EXPLOITING CONCEPT MAP MINING PROCESS FOR E-CONTENT DEVELOPMENT

Chara Panopoulou[1], Natalia Spyropoulou[1], Christos Pierrakeas[1,2], Achilles Kameas[1]

[1]Hellenic Open University (HOU) (Greece)
[2]Dept. of Business Administration, Technological Educational Institute (TEI) of Western Greece (Greece)

Abstract

E-learning has revolutionized education all over the world, defining a different and promising aspect of education, reinforcing the perception that it builds inclusive knowledge societies. Higher Education Institutes (HEI) adopted this innovative education model in order to provide to their students the option of distance education. Since the most important component of e-learning is e-content, its development is a popular research topic in the educational community. Due to the fact that e-content reusability can be increased by using an approach based on Learning Objects (LOs), many methodologies of e-content design introduce guidelines for creating LOs. LOs can make the process of e-learning effective and can offer high quality e-Learning experience to students. The Hellenic Open University (HOU) is introducing LOs in the educational process, using teaching subject domain ontologies to describe them. Ontologies provide a simple way of identifying the knowledge domains covered by LOs, while facilitating its reusability. The preliminary step to create these domain ontologies is the design of the Concept Map (CM), that is a diagram for representing knowledge in a structured form. Concept Maps foster meaningful learning and serve both as a knowledge base for building domain ontologies and as a frame for composing more detailed LOs. Concept Map Mining (CMM), a process for automatic or semi-automatic creation of Concept Maps from documents, is used to facilitate the construction and sharing of Concept Maps. In this paper, we propose a methodological framework and a semi-automatic method for Concept Map creation from unstructured text, which can even handle the morphologically rich Greek language skillfully. The proposed approach combines language processing tools and the knowledge of domain experts. In addition, a case study based on a HOU teaching domain is presented, illustrating the process of Concept Map Mining and showing encouraging results.

Keywords: e-content, e-Learning, Concept Map, Concept Map Mining

1 INTRODUCTION

The digital age has brought upheaval in the fields of education and educational content development. It is a widely held view that the educational content is the main driver of the educational process, especially in distance education, considering that the educational content aims to undertake the largest possible part of the instructor's role in the distance education process. Thus, the focus of many European institutions' interest turned towards the design and development of qualitative digital educational content [1]. Any form of digitized content that can facilitate the learning process can be defined as e-content [2].

E-Content can effectively enhance the online academic course creation offering flexible access to learning opportunities without the time and distance barriers. Higher Education Institutions (HEIs) incorporate online courses in their curriculum [3] and develop e-content in order to complement these courses [24]. As the content development plays a key role in e-learning, the e-content must be designed properly [4]. One way is the e-content to be designed and developed in smaller, modular, discrete units of learning known as Learning Objects (LOs) [24]. Therefore, HEIs have adopted instructional technologies in e-content development such as the Learning Objects.

Learning Objects could enable higher education to capitalize on the promise of e-learning, as their size and manageable format allow the re-use of material and offers the possibility to be searchable and accessible from everywhere to a wide audience [5]. The specification of a standardized set of metadata contributes to improving LOs reusability [6].

With the usage of ontologies, the representation of learning resources is attributed to a finer level of detail, providing accuracy and flexibility for LOs metadata [8]. However, ontologies are used not only to support LOs' interoperability describing their structure [7], but also to build Learning Object content [6]. Furthermore, ontologies are often used to capture and share the knowledge that is revealed through the Concept Maps (CMs). CMs have traditionally been used as cognitive tools in the educational process, organizing and representing the content of a knowledge domain through concepts and relationships linking them. [9]. Moreover, CMs facilitate the complexity of formal ontology design in the process of an educational content representation because they can serve as a knowledge base for building domain ontologies [10]. Consequently, due to the important role of Concept Maps in the e-content development, the highlight of the Concept Map Mining process is essential. The Concept Map Mining process is termed as automatic or semi-automatic extraction of Concept Maps from text [11].

Based on the aforementioned educational framework for e-content development, the purpose of this paper is to outline an approach in Concept Map Mining process followed by the Hellenic Open University (HOU), in order to create Concept Maps for the educational content design process. The combined use of Concept Maps, ontologies and LOs facilitate the creation and distribution of e-Content.

