Bootstrapping Domain Knowledge Exploration using Conceptual Mapping of Wikipedia

Mai Eldefrawi
Information System Department,
Faculty of computers and information, Helwan University
Cairo, Egypt

Ahmed Sharaf eldin Ahmed
Information System Department,
Faculty of computers and information, Helwan University
Cairo, Egypt

Adel Elsayed
Learning Systems International,
Leeds, UK, Formerly Research Leader, M3C Lab, University of Bolton, UK

***Abstract*—Wikipedia is one of the largest online encyclopedias that exist in a hypertext form. This nature prevents Wikipedia's potential to be fully discovered. Therefore the focus of this paper is on the role of domain knowledge in supporting the exploration of classical encyclopedic content, which in this case is Wikipedia. A main contribution provided by the author of this work is a methodology for identifying the nature, the form and the role of domain knowledge expressed in conceptual form. It's also a method of representation and analysis for describing the domain knowledge and for the extraction of the logical representation of a raw form of the domain knowledge. Such logical representation is of limited value in describing the real nature of domain knowledge. Hence we transform it into an adequate graphical representation, mostly of an arc-node form which is called conceptual representation.**

Keywords—Conceptual Mapping; Conceptual Representation; Domain knowledge; Wikipedia; Self-regulated learners.

I. INTRODUCTION

Domain knowledge exploration is one of key elements for learners specially when exploring open sources of knowledge. This exploration is not only a natural step for learners wishing to acquire new information related to their exploration interest, but it's a gate for adding new perspectives to their current knowledge from the wide range of available resources. The rapid technological age we are living has made it more demanding to develop new approaches and methodologies for exploring the knowledge available on the web. This requires a proper support for learners when exploring the web especially when they only have limited knowledge of a subject domain [11]. Learners who depend on online resources are usually self-regulated learners who need proper guidance. These online resources provide a rich and prosperous environment, but if not well managed learners can face a cognitive overload and distraction [13].

One of the online resources that are already available to help users get introductory information about specific domain is Wikipedia; Wikipedia is an online encyclopedia with around 30 million articles in 286 languages [22]. It's written collaboratively by volunteers around the word. Wikipedia has become very popular on the internet, with 365 million readers worldwide and ranked sixth among all worldwide websites. Wikipedia is available in a hypertext mode. The main disadvantage of Wikipedia that when navigating through its links each link takes the user to a different context far from the main idea being explored. As a result, there is no clear linkage between each topic and other external topics that could be of great importance to the reader. However the main power of Wikipedia is in containing a vast amount of concepts.

The open nature of Wikipedia raised concerns regarding the quality and consistent of information. These concerns led to an investigation conducted by Nature journal that showed that science articles have a very close accuracy rate to that of Encyclopedia Britannica [22]. Wikipedia articles are loosely organized; however any article usually starts with a short paragraph that summarizes the main features of an article with some definitions and hyperlinks. The scope of this paper considers this introductory paragraph of scientific domains within the analysis.

Wikipedia offers two features for navigating across article, the first feature is hyperlinks. Hyperlinks are merely an entry for other articles that exist in Wikipedia and mentioned by the name in the current article. The other feature is Categories [14], most users who are interested in finding the relations between topics and each other investigate the Wikipedia category tree.

Wikipedia category trees are arranged in the form of: main category>> a number of sub-categorical levels >> Pages. This hierarchy is the only arrangement for Wikipedia category tree, so an article is placed under another if according to the author of the article they are related. These articles may not necessarily be having a hierarchal structure according to their content modeling.
However this categorization techniques isn't efficient enough for the following reasons.

- These categories are arranged by the author's point of view, this means that important articles may not be included in a category or vice verse.
- There is no clear criterion for including topics within categories and no provision over the referencing.
- Categories include entries under different contexts that are not explicitly stated, the name of an article is only mentioned with no elaboration [14].

Wikipedia categories trees are considered the most appropriate method for navigating a certain domain of knowledge within Wikipedia. Thus Wikipedia Categories are the focus of the analysis conducted in this paper.

Conceptual representation is an area that has been attracting researchers recently. As concepts form the base for human's cognition and graphical representation is usually more appealing for learners to read and understand, the use of graphical representation for concepts to support cognitive process has been the core of several studies. Novak defines concepts map as a graphical tools for representing knowledge. A concept map includes concepts contained in circles or boxes of some type. Relationships exist between pairs of concepts; these relationships are represented as a link connecting the related concepts. On the link the connecting word or phrase is placed which indicated the type of the relationship between the two concepts [21].

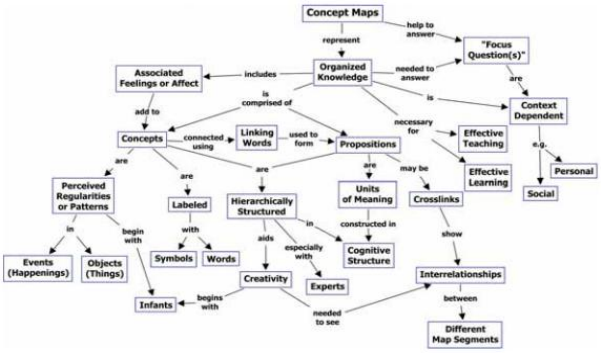


Fig. 1. A concept map showing the key features of concept maps. Concept maps tend to be read progressing from the top downward [21]

Our motivation is to benefit from Wikipedia by combining the advantage of it along with the advantage of conceptual representation. We are proving that using conceptual mapping for the analysis and exploration of a certain domain of knowledge within Wikipedia provide the user with the ability to navigate and explore through concepts, in a manner that couldn't be achieved by normal means. We are also comparing the findings with what the current navigational method of Wikipedia Provide, to show the potential that conceptual representation adds to the navigational and representation capabilities of Wikipedia.

