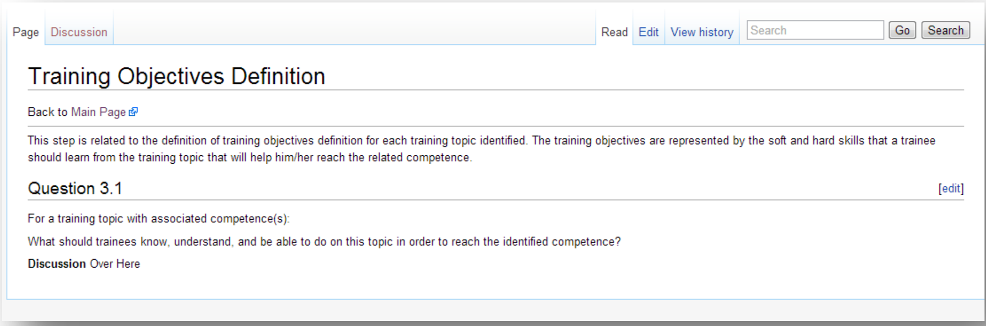

Figure 5.18 - Question 3.1 page

Here each training unit member gives his opinion on which should be the three or more training objectives for each group of topic ideas with associated competences. Then after both voting processes as seen on the diagram previously shown, the results are saved so the final stage of the training course process can be done.

5.1.5. STEP 4

This is the step where the entire construction of the course will be made. It is where the collaborators will organize (topic ideas, statements, competences, and training objectives), identify the general theme and the training modules and finally arrange them in a certain order around the chosen theme. Using all the information gathered on the first three steps, the course construction will be made following sub-steps presented on Figure 5.19.

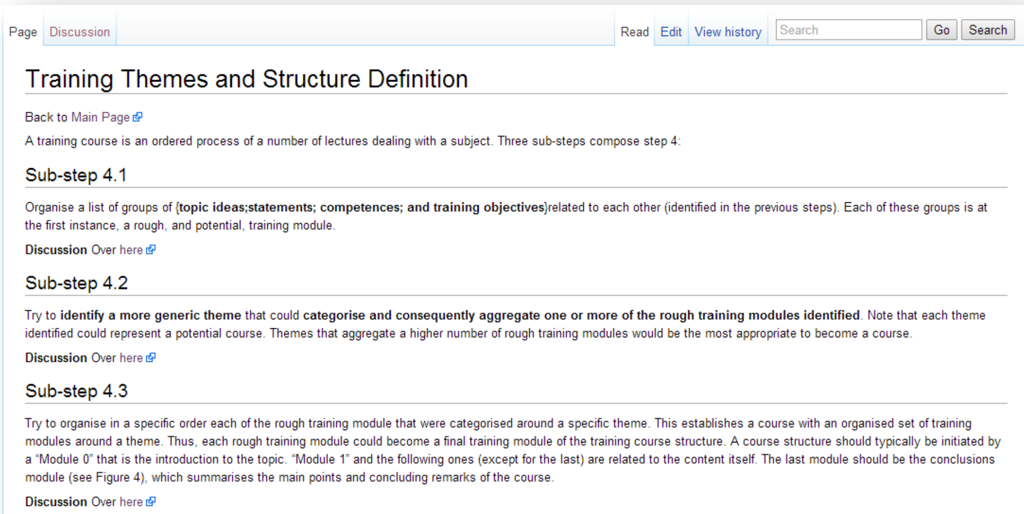

Figure 5.19 - Training Structure Definition

After having the course structure completed, a synopsis must be filled. This is the step 5, where all the course information is transferred to the synopsis model so it can be used on the LMS.

5.2. Prototype and Methodology Deployment

The presented methodology was applied in the CRESCENDO project, which results were presented in the deliverable D1.5.3 [59] of the same project. It presents the CRESCENDO training implementation and an analysis made since the CRESCENDO training courses definition, until the delivery or distribution of such instructional materials, which objective pursued to ensure that training sessions were able to captures trainees' interest.

As all the courses were reviewed by the training unit members, in the following are presented some of the comments and interaction received in each of the steps of the proposed methodology to conduct its training development and implementation:

1. "*…this course (SEDb)[1] doesn't clearly describe its contents, due to the absence of descriptions (narrative text) on the SEDB viewer presentation…*"
2. "*…the 'sota[2] template' training module has to have a small sentence to introduce the module to clearly introduce its contents* …"
3. *"…this training module (BDA Quality Lab - General presentation) starts by presenting some questions to be answered without give any explanations about its purpose. It should start describing something about the training module and only then present those questions…"*

Consequently authors improve their courses to solve all the comments provided by the training unit.