Following this introduction, the paper is structured as follows. The second section is addressed to concept map mining process and the several aspects to it. In the third section, we present the proposed approach for the CMM process in order to build a concept map. The implementation of the proposed approach is analyzed in the fourth section, as well as the problems we faced for the Greek language throughout the process. Section five describes the evaluation framework applied, aiming to evaluate the performance of the proposed method. A brief summary of the proposed method and a plan for future research activities are given in the last section.

2 A BRIEF OVERVIEW IN CONCEPT MAP MINING

Bearing in mind the previous points, in this section we focus on the importance of the concept maps in the e-content design. CMs are a knowledge representation tool, which provides an overview of the course content [12], as well as it is an alternative way to structure content according to Baylen et al. [13] who created concept maps to organize content for a graduate-level course. In this way, CMs can contribute to developing e-Content.

Concept map mining is a process applied to unstructured textual sources and as a result, extracts concept maps from them. The CM should be an accurate visual abstract of a source text [14]. There are several approaches in the Concept map mining process, such as automatic and semi-automatic creation of concept maps from textual and even non-textual sources [15]. In a semi-automatic process, the system finds and suggests elements of a map, and a person manually has to finish the map using the provided information. In this process, several statistical and data mining techniques are used in combination with linguistic tools. In the automatic construction process, the user's assistance is not required and the process creates the map automatically from the available resources. Although the phases of a CMM process are almost specific, adjustments are allowed accordingly the approach of concept map construction.

Describing the CMM process, Žubrini´c [15] presented the document as a superset of the sets of concepts and relationships used to connect these concepts. According to a strictly mathematical viewpoint, a document can be formalized as follows: $\mathrm{D} = \{\mathrm{C}, \mathrm{R}\}$. The element C, $\mathrm{C} = \{\mathrm{c}_1, \mathrm{c}_2, \dots, \mathrm{c}_\mathrm{n}\}$, is a set of all concepts that exist in the document, whereas the element R, $\mathrm{R} = \{\mathrm{r}_1, \mathrm{r}_2, \dots, \mathrm{r}_\mathrm{m}\}$, is a set of all relationships that link the concepts between them. In the first phase of the process, the identification and extraction of the concepts and relationships are completed, thus defining the sets C and R. The members of C set are concepts, usually nouns or noun phrases. Likewise, the members of R set are the extracted links or, in other words, the relationships. Each member of that set connects two members of the first set indicating the existence of a proposition in the document. Consequently, the sets C and R include the candidate terms, concepts and verbs, respectively, that will be the constructional materials of the CM. The next phase of the CMM process is a summarization of the information extracted from document D which aims to identify the subsets C_d and R_d. These subsets contain the terms that finally chosen to be used in the CM design by the total of the potential terms included in the sets C and R.

To visualize the previous mathematical viewpoint, it can be considered that the concepts serve as nodes and the verbs as arcs that connect the concepts creating a web of concepts. CMs must also have a topology [11], [15]: more general concepts should be higher on the map, and more specific concepts on lower levels. Concepts with the same level of generalization should be at the same level of the topology [11]. Therefore, the output of the CMM process should comprise: Concepts, Relationships and a Topology [11]. The final phase of the CMM process is the creation of hierarchical CM, which can be defined as a set $\mathrm{CM} = \{\mathrm{C}, \mathrm{R}, \mathrm{T}\}$ that contains concepts, relationships and topological information of the map, where the set $\mathrm{T} = \{\mathrm{t}_1, \mathrm{t}_2, \ldots, \mathrm{t}_\mathrm{k}\}$ is a sorted set of generalization levels [11].

Briefly, a CM construction implies the following steps: extraction of the terms of the document and after being filtered selecting the most important ones (Summarization), then organizing them in a graphical representation. In this graphical representation, pairs of concepts and their linking phrases frame propositions in the form of a triplet, $\mathrm{P} = \{\mathrm{c}_\mathrm{p}, \mathrm{r}_\mathrm{i}, \mathrm{c}_\mathrm{q}\}$, where c_p and c_q are elements (concepts) of the C set, and r_i is element (relationship) of the R set. These triplets are considered to be the basic units of meaning [16]. For a visual representation of the CMM process, the phases are illustrated in Fig. 1.