II. Related Work

The process of learning in a self-regulated manner is challenging as the learner can find himself overwhelmed with a flow of information in several directions and in many fields. The increased volume of knowledge, particularly when related to many fields makes the cognitive process more difficult to manage. As a result, the need for an adequate cognitive tool that allows users to easily construct their growing knowledge schema and navigate through large amount of available online resources become very demanding.

Knowledge based work demands ways of externally representing knowledge to complement the memory's functions. As a result, the idea of using concept maps in education was introduced. Conceptual mapping has a long history with education; early research in this field was conducted in [4]. They invented concept mapping technique to follow and understand changes in children's knowledge of science. Depending on the learning psychology of David Ausubel, that learning takes place by incorporating new concepts with existing concepts in the learner brain and generating relationship between them to make it easier for information to be memorized and retrieved when needed. This is known as "individual's cognitive structure". This idea was the basis for a lot of research especially in teaching, learning and assessment.

Empowering concepts mapping by incorporating it with the power of computers was the focus of several studies. For instance [5] suggested the use of advanced computer-based concept maps to help students in managing knowledge when they cope with the complexity of knowledge and Knowledge resources in many domains particularly in resource-based learning scenarios.

Experiments then were conducted to measure the effect of using Concept Maps as a knowledge construction tool. One of these experiments was conducted on a graduate online course [6]. In this course students were asked to construct a Concept Maps after reading four texts that are related in topic to one another. Students were asked to synthesize their understanding of the texts by representing graphically at least fifteen key knowledge objects. Overall observations show that students generally found the concept mapping activity useful but its full potential as a knowledge construction support tool was far from optimal. A similar experiment was conducted in [7], in which students were asked to study issues from their everyday life in a self-regulated way. Then the issues were organized and simulated by the cognitive approach with the use of computer-based conceptual mapping software.

The association of concepts to improve domain knowledge exploration is proposed in [11], in which a methodology that makes use of concepts association bank is introduced. This bank is used then to recommend new concepts within certain subject domain to knowledge explorers. The concepts association bank utilities the use of mind map to extracts concept associations from collective domain databases.

Domain knowledge exploration can also be supported through the enhancement of knowledge and knowledge resources themselves. With the availability of advanced computer-based concept-mapping tools, several researchers have discussed the potential of using digital concept maps to support spatial learning strategies and processes of individual knowledge management. In [13] the researchers not only showed that digital concept maps can support knowledge management and learning strategies, but he also showed that these tools can be used to represent content knowledge about a domain, as well as knowledge resources.

Representation and visualization of knowledge is another issue that faces self-regulated learners and knowledge explorers. One of the contributions of Concept Maps that they depend on the visual perspective of the receiver as the cognitive process becomes easier when concepts are plotted out in the form of map. However, this same property could result in a distortion as maps can get more and more complex due to the amount of information and concepts contained. For that reason visualization of concept maps and knowledge domain has become a challenge and a rich field for research.

A survey was conducted by [8] on visualization techniques of knowledge domains. The importance of this research that it doesn't only list previous work but it introduces bibliographic data set. This dataset includes articles from the citation analysis, bibliometrics, semantics, and visualization literatures. It applied different visualization techniques on the dataset and compared the results.

Also one of the interesting visualization of conceptual information space techniques is that discussed in [9]. It used the infolead technology to navigate a large number of web pages simultaneously in a 3D visual presentation. The research in this area is still open and evolving the main goal is how to make the interaction between the human and the machine more flexible and simulate the way the human brain works.

This is what led University of Wisconsin-Madison to adopt the project of CoMPASS (Concept Mapped Project-based Activity Scaffolding System) [1]. The CoMPASS is a hypertext system that uses two representations (concept maps and text) to enable multiple passes through the same material and to support inquiry and learning. The CoMPASS hypertext is used to help students generate ideas and learn about science concepts that will help them to solve their design challenges. Based on this system a study was made on middle school students [10]. The study showed that students who used the map version of the software their navigation was more focused and did better in the conceptual map and the essay test.

A. But how conceptual maps can be constructed?

Two main approaches have been suggested in [15] and [16] for producing conceptual maps. The first approach is a text-based approach [15]. As the name states text-based maps are generated closely related to certain piece of text. It is based on text-charting and Rhetorical Structure Theory (RST) and it's done in several stages. Starting with taking notes and summarizes of the text, Then a text charting is done on the text while considering the RST of the text to produce an initial map. The final stage results in the final concept map in which concepts and relationships between them is identified and plotted.

A method was suggested in [12] to facilitate the process of producing a text-based concept map; it is called Text-to-conceptual representation. The aim of this method is to facilitate the transformation of text-based information into a graphical arc-node conceptual representation. Traditionally the graphical conceptual representation was done directly while extracting concepts from a piece of text. So the conceptual map is being drawn while reading a piece of text. This required several re-correction and re-formulation and also makes it harder for users to trace back concepts in the map to the original text. Once the map is plotted there is no way to identify where each specific concept is extracted from. As the user depend only on his memory and brain to organized and extract the concepts without recording why and how he draw the map in this certain way.

Text-to- conceptual representation allowed for enriching the conceptual representation process. The conceptual outline extraction is done in a table of the following form.

TABLE I. TEXT TO CONCEPTUAL REPS CONVERSION – TEMPLATE [12].

Original text	*Conceptual Outline*	*Complementary Info*	*Media Assets*	*Supplementary Info*

- Original text: Sentence by sentence of the original text is placed in this column. The original text is placed here without change.
- Conceptual outline: In this section the conceptual outline is extracted and placed. It is extracted in the form of <Concept> predicate <Concept>.
- Complementary Info: This section contains any further information not expressed in the conceptual outline. This information is seen by the user as complementary and he can always go back to it for enhancing the understanding of the conceptual map.
- Media Assets: This column for any media (images, video) asset extracted from the text and adds to the understanding of it.
- Supplementary Info: any external information from sources other than the text in hand is added in this column.