After training unit validation the courses were conducted to the next step, the CRESCENDO Experts validation. On this step some discussions took place. The main contributions were related to the correctness, reliability and clearness of the contents.

After having validated the courses, a pilot training execution was performed. Thus, people that attended the courses provided their feedback by filling up the feedback questionnaire. From the feedback results an overall analysis based on the percentage distribution of the answers was made.

Next figures present some of its feedback questions results, using a percentage distribution.

[1] Systems Engineering Database

[2] State of the Art

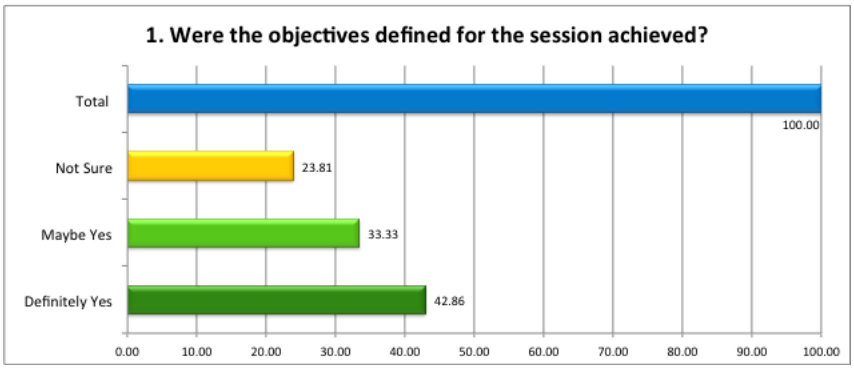

Figure 5.20 - Feedback Question 1

It could be concluded that from the trainees or reviewers point of view analysis, the courses achieve their objectives to some extent (Figure 5.20) and that they are quite well structured (Figure 5.21).

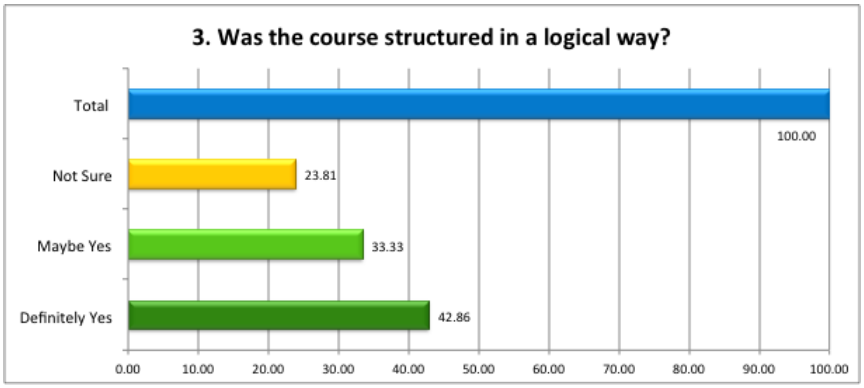

Figure 5.21 - Feedback Question 3

Most of them considered that the concepts and techniques presented are clearly explained (Figure 5.22) and the courses are easy to follow (Figure 5.23).