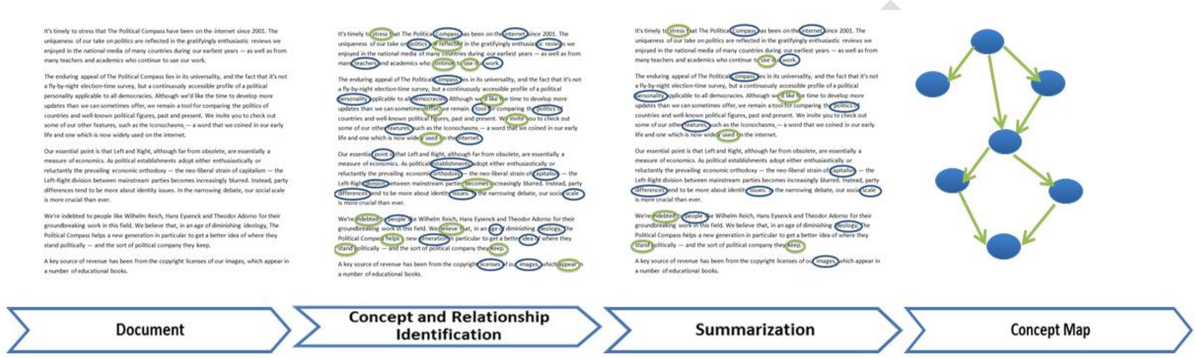

Figure 1: The phases of CMM process

3 THE PROPOSED APPROACH FOR CMM

The efficient management of terminology is a key contributor to quality and consistency of content development [17]. Due to the fact that the terminology used in a document is important in the educational context, we create CM representations using terms that are extracted from the original text based on their incidence rate. We advocate that the incidence rate of a term in a document, possibly, indicates that this term could be a key concept of the knowledge domain the document associated. The next section proposes and describes an approach of CMM procedure for semi-automatic creation of CMs from unstructured texts, which after being applied to a document, a list of the most frequently used concepts will be resulted. The resulted list will contain potential concepts that could be used in the CM construction.

3.1 General Procedure

Cañas et al. (2004) [18] suggests an approach of the CMM process in which the terms are automatically extracted, and are candidates for concepts, following a manual construction of the CM by the selected concepts. Towards Cañas's propose and having considered the general phases of CMM procedure, we outline our approach in the CMM process applying some adjustments to the phases accordingly the principles followed to create CMs from unstructured texts.

Our methodological framework of CMM process is based on two main principles: a) the terminology significance and b) the convenience offered in the process by the list of the suggested concepts of a teaching domain which resulted from the educational material (digital format) provided for this teaching domain. Based on the assumption that the terminology used in a document concerns the key concepts of a domain and the fact that the key concepts are the most frequently used terms in a document because of the important role in this [25], we propose a methodological framework, which suggests the use of a list of concepts extracted from a document during the CMM process. This list of concepts allows the CM constructor to concentrate on the meaning-making process of linking the concepts to form propositions and the CM's structure, reducing the concern that some key concepts have probably been missed.

On the other hand, statistical and data mining techniques are usually used in many approaches of the CMM process from unstructured text, and are computationally efficient in the frequency measurement of terms and their co-occurrence in a document. Perhaps, the most major advantage of these techniques is that they are language- and domain- independent, giving the opportunity to these techniques to help in CMM of documents in different languages and different knowledge areas. However, these methods are often not enough precise and the use of linguistics tools is recommended for their precision improvement [15].

Therefore, the proposed approach is hybrid because combines the general phases of CMM process, text mining techniques, the knowledge of domain experts and a linguistic tool for a more precise extraction of concepts and relationships from unstructured text. The proposed CMM procedure creates CMs taking as input one document and obtaining a single CM as an outcome. Also, this approach is suitable for the mining of texts independent of their language origin. The main steps of the proposed procedure are shown in the Fig 2. and a description of them follows.

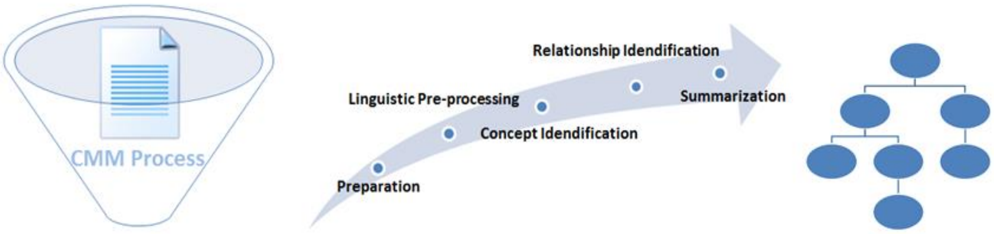

Figure 2: The main steps of CMM procedure for unstructured textual sources

A. Preparation Phase:

In the first phase of the proposed procedure, we identify the documents that are suitable for analysis. Due to the purpose of this research, data sources used for the analysis are digital educational material provided in the HOU's curriculum.