As noticed from the above explanation of each column role in the creation of conceptual representation, the conceptual outline is recorded in this table. This recording allows the users to trace what they drew on the map back to the text. And also determine the origin of each concept from the text.

The other approach is domain-based approach; this approach is discussed in [16]. As our thoughts not always expressed in linguistic form; the domain-based approach is used in representing non-linguistic acquired knowledge. Although there is no specific definition for this approach, some main features that characterize this type of representation can be found in [16].

- It should cover the main attributes and features of certain domain of knowledge.
- It is not specifically related to certain cognitive schema.
- It is not specific to certain piece of text; it's a text free representation.
- The developed domain map should be of minimal representation and yet captures the main features of a domain.
- Graphical and visual icons can be used to describe the dynamics of certain domain of knowledge.
- The produced domain map should be incontestable to experts in the same domain.

The analysis of Wikipedia in this paper is done using text-based approach. However the resulting conceptual maps can be considered a domain map as explained below in the proposed method.

III. PROPOSED METHOD

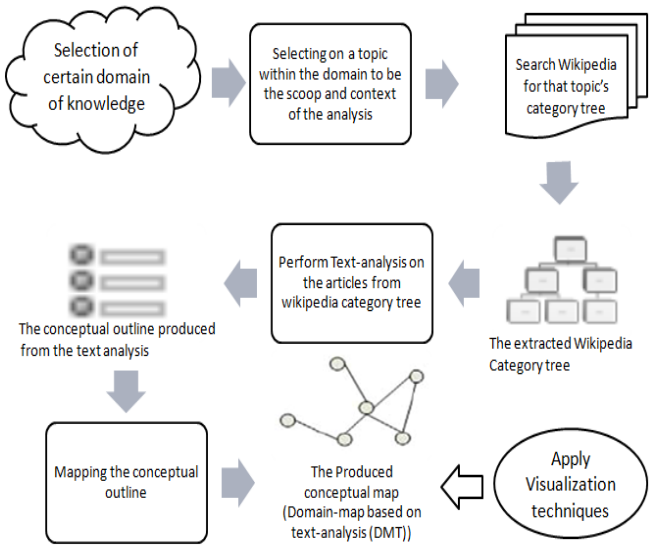


Fig. 2. Proposed method for analyzing Wikipedia domains

Although the use of conceptual maps to support self-regulated learners and to improve the navigation and exploration capabilities for learners has been addressed in several researches, the use of conceptual maps hasn't been applied before on this type of encyclopedic knowledge. Therefore it was necessarily to develop a new methodology to demonstrate how concept maps can add significantly to the representation and navigation process through encyclopedic domains of knowledge. The stages of the proposed methodology shown in Fig.2 the details of each stage are discussed below.

Selection of certain domain of knowledge

In this analysis the focus is scientific domains. Three domains were selected for the analysis, the criteria suggested for specifying domains is as follow: The first subject domain selected is a single subject single discipline, the second one is single subject multiple disciplines, the third one is a multiple subjects multiple disciplines.

Selecting a topic within a domain to be the scoop and context of the analysis

Within each domain a context should be stated clearly while navigating through Wikipedia. Such context makes it easier to judge on which concepts to include and which not in the analysis. This makes the domain of a manageable size. Without clear context so many concepts can be included which makes the analysis a daunting task. As when the chosen topic is subject group or principle subject, the navigated domain becomes so wide and harder to analyze and investigate.

Search Wikipedia for that topic's category tree

Each selected domain is searched in Wikipedia to extract the corresponding category tree of that domain. This extraction is simply done by finding the context related upper and lower categories of the selected topic along with any related pages within the categories' hierarchy. Categories or pages that are related to our context are included in the analysis that is done on the next phase. This relatedness is judged by the help of the experts in each domain.

Perform text analysis on the articles from Wikipedia category tree.

The text analysis performed on Wikipedia category tree articles is a text-based analysis done by Text-to- conceptual representation method mentioned above.

Mapping the conceptual outline

The produced conceptual outline from the previous stage is transformed into a conceptual map. The produced map is called Domain-map based on text-analysis (DMT).

Apply visualization techniques

It's widely known that when conceptual maps get larger in the number of displayed concepts and relationships, they began to look cluttered and more difficult to read. Another dimension of our research was to propose a method for the displaying of such large and nested maps. This leads us to the final stage of applying the visualization techniques. Visualization techniques are applied on the DMT map produced from the previous stage. These techniques are suggested to improve the visualization of the produced DMT map. In order to facilitate the exploration of such large and entangled maps an organization is of the map is proposed. The organization result in a 4 layers of the same map. These 4 layers end with the "Top cluster Level".

Layer4: is the bottom layer, this layer contains the DMT map as it is in its original form.

Layer3: in the 3rd layer, concepts are arranged according to their topical classifications. According to [18] topical classification is abstract structured spaces for arranging material spaces in which material or immaterial objects can get a location. Such immaterial objects can be concepts of certain domain or discipline or generally subjects of documents that are abstractly taken as information units. In this layer the relationships that join the concepts remain apparent.

Layer2: the resulting map from layer3 is displayed but the relationships are removed. This allows the user to concentrate on the concepts and see them without the interference of the relationships.

Layer1: the cluster layer, in this layer concepts are grouped into clusters with the header of each cluster is shown in the top of the class. Each cluster is about certain topic, the header is the main concept and the rest of the cluster is the concepts that are nesting from this main concept.

Domain Expert support

In all of the three explored domains experts supported the analysis in two main steps. Experts help essentially in reviewing the extracted Wikipedia category tree and the produced DMT map. In case of the category tree, evaluating the extracted map can cause the addition or removal of any concepts to the map according to their relatedness to the context in hand. In the case of the DMT map, they help in the comparison with the original category map.