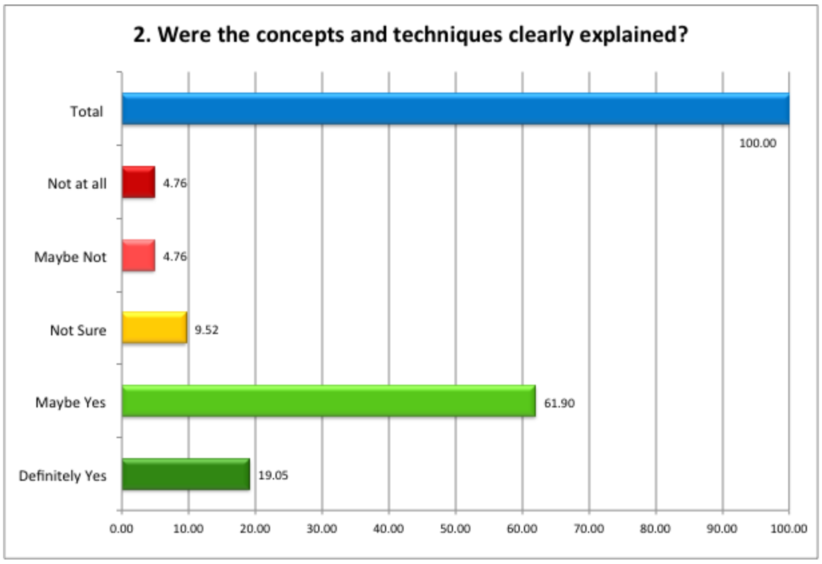

Figure 5.22 - Feedback Question 2

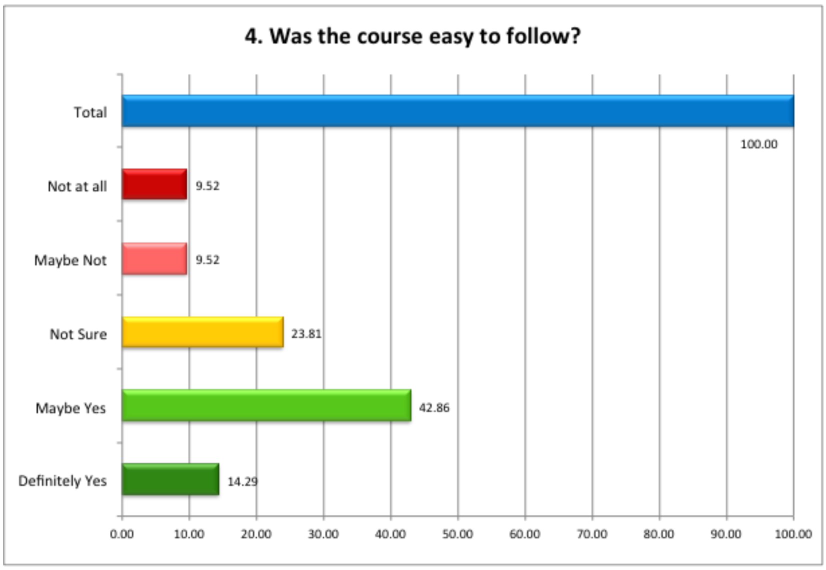

Figure 5.23 - Feedback Question 4

However, trainees are not so much encouraged to actively participate during the course (Figure 5.24). But since the courses are web based this is a quite normal/acceptable result.

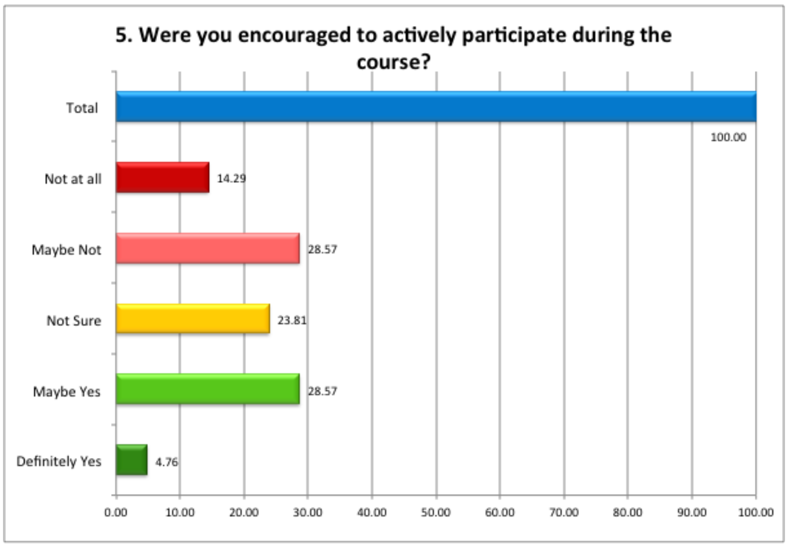

Figure 5.24 - Feedback Question 5

5.3. Dissemination and Hypothesis Validation

The usage of this methodology in CRESCENDO Project proves that it is possible to support the creation of e-learning materials which would be easily adapted for an effective generation of custom-made e-learning courses or programmes. Also the publication of the paper “E-training development approach for enterprise knowledge evolution”[60] serves as a validation of the developed methodology.

It embraces solutions for knowledge management, namely information extraction and formalization from textual documents and methodologies for collaborative e-learning courses development, where main objective is to enable multiple organizations to actively participate on its production. This also pursues the challenge of promoting the development of competencies, which would result from an efficient knowledge-transfer from research to industry.