The simple form of digital educational material, such as a digital book, lags behind the Learning Objects, which is the most recent way of thinking about educational content. Moreover, CMs can be an important intermediate stage of the educational material conversion from one form (simple digital form) to another (Learning Object form). CMs can support a better understanding of the specific documents, and lead to upgrading the quality of educational material.

Consequently, in order to be a document valuable source of information for CMM processing, some preceding tasks must be completed. Some of these tasks are the remove of all elements without information such as pictures, tables and references. One even task is the removal of text formatting in order not to affect the results of the process. Moreover, all words are converted to lowercase.

Finally, the resulting text, enriched with semantic information, is stored and ready for processing.

B. Linguistic Pre-processing Phase:

During the linguistic pre-processing phase, a linguistic tool and a stemming algorithm cooperate in order to be completed some linguistic tasks regards the words of the document.

At the beginning of this phase, a stop list must be created in order to remove all the remaining words without information value. Among these words are parts of speech, such as pronouns, conjunctions, articles, quantifiers, adverbs and prepositions. Although, the verbs usually give important information in the Relationships Idendification Phase, in the framework of our research, the verbs are also removed.

The linguistic tool used called Atlas.ti [20]. Atlas.ti is a versatile workbench that offers a variety of tools for qualitative analysis of data, accomplishing the tasks associated with any systematic approach to unstructured data that the formal, statistical approaches cannot meaningfully analyze. Among the tools provided by Atlas.ti in textual information processing level, there is the "Word Cruncer". This tool creates a list of words showing their frequencies within the preprocessed document, offering a simple quantitative content analysis. Additionally, the results can be directly exported into an Excel spreadsheet. As it shown in Fig. 3, each column of Excel spreadsheet represents how many times each word (Column A) occurred in each document (Columns B, C, etc…).

	A	B	C	D	E	F	G	H	I
1	Words	P 1	P 2	P 3	P 4	P 5	P 7	P 8	P 9
2	AAHSL	0	0	0	0	0	0	0	0
3	AAT	0	0	0	0	0	1	0	0
4	AAUP	0	0	0	0	0	1	0	0
5	ABATING	0	0	0	0	0	0	0	0
6	ABBAS	0	0	0	0	0	0	0	0
7	ABBOTT	0	0	0	0	0	0	0	0
8	ABBREVIATION	0	0	1	0	0	0	0	0
9	ABC	0	0	0	0	0	0	0	0
10	ABDICATE”	0	0	0	0	0	0	0	0
11	ABI	0	0	0	0	0	0	0	0
12	ABILITIES	0	0	0	0	0	0	0	0
13	ABILITY	0	0	1	1	4	5	3	4
14	ABLE	0	0	1	1	1	0	5	2
15	ABLEX	0	0	0	0	0	0	0	0
16	ABLY	1	0	0	0	0	0	0	0
17	ABOUND	0	0	0	0	0	0	0	0
18	ABOUT	2	8	13	7	11	20	6	8
19	ABOUT”	0	0	0	0	0	0	0	0

Figure 3: ATLAS.ti’s resulted Excel spreadsheet

Furthermore, Word Cruncer gives the option either to use the predetermined ATLAS.ti stop list or to create a new one, which excludes the stop words without information value, such as punctuation marks, certain words, characters or patterns, from the frequency count. The predetermined ATLAS.ti stop list appears to constitute the common English stop list. To create a new stop list suffices to formulate both words and patterns in regular expressions (GREP), which allow to remove even more complicated patterns. Towards this direction, the stop list used is that created at the beginning of this phase, which contains all the words and patterns that we want to be excluded from the results.