It's been found according to the experts' provided support that the produced conceptual map can be considered a reasonable domain map. The analysis starts with text-based approach for the articles extracted from Wikipedia category trees and ends with producing a domain map that satisfies the domain map characteristics mentioned earlier. This solves an

important issue regarding the creation of domain maps that require an expert to create them. So without the need for experts a domain map can be obtained through the text created by Wikipedia users, the reason for seeking a domain map is that these maps captures the main essential features of a domain without relying on a certain piece of text. Therefore the resulting conceptual map is called Domain-map based on text-analysis (DMT).

In each case we are assisted by different expert. In the first case the experts are teacher assistants and the authors of this work who work in the faculty of computers and information systems. In the second case the expert support is given by a domain map created in [23]. This map was created as teaching aid in the course of "introductory to signal processing". In the final case the expert support is the category tree extracted from EduTechwiki. EduTechWiki is concerned with Educational Technology and related fields and hosted by TECFA - an educational technology research and teaching unit at University of Geneva. EdutechWiki is a resource for educational technology teaching and research. It also provide some (technical) tutorials for self-learners that or to be used in classes around the world [20]. EduTechWiki is organized in the same manner as Wikipedia, it include categories and sub-categories. As noted from the above experts are not necessarily human experts. The support can be given by a map or a reaserach conducted by some expert in certain domain. They act as experts by helping the analysist of this work to evaluate the results in comparison to artifacts in hands.

IV. RESULTS AND DISCUSSION

In this section the proposed method is applied on three cases, these three cases are discussed below.

A. Relational Database model case

1) Stage one: Selection of certain domain of knowledge

The domain selected in the first case is Database; there are two main reasons for choosing this domain. First this domain satisfies the predefined criteria in which database is considered a single subject, single principle domain

The other reason is that the authors of this work are professors and teacher assistant in the faculty of computers and information systems in the Information system department, and so considered experts in this area. We also consulted with another teacher assistant in the same department.

2) Stage two: Selecting on a topic within a domain to be the scoop and context of the analysis

For our analysis we selected relational database model [24] to be the context of this case.

3) Stage three: Search Wikipedia for that topic category tree

According to Wikipedia Relational Database model category tree is as shown in Fig.3.

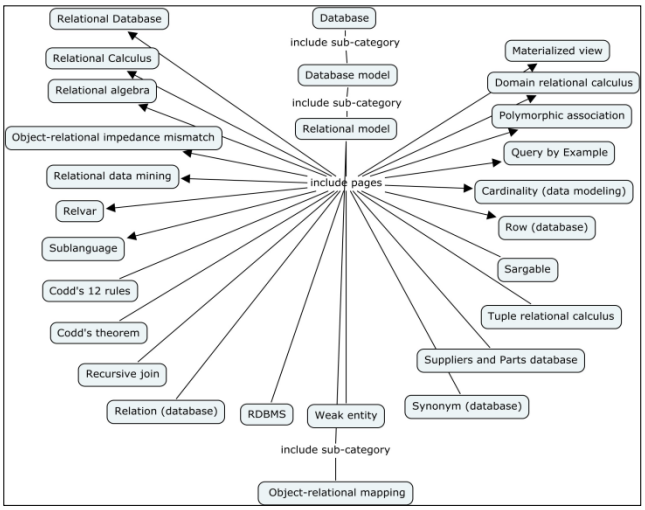
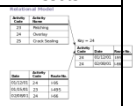


Fig. 3. Wikipedia Relational model Category tree

4) Stage four: Perform text analysis on the articles from the category tree

Below is sample of the text analysis done on the Relational Database model article.

TABLE 2. TEXT TO CONCEPTUAL REPRESENTATION OF RELATIONAL DATABASE MODEL ARTICLE

Relational model Article analysis			
Original text	***Conceptual Outline***	***Complementary Info***	***Media Assets***
The relational model for database management is a database model based on first-order predicate logic, first formulated and proposed in 1969 by Edgar F. Codd.[1][2]	<Relational model> is a <DB model>		
In the relational model of a database, all data is represented in terms of tuples, grouped into relations	<Relational model> represent data in <Tuples> grouped into <relations>		
A database organized in terms of the relational model is a relational database.	<relational DB> organized according to <Relational model>		
The purpose of the relational model is to provide a declarative method for specifying data and queries	<relational model> is a <declarative method> for <data>		

	and <queries>		
users directly state what information the database contains and what information they want from it, and let the database management system software take care of describing data structures for storing the data and retrieval procedures for answering queries	<DBMS> describe <Data Structure > for <sorting> and <retrieval > of <data>	Users use DBMS to state what should be contained in the Database and what is the type of information they want to retrieve from it	
Most implementations of the relational model use the SQL data definition and query language	<relational model> implemented by <Sql DD> and <Query Language >		
However, SQL databases, including DB2, deviate from the relational model in many details; Codd fiercely argued against deviations that compromise the original principles		Another type of Sql Databases is DB2 that is considered a deviation from the relational model	

5) Stage Five: Mapping the conceptual outline

This process includes **transforming** the conceptual outline from the previous stage into a conceptual representation. The produced map from this stage is showed below for the relational model case, only a snap shot is shown below due to the size of the map.