Concerning the research question, the author gave evidence that after some knowledge acquisition from different sources it is possible to semi-automatically generate a course using the proposed framework, and consequently the presented methodology, that was proven by the Wiki based application prototype that was explained on the previous section.

In line with the hypothesis of this dissertation, by following the methodology and using the wiki platform is possible to facilitate knowledge transfer using a proper framework with the support of knowledge extraction and building of structured courses to guide the training implementation.

The implementation of a semi-automatic application of structured courses creation according to the user needs together with the suggested methodology, allows a capable development of training courses and materials. Thus, conceding a significant flexibility in e-training development, deployment and reuse of the materials available. These factors lead to an easier maintenance of the training system as well as versatility, which are key factors in the e-training acceptance of industry.

6. Conclusions and Future Work

The presented methodology for building e-training courses facilitates the process of creating training materials because it aggregates knowledge acquisition through qualitative information collection methods. Such methods facilitate organizations to participate on the knowledge transfer from research projects results to its identified target audience. The outcome is an approach able to facilitate enterprises to collaboratively create and transfer knowledge between its communities and outside.

The training impact in these specific programmes is proportionally related to the ability, skills and competences that such training could give to an worker, which could directly influence on a job performance, such as operations, human resources policies, or management and leadership [61]. The CRESCENDO training programme is aligned with the previous statement, since it developed a set of course able to offer skills and competences capable of being used to improve a job or a specific task in its domain.

The way as the knowledge is organized in the proposed synopsis facilitates its representation in ontology, which as proposed by Sarraipa et Al. in [52] enables the creation of adaptable e-training services. This provides to the system that could have implemented the proposed methodology, an additional functionality. It will afford enterprises communities that use such system, with a more efficient knowledge transfer instrument to transmit research projects' results to its target audiences.

The proposed knowledge-based framework was developed with the idea to support the creation of e-learning materials, which would be easily adapted for an effective generation of custom-made e-learning courses or programmes and where the main objective was to enable multiple organizations to actively participate on its production. It embraced solutions for knowledge management, namely extraction from text & formalization and methodologies for collaborative e-learning courses development.

As future work author would like to implement this approach in the IMAGINE research project but accomplished with a training evaluation measurement procedure able to quantify the effectiveness of such kind of knowledge transfers.

Also this process is still not fully automatic, since a moderator is needed to keep up with the process, so a fully automatic approach would be an improve on the success and develop of this thesis.

7. REFERENCES

[1] P. Marc, “1st CoSpaces Newlstter,” 2007. [Online]. Available: http://www.cospaces.org/downloads/CoSpaces.

[2] H. Afsarmanesh, L. M. Camarinha-Matos, “ECOLEAD: A holistic approach to creation and management of dynamic virtual organizations,” in *Collaborative networks and Their Breeding Environments*, 2005.

[3] A. Schmidt, “Enabling Learning on Demand in Semantic Work Environments: The Learning in Process Approach,” in *Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications*, J. Rech, B. Decker, and E. Ras, Eds. IGI Publishing, 2008.

[4] J. G. Ruiz, M. J. Mintzer, and R. M. Leipzig, “The impact of E-learning in medical education.,” *Acad. Med.*, vol. 81, no. 3, pp. 207–212, 2006.

[5] C. T. dos Santos and F. S. Osório, “An Intelligent and Adaptive Virtual Environment and Its Application in Distance Learning,” in *Proceedings of the Working Conference on Advanced Visual Interfaces*, 2004, pp. 362–365.

[6] L. M. Camarinha-matos, “SCIENTIFIC RESEARCH METHODOLOGIES AND TECHNIQUES Unit 2: SCIENTIFIC METHOD,” no. PhD Program in Electrical and Computer Engineering, 2012.

[7] “Knowledge Definition.” [Online]. Available: http://www.oxforddictionaries.com/definition/english /knowledge.

[8] S. Grimm, P. Hitzler, and A. Abecker, “Knowledge Representation and Ontologies -- Logic, Ontologies and Semantic Web Languages.” .

[9] J. Sarraipa and R. Jardim-Goncalves, “Semantic interoperability,” *ACM Computing Surveys*, vol. 27, no. 2. pp. 271–273, 01-Jun-1995.

[10] “Typology,” 2013. [Online]. Available: http://en.wikipedia.org/wiki/Terminology#Typology.