In the next step, all words must be normalized. This is achieved by reducing inflected words in their common base form using lemmatization or stemming techniques. The applied stemming algorithm follows the lines of Ntais algorithm [19], which is a rule-based system serving the requirements of the Greek Language. Since the words are reduced to their base form, duplicate words are generated. If the repeated words of the text will be merged and their frequencies will be added, then only different words will compose the list.

By removing all these words and using this light-stemming technique, the computational cost of the stemming algorithm is reduced, reaching higher precision in the results. If the target language is different from Greek, there are implementations of stemming algorithms that are available in other languages.

At the completion of this phase, sorting the results of Word Cruncher according to their frequency, a sorted list of words in descending order is created. The terms of this list are candidates in CM construction.

C. Concept identification Phase:

In the concept identification phase, we identify the candidate concepts for CM construction through the list of the most occurring words emerged in the previous phase. All words that frequently occur in a document are marked as potential candidate concepts in CM construction. Potential candidate concepts are the terms with a frequency score higher than a manually determined threshold. All the words with a smaller frequency score than 20 occurrences in the document are removed. All the remaining words of the list are candidate concepts.

Afterwards, this list is recommended to the domain experts who are asked to derive concepts from this list in order to create the CM, which will be used in their domain knowledge representation. However, this list is not exhaustive. The domain experts have the option to add a word that according to them should be used for the CM development, but not included in the recommended list.

D. Relationships identification Phase:

Cañas et al. (2004) [18] refers that the most difficult part of constructing a concept map is to link the concepts through a relationship forming units of meaning. By linking the concepts, a coherent structure is created that reflects understanding of a specific domain. In the relationship identification phase, candidate relationships are usually verbs of the document that semantically connect the extracted concepts (that are concepts of the document). Specifically, the main verb used in a sentence is more likely to be a candidate relationship between the concepts that exist in the same sentence. The use of POS (Part-of-speech) tagging process and TF-IDF (Term Frequency–Inverse Document Frequency) indices calculate the frequency of relationships' appearance among specific concepts, helping in the relationship extraction, however these processes are complex to work in Greek language. Therefore, we omit the standard procedure of this phase and instead, we seek for the help of domain experts to link the concepts that will have already selected from the recommended list. Instead, we provide domain experts with a predetermined list of recommended relationships which resulted from an effort of the HOU to standardize and define the relations that have been used for knowledge representation of numerous cognitive (teaching) domains [21]. The use of the list is mainly provided to assist the domain experts due to the important information offered.

E. Summarization Phase:

During the final phase of summarization, the domain experts are responsible for the final selection of the most important terms from the lists given to them in the phases of Concept and Relationship identification. The concepts and relationships are chosen according to the domain expert opinion; and the selected terms serve as a base for CM construction. We involve the domain expert in CMM process, as we consider that they are the most critical factor which could lead the process, providing their special skills and relevant conceptual knowledge in a particular knowledge area.

4 IMPLEMENTATION OF THE PROPOSED APPROACH IN GREEK LANGUAGE

The proposed approach can operate in several languages by creating a new stop list and by simply adjusting the rules of stemming algorithm. However, this paper proposes and describes a CMM procedure for semi-automatic creation of CMs from unstructured texts in morphologically rich Greek language. As mentioned earlier, the documents processed are teaching (cognitive) domains of HOU's curriculum and, specifically, this implementation is based on the study of the cognitive domain "The field, the learning principles and the stakeholders of adult education" of HOU's teaching module EKP64 "Introduction to Adult Education". Applying the proposed method to this cognitive domain, a list of 50 potential candidate concepts emerged and the domain expert selected 31 of them. During the implementation of the methodology, many problems are mainly caused due to the Greek language, which is a highly inflected language.

4.1 Problems in Greek Language

As far as the implementation of this method in the Greek language is concerned, some problems arose during the Linguistic pre-processing Phase. Although, the use of language processing tools, such as tokenizers, stemmers, part-of-speech (POS) taggers and parsers, improves the processing results, there is a limiting factor that prevents the use of such tools and techniques. The majority of available linguistic tools and methods used in CMM process are mainly based on the English language [15], creating problems during the analysis of other languages. The suggested method is easily customizable in order to operate in other languages, adjusting only the rules of the used linguistic techniques to serve the needs of each language.

5 EVALUATION

The most well known, commonly accepted, and widely applied performance metrics in the Information Retrieval (IR) fields are the Precision, Recall and F-Measure [22]. In order to evaluate the described method, we use the aforementioned metrics which are defined to serve the suggested CMM framework.