Layer 4:

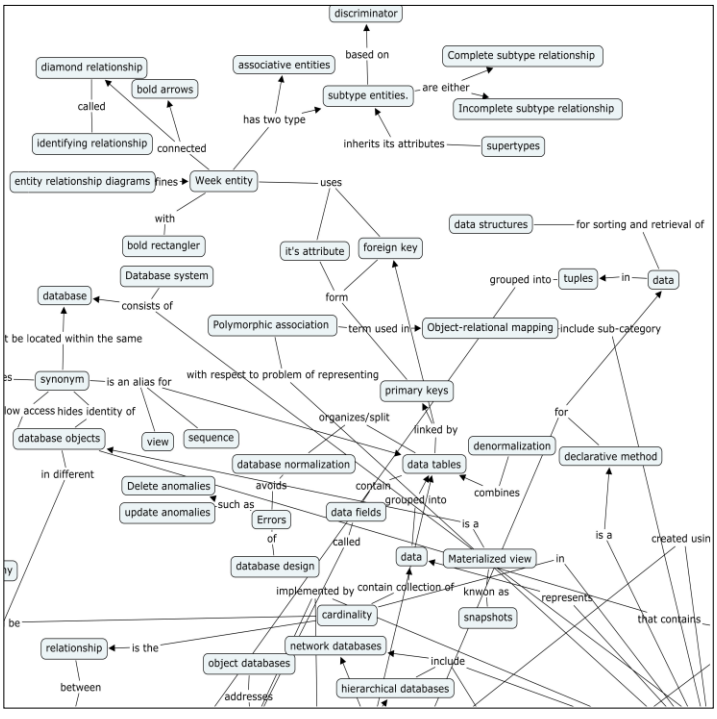


Fig. 4. Layer4 map, Conceptual representation of Relational model case (DMT map)

I. *Stage Six: Apply visualization techniques*

Layer 3:

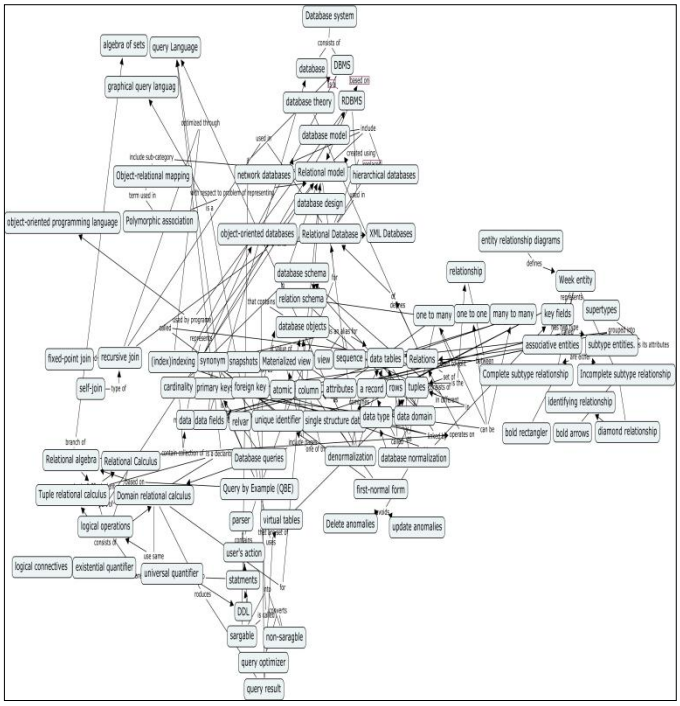


Fig. 5. Layer3 map, DMT map arranged according to concepts topical classifications

Layer 2:

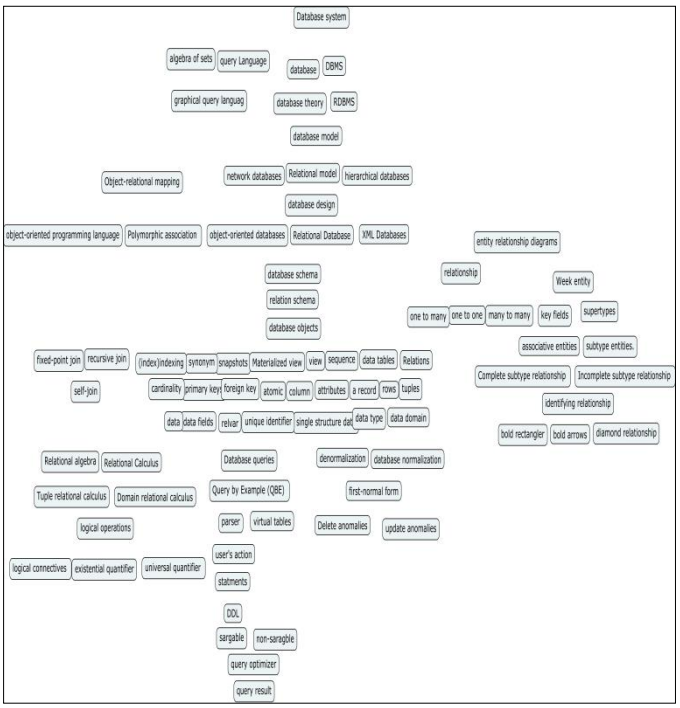


Fig. 6. Layer2 map, DMT map arranged according to concepts object classifications without relationships

Layer 1:

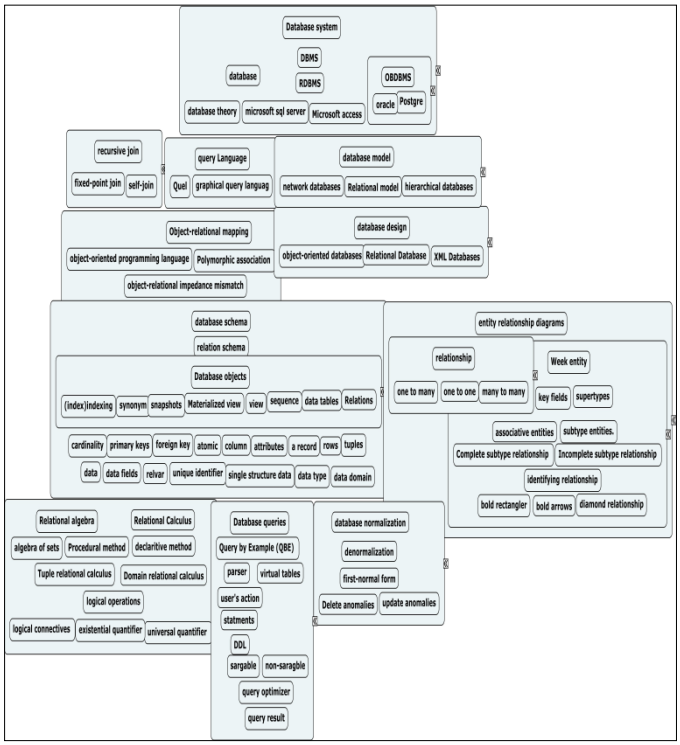


Fig. 7. Layer1 map, DMT map arranged into clusters of topics

Domain analysis:
Wikipedia category Tree:

Wikipedia Categories are arranged in a hierarchy which could lead to misleading information about what is related to what. For ex: in Fig.3, topic "RDBMS" is placed under "Relational model" as a subsidiary of it, while it's supposed to go under "DBMS" which is a software to implement and maintain Database. Note that "DBMS" along with "Database" form "Database system" which is in a higher level than "Relational model" in the Database hierarchy tree.

- An example of the misplacing of concepts under different categories. For ex: the topic "Week entity" exists while the "Entity Relationship diagram" that describes week entities and other types of entities isn't included.
- Some entries are placed under different contexts as in mentioning "Relational Algebra" and "Relational calculus"; these two topics form a theoretical foundation for query languages. A context where they are mentioned is usually when we are talking about Query Languages and the power of using them with "Relational Database". Other topics take us to different contexts like "Week entities" topic that is concerned with the establishment of a database schema.
- Fig.3 contains the topic "Relational Data mining", this topic describes a data mining technique applied on relational database; it must be place under "Data mining" category for those interested in data mining and its techniques. This is evidence that there is no clear criterion for including topics within categories.

The Resulting DMT maps: The maps produced by the proposed method shows significantly the difference between Wikipedia representation of domain knowledge and between the resulting conceptual representations. Even though our representation began by the analysis for articles from Wikipedia category tree and not from other sources, the results show wider views and interconnectivity between concepts within a domain.

The resulting cluster map in Fig.7 shows a variety of topics inside Relational model domain. For instance, Entity relationship diagram (ERD) that contains different types of entities including normal entities and week entities and the type of relationships between these entities are described clearly. Such information didn't appear in the original Category tree map, only mentioning week entity article as a page within the relational model category. Note that ERD is an essential concept in the creation and development of any relational database model. Fig.7 also contains the different types of Database objects, database models, query languages and concepts of normalization and denormalization. The different operations allowed by relational algebra and relational calculus are also placed in the map in Fig.7.

The Role of domain expert support: proving how well domain knowledge is represented is judged by the domain expert. Experts also help analyzing the included topics on both the category tree and the resulting DMT map. As explained above the resulting conceptual DMT maps provide a much more comprehensive view for the domain in hand. Also the

relationships between the concepts are much easier to locate and understand.

In this case the evaluation process of Wikipedia category tree resulted in exluding some article like Relational data mining, Relver and other unrelated concepts.

B. Simple Harmonic motion case

The scope of the second case is Simple Harmonic Motion (SHM). SHM is a single subject, multiple disciplines. The concept of SHM is related to several areas; it's used in Physics, in mechanics, in engineering. It's also involved in so many dynamical systems for ex: "Pendulums", "mass-spring". And it's related to other motions like Circular motion. In the analysis of this case the expert support is a SHM map (expert map [23]) Fig.8.

after searching for SHM in Wikipedia, the extracted category tree is drawn as in Fig.9.

Then guided by the expert map the first category tree is expanded in which we searched Wikipedia for topics that exist in the expert map and add these articles to the first extracted category tree Fig.10. After applying all the six stages of our proposed method the resulting DMT map in the form of clusters (Layer 1) is then drawn in Fig.11.

Domain analysis:

Wikipedia category tree: For any user investigating SHM topic, the first category tree Fig.9 is what he will find in Wikipedia. This shows the difference between a map truly describing the domain and what users who don't have enough knowledge will find.

In the expanded category tree in Fig.11 it included important articles like, oscillation, waveform, amplitude, damping, sine wave, circular motion, spring pendulum and pendulum. This shows the deficiency of navigating Wikipedia without previous knowledge in a domain. Without the guide map such articles that are according to the guide map form a very important aspects of SHM topic wouldn't been reached. For users who are reading for the first time about SHM, these topics will look like any other topic on mechanics category that contains over 200 categories, sub-categories and hundreds of pages under these categories.

- Wikipedia categorization trees have ad hoc organization of topics insufficient to fully and satisfactory characterize a domain. For ex: topic like "Oscillations" is categorized in the same level with classical mechanics, and treated as a topic of higher level than SHM. While in fact oscillation is one of the behavioral features that describe SHM it is part of it.
- Concepts that are supposed to be linked to each other appear here as separate entities not related in the Wikipedia category tree. For ex: dynamical systems that contains systems such as "Mass-spring", "Pendulums", and "spring pendulums".
- When searching for concepts related to engineering and SHM, the only results are instruments related to the control part of engineering and not the motion.Even when searching for the two concepts together "Pendulum+ dynamical systems" although the pendulum concept exist in "dynamical system" it didn't show in the results and didn't reflect in the category tree of them together.
- Also the steady state and transient states of system behavioral features exist under different categories far from SHM

The Resulting DMT maps: This classification shows the domain of SHM in a more detailed and clearer representation. It specifies the natural phenomenon's affecting any system under motion like (frequency, gravity, oscillation), different types of dynamical systems that are subject to motion like (springs, pendulums, oscillators), different types of periodic motion like (cicular motion, SHM, molecular motion), damping and their types that affects the motion of dynamical systems and the states these systems can be under.

The resulting conceptual maps don't simply list the concepts, but they also arrange them and specify the type or relationships between them. This shows the importance of the conceptual representation in detecting concepts that are highly related to the domain and yet are not included or included in ambiguous way.