[11] “CMU (Carnegie Mellon University)-Domain Dictionary,” 2007. [Online]. Available: www.sei.cmu.edu/domain-engineering/domain_dict.html.

[12] T. R. Gruber, “A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications,” vol. 5, no. April, pp. 199–220, 1993.

[13] H. Mizuguchi and D. Kusui, “An information extraction method for multiple data sources,” pp. 348–353, 2010.

[14] M. Yankova, H. Saggion, and H. Cunningham, “A Framework for Identity Resolution and Merging for Multi-source Information Extraction,” pp. 1367–1372.

[15] M. Labsk, “Combining Multiple Sources of Evidence in Web Information Extraction,” pp. 471–476, 2008.

[16] R. Costa, C. Lima, J. Sarraipa, and R. Jardim-Gonçalves, “Facilitating knowledge sharing and reuse in building and construction domain: an ontology-based approach,” *J. Intell. Manuf.*, pp. 1–20, 2013.

[17] L. Paiva, R. Costa, P. Figueiras, and C. Lima, "Discovering Semantic Relations from Unstructured Data for Ontology Enrichment - Association rules based approach," in *8th Iberian Conference on Information Systems and Technologies*, 2013.

[18] L. Sheng, "A Semantic Vector Retrieval Model for Desktop Documents," *J. Softw. Eng. Appl. 2*, pp. 55–59, 2009.

[19] R. Costa, P. Figueiras, L. Paiva, R. Jardim-Gonçalves, and C. Lima, "Capturing Knowledge Representations Using Semantic Relationships," in *The Sixth International Conference on Advances in Semantic Processing*, 2012.

[20] G. R. Librelotto, P. R. Henriques, and J. C. Ramalho, "TM-Builder : An Ontology Builder based on XML Topic Maps," vol. 7, no. 2, 2004.

[21] J. Isbell and M. H. Butler, "Extracting and Re-using Structured Data from Wikis Extracting and Re-using Structured Data from Wikis," 2007.

[22] B. Leuf and W. Cunningham, *The Wiki Way: Collaboration and Sharing on the Internet*. Addison-Wesley Professional, 2001.

[23] O. Medelyan, D. Milne, C. Legg, and I. H. Witten, "Mining meaning from Wikipedia," *Int. J. Hum. Comput. Stud.*, vol. 67, no. 9, pp. 716–754, Sep. 2009.

[24] R. Navigli and P. Velardi, "Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites," *Comput. Linguist.*, vol. 30, no. 2, pp. 151–179, 2004.

[25] A. Zouaq, R. Nkambou, and C. Frasson, "An Integrated Approach for Automatic Aggregation of Learning Knowledge Objects," vol. 3, 2007.

[26] K. K. Thyagharajan and R. Nayak, "Adaptive Content Creation for Personalized e-Learning Using Web Services," vol. 3, no. 9, pp. 828–836, 2007.

[27] S. M. Asiimwe, "A Knowledge Acquisition Tool to Assist Case Authoring from Texts," 2009.

[28] B. K. C. Ma, "Making the Most of PowerPoint Slides for Rapid eLearning," pp. 1–24.

[29] B. Jaksch, S. Kepp, and C. Womser-hacker, "Integration of a Wiki for Collaborative Knowledge Development in an E-Learning Context for University Teaching," pp. 77–96.

[30] "7 Things You Should Know About Wikis," 2005. [Online]. Available: https://net.educause.edu/ir/library/pdf/ELI7004.pdf.

[31] M. Ebner, M. Kickmeier-Rust, and A. Holzinger, "Utilizing Wiki-Systems in higher education classes: a chance for universal access?," *Univers. Access Inf. Soc.*, vol. 7, no. 4, pp. 199–207.

[32] S. Reinhold, "WikiTrails: augmenting Wiki structure for collaborative, interdisciplinary learning," in *Int. Sym. Wikis*, 2006, pp. 47–58.

[33] D. L. Dalmon, A. A. F. Brandão, S. Isotani, and L. O. Brandão, "Work in Progress - A Framework for Building Interactive Learning Modules," pp. 11–12, 2011.

[34] K. M. Al-shawkani, "E-Learning Authoring Tools : The Present and Future Vision."

[35] M. Haghshenas, M. Khademi, and H. Kabir, "E-LEARNING AND AUTHORING TOOLS : At a Glance," vol. 10, no. 2, pp. 259–263, 2012.