Precision (equation 1) is defined as the fraction of the number of selected concepts from the list of the recommended concepts divided by the total number of recommended concepts of the list. On the other hand, Recall (equation 2) is defined as the fraction of the number of selected concepts from the list of the recommended concepts divided by the total number of concepts used in the CM construction.

$$Precision = \frac{a}{a+b} \quad (1) \qquad Recall = \frac{a}{a+c} \quad (2)$$

These equations use the variables “a”, “b” and “c”, where, the “a” is defined as the number of the correctly appeared concepts of the list that used in the CM construction, the “b” is defined as the number of incorrectly appeared concepts of the list, which not used in the CM construction and finally the “c” is defined as the number of the concepts used in the CM construction, but not appeared in the list. In Fig.4 the variables “a”, “b” and “c” are depicted trough the modified figure of Wikipedia[1], which describes the metrics Precision and Recall.

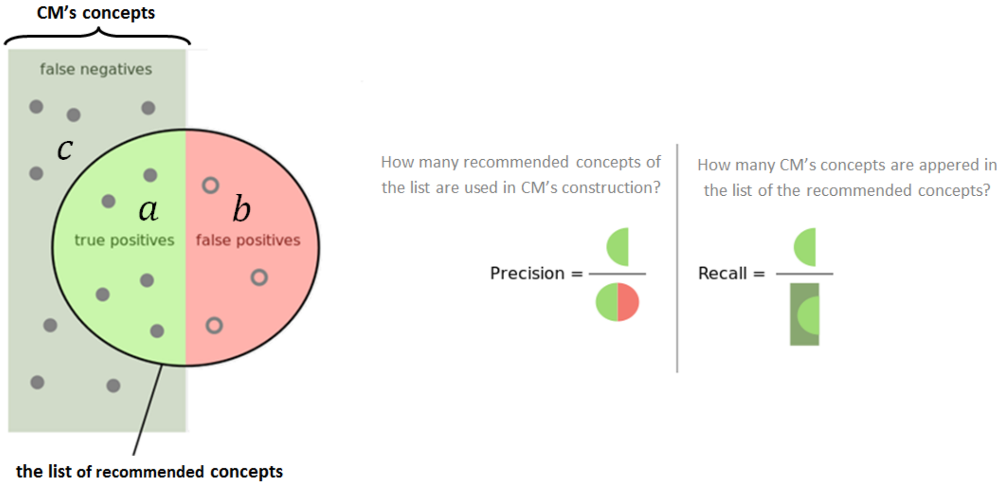

Figure 4: Evaluating the suggested CMM process

The above quantitative measures, Precision and Recall, were computed and the results indicated that the proposed method can achieve a satisfactory performance in CMM process. Precision and Recall are calculated 0.62% and 0.78%

For a good balance between the precision and recall values, the F-measure [23] (equation 3), combines through a function both the precision and recall, providing a harmonic mean of them. As higher the F-measure value, the better the result of the suggested CMM process.

$$F_{measure} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)$$

6 CONCLUSION

This paper is a modest contribution to the ongoing research on concept mapping. Initially, we point out the significant importance of e-content in the e-learning process, mentioning the emerging technologies used in e-content development. One of them is the concept maps which is an intuitive form for representing and organizing knowledge. Thus, our attention was focused on CMM process, suggesting an approach of CMM process for the semi-automatic creation of a CM from unstructured text, as well as describing the problems being faced during the implementation of this in the Greek language. Based on the obtained results, it can be concluded that the proposed method in CMM process, which presents an approach to CMs creation, combining the knowledge of the domain experts in a specific domain with linguistic tools, is proved very promising. On the basis of the promising results presented in this paper, the research into the CMM process is continuing, applying the suggested approach to the other teaching (cognitive) domains of HOU and presenting more results in future papers.

ACKNOWLEDGMENTS

This research described in this paper has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) (Funding Program: “HOU”).

---

[1] (http://commons.wikimedia.org/wiki/File:Precisionrecall.svg)

REFERENCES

[1] Fernández-Manjón, B., Sánchez-Pérez, J. M., Gómez-Pulido, J. A., Vega-Rodríguez, M. A., & Bravo-Rodríguez, J. (2007). Computers and Education: E-learning, From Theory to Practice (6), pp. 67–76.