The role of domain expert support: The first category map in Fig.9 shows that normal users will only find articles about simple harmonic motion, classical mechanics and pendulums. It is almost impossible for users with no knowledge about the domain to find the rest of the topics that acquired by using the expert map in Fig.8.

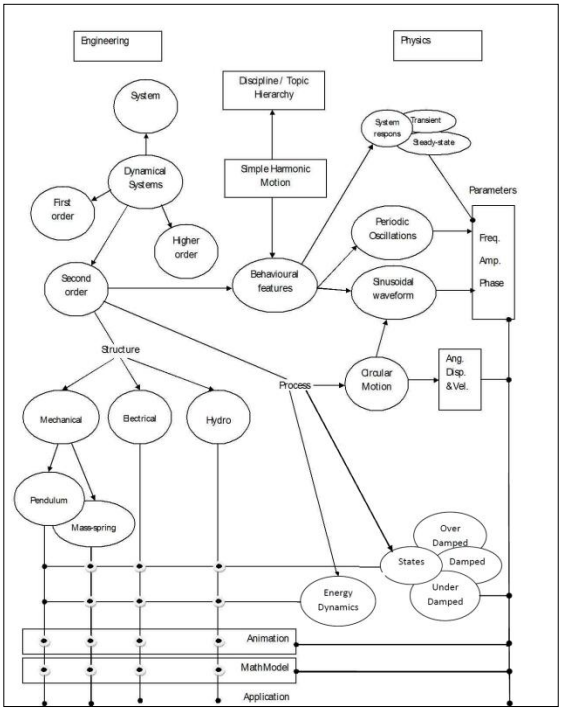


Fig. 8. SHM guide map, created by domain expert

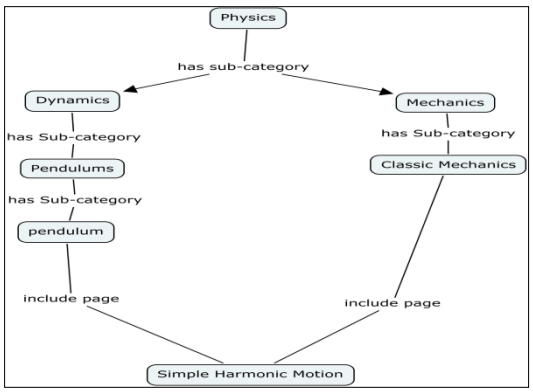


Fig. 9. Wikipedia SHM Category tree

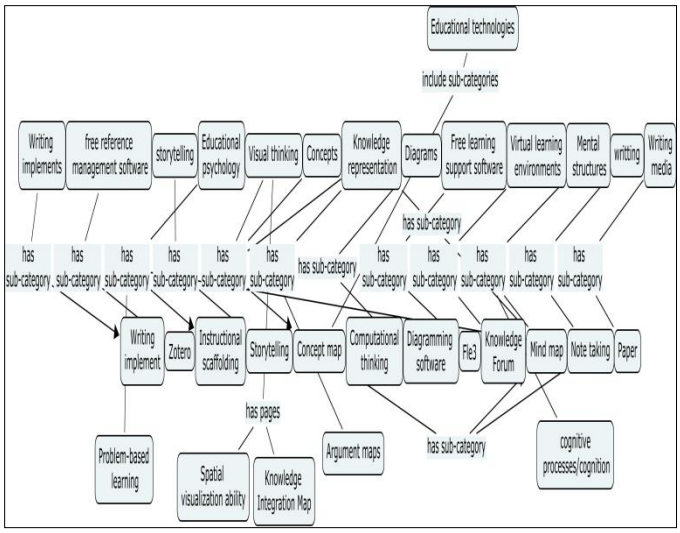


Fig. 10. SHM Category based on SHM Guide map in whole of Wikipedia

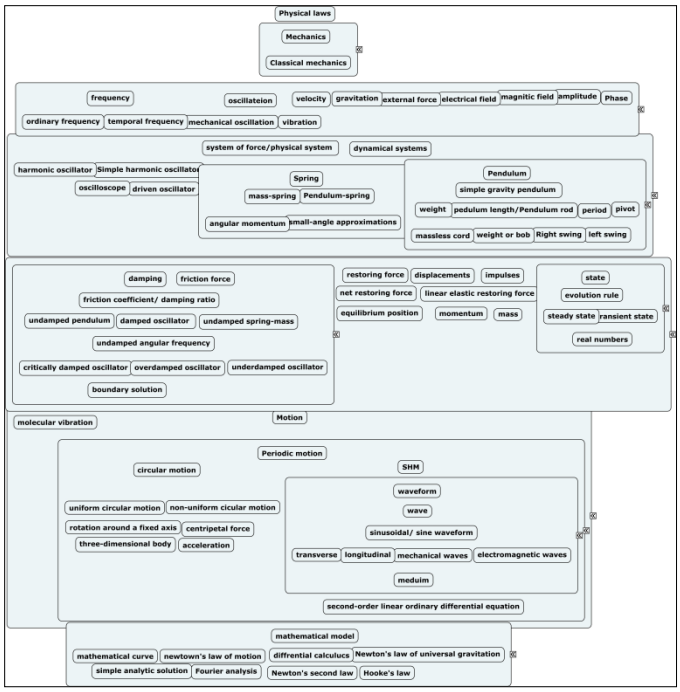


Fig. 11. Layer1 map, DMT map arranged into clusters of topics

C. Cognitive support tools

The third case is Cognitive support topic. Following same selection criteria as done on the previous 2 cases, this case is a multiple subjects, multiple disciplines. The selected topic within cognitive support domain is cognitive support tools.