[36] T. Murray, "Authoring Intelligent Tutoring Systems : An Analysis of the State of the Art," vol. 10, pp. 98–129, 1999.

[37] M. Virvou and M. Moundridou, "Adding an Instructor Modelling Component to the Architecture of ITS Authoring Tools," pp. 185–211, 2001.

[38] W. Emigh and S. C. Herring, "Collaborative Authoring on the Web : A Genre Analysis of Online Encyclopedias," vol. 00, no. C, pp. 1–11, 2005.

[39] S. C. Babu, "e -Learning Standards," pp. 1–11.

[40] P. Buitelaar and B. Magnini, "Ontology Learning from Text: An Overview," in *In Paul Buitelaar, P., Cimiano, P., Magnini B. (Eds.), Ontology Learning from Text: Methods, Applications and Evaluation*, 2005, pp. 3–12.

[41] A. Maedche and S. Staab, "Semi-Automatic Engineering of Ontologies from Text," in *In Proceedings of the 12th Internal Conference on Software and Knowledge Engineering*, 2000, pp. 231–239.

[42] D. Faure and C. Nedellec, "Knowledge acquisition of predicate argument structures from technical texts using Machine Learning: the system Asium." 1999.

[43] P. Velardi, R. Navigli, Michele, and Missikoff, "Integrated Approach to Web Ontology Learning and Engineering," 2002.

[44] P. Buitelaar, D. Olejnik, and M. Sintek, "A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis," vol. 5, pp. 31–44.

[45] P. Cimiano and J. Völker, "Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery," in *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, 2005, vol. 3513, pp. 227–238.

[46] "Framework Definition." [Online]. Available: http://whatis.techtarget.com/definition/framework.

[47] D. Levine and A. Gilbert, "Managerial Practices Underlying One Piece of the Learning Organization," *Knowledge Transfer website of Institute of Industrial Relations, University of California, Berkeley*, 1999. [Online]. Available: http://www.irle.berkeley.edu/cohre/knowledge.html.

[48] C. Borysowich, "Using the Nominal Group Method to Drive a Team Decision," 2006. [Online]. Available: http://it.toolbox.com/blogs/enterprisesolutions/.

[49] J. A. Sample, "Nominal Group Technique: An Alternative to Brainstorming," *J. Ext.*, vol. 22, no. 2, 1984.

[50] J.- INTERNET, "Metaplan," 2010. [Online]. Available: http://elearn.jku.at/wiki/index.php/Metaplan.

[51] W. Dick, L. Carey, and J. O. Carey, *The Systematic Design of Instruction*, 5th ed. Longman, 2001.

[52] J. Sarraipa, S. Baldiris, R. Fabregat, and R. Jardim-Gonçalves, "Knowledge Representation in Support of Adaptable eLearning Services for All.," *Procedia CS*, vol. 14, pp. 391–402, 2012.

[53] D. Gazevic, J. Jovanovic, V. Devedzi, and M.Boskovic, "Ontologies for Reusing Learning Object Content," in *International Conference on Advanced Learning Technologies*, 2005.

[54] “OWL 2 Web Ontology Language Overview.” [Online]. Available: http://www.w3.org/TR/owl2-overview/. [Accessed: 03-Oct-2013].

[55] D. Figueiredo, “Customisable e-Training Programmes based on Trainees Profiles,” no. February, 2011.

[56] J. Sarraipa, “Semantic Adaptability for the Systems Interoperability,” 2012.

[57] “University Kentucky -Course Synopsis,” 2012. [Online]. Available: http://www.uky.edu/BusinessEconomics/dssakba/instman/prepare.htm.

[58] Y. Baudier, *BDA Training Policy for the training system and related training material*. CRESCENDO Deliverable D1.5.1., R1.3, 2011.

[59] J. Sarraipa, Y. Baudier, C. Johansson, C. Romesis, C. Pessa. et Al. “Impact of BDA Training Session Reports.” Crescendo Deliverable D1.5.3, 2012.

[60] J. Sarraipa, P. Gomes-de-Oliveira, C. Marques-Lucena, R. Jardim-Goncalves, and J. Mendonça-da-Silva, “E-TRAINING DEVELOPMENT APPROACH FOR ENTERPRISE KNOWLEDGE EVOLUTION,” in *Proceedings of the ASME Congress2013 ASME International Mechanical Engineering Congress and Exposition IMECE 2013*, 2013.

[61] B. Nielson, “Relevance of Learning versus Relevance of Training and Development,” 2009.