[2] Lizeron Eremias, & Subash, R. (2013). e-Content Development: A Milestone in the dynamic progress of e-Learning. International Journal of Teacher Educational Research (IJTER), Volume (2), No.(1), pp. 43-49.

[3] Kim, K., & Bonk, C. J. (2006). The future of online teaching and learning in higher education: The survey says. Educause quarterly 29(4), pp. 22-30.

[4] Ally, M. (2004). Foundations of educational theory for online learning. Theory and practice of online learnin 18(1), pp. 15-44.

[5] Griffith, R. (2003). Learning objects in higher education. Academic Advanced Distributed Learning Co-Lab publication, University of Wisconsin (April 2003), http://www. academiccolab. org/resources/webct_learningobjects. pdf.

[6] Gasevic, D., Jovanovic, J., Devedzic, V., & Boskovic, M. (2005, July). Ontologies for reusing learning object content. In Advanced Learning Technologies. ICALT 2005. Fifth IEEE International Conference, pp. 944-945. IEEE.

[7] Behr, A., Primo, T. T., & Vicari, R. (2014). OBAA-LEME: A Learning Object Metadata Content Editor supported by Application Profiles and Educational Metadata Ontologies. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação, Volume (3), No.(1).

[8] Dicheva, D., Mizoguchi, R., & Greer, J. E. (Eds.). (2009). Semantic web technologies for e-learning, Volume (4), 15(11), pp. 19-218. Ios Press.

[9] Carnot, M. J., Feltovich, P., Hoffman, R. R., Feltovich, J., & Novak, J. D. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. Pensacola, FL.

[10] da Silva, A. A., Padilha, N., Siqueira, S., Baião, F., & Revoredo, K. (2012, September) . Using Concept Maps and Ontology Alignment for Learning Assessment.

[11] Villalon, J. J., & Calvo, R. A. (2008, December). Concept Map Mining: A definition and a framework for its evaluation. In Web Intelligence and Intelligent Agent Technology. WI-IAT'08. IEEE/WIC/ACM International Conference on, Volume (3), pp. 357-360. IEEE.

[12] Hartsell, T. (2015). Mapping Concepts for Learning. In Essentials of Teaching and Integrating Visual and Media Literacy, pp. 203-215. Springer International Publishing.

[13] Baylen, D. M., & D'Alba, A. (2015). Essentials of Teaching and Integrating Visual and Media Literacy: Visualizing Learning. Springer.

[14] Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. Florida Institute for Human and Machine Cognition Pensacola Fl, www. ihmc. us.[http://cmap.ihmc.us/docs/theory-of-concept-maps], 284, 16.

[15] Zubrinic, K., Kalpic, D., and Milicevic, M. (2012). The automatic creation of concept maps from documents written using morphologically rich languages. Expert Systems with Applications, 39(16), 12709–12718. doi: http://dx.doi.org/10.1016/j.eswa.2012.04.065

[16] Broers, N. J. (2009). Using propositions for the assessment of structural knowledge. Journal of Statistics Education, 17(2), n2.

[17] Gornostay, T. (2010, March). Terminology management in real use. In Proceedings of the 5th International Conference Applied Linguistics in Science and Education, pp. 25-26.

[18] Cañas, A. J., Carvalho, M., Arguedas, M., Leake, D. B., Maguitman, A. & Reichherzer, T. (2004), Mining the web to suggest concepts during concep map construction. In Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, pp. 135–142.

[19] Ntais, G. (2006). Development of a Stemmer for the Greek Language (Doctoral dissertation, Royal Institute of Technology).

[20] Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. Applications of social research methods to questions in information and library science, pp.308-319.

[21] Panopoulou, C., Pierrakeas, C., & Kameas, A. (2014, October). The Extented Correllation Verb Ontology: A New Approach for Object Properties Specification within Domain Knowledge Representation. AC 2014: 11th International Conference Applied Computing, pp. 11-19.

[22] Su, L. T. (1992). Evaluation measures for interactive information retrieval. Information Processing & Management, 28(4), pp. 503-516.

[23] Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association, 12(3), pp. 296-298.

[24] Govindasamy, T. (2002). Successful implementation of e-Learning Pedagogical considerations. Internet and Higher Education, 4, pp. 287-299.

[25] Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. International Journal on Artificial Intelligence Tools, 13(01), pp.157-169.