When searching for this topic 'cognitive tools' in Wikipedia, this topic wasn't found in Wikipedia, not even the topic 'cognitive support'. In this case, the expert support that will guide the search in Wikipedia is EduTechWiki [19]. Guided by category tree of EdutectWiki map in Fig.12 a search is conducted in Wikipedia for finding topics that matches those in Fig.12. The resulting Wikipedia category tree for cognitive support tools is drawn in Fig.13. Although the main concept 'cognitive tool' doesn't exist in Wikipedia, some of the articles in EduTechWiki category tree exist in Wikipedia but under different categories. After applying all the six stages of the proposed method the resulting DMT map in the form of clusters (Layer 1) is drawn in Fig.14.

Domain analysis:

Wikipedia category tree: Same issues regarding the organization of Wikipedia categories can be seen in this case too. Fig.19 shows that concepts that are related to cognitive tools appear here to be scattered all over Wikipedia under different categories. Only few articles seem to have some sort of a connection as Note taking, Mind map and Concept map, Instructional scaffolding and Note taking.

The Resulting DMT maps: Although the concepts of Cognitive tools and cognitive support are not explicitly mentioned in the articles extracted from Wikipedia but after clustering these concepts start to surface. Cluster like Cognitive skills include concepts that describes cognition and skills acquired by the brain. Also the Cognitive support cluster shows type of support that can be given to learners in order to support the cognitive skills. It also shows the type of mental processes and functions that human brain is capable of.

The cognitive tools cluster is a group of the cognitive tools that can support the cognitive process performed by the brain. Some of these tools use visualization techniques like this approach in hand 'conceptual map', so another cluster appears that include some of the visualization features used in the cognitive process.

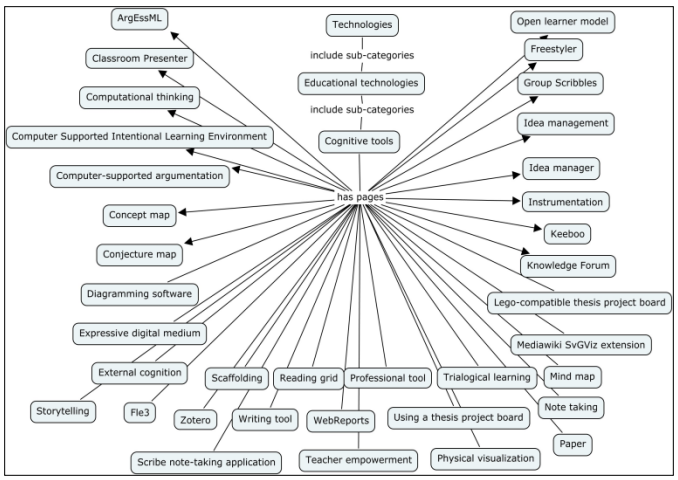


Fig. 12. EduTechWiki cognitive tools category tree

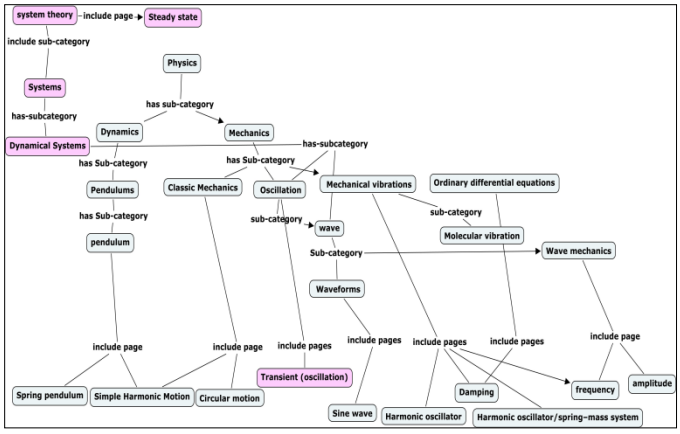


Fig. 13. Wikipedia category tree of Cognitive support tools

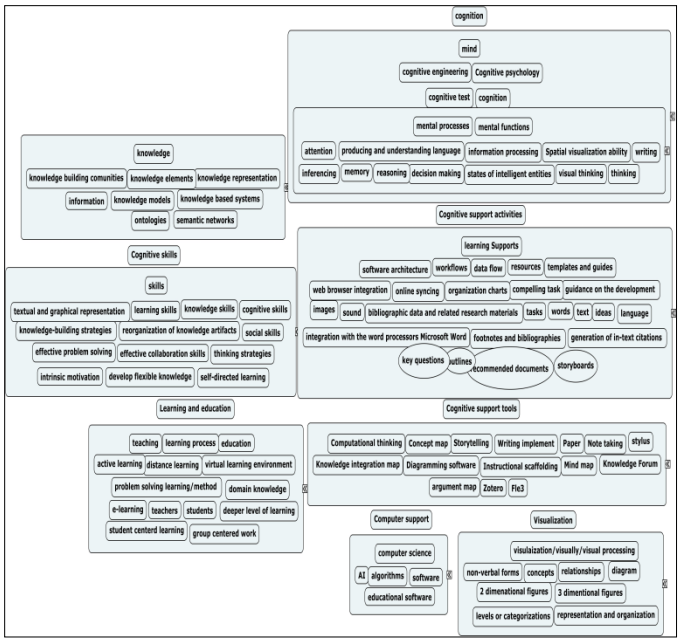


Fig. 14. Layer1 map, DMT map arranged into clusters of topics

Also a new tool like argument map and Knowledge integration map weren't mentioned in the EduTechwiki category tree. But it appeared in the conceptual analysis of the articles as tools that support the cognitive process.

The Role of domain Expert: This topic in particular is very important in this analysis, as the topic in hand doesn't exist in Wikipedia as a topic. Obviously this is a dead end for any knowledge explorer who would simply stop searching. Although cognitive tools are not explicitly stated in Wikipedia but several tools are mentioned under different categories. Guided by the Edutechwiki category tree the resulting DMT map shows the topic in a clear way with concepts didn't appear in the original category tree.

V. CONCLUSION AND FUTURE WORK

In this paper a new methodology is introduced for representing and navigating online encyclopedic knowledge using conceptual representation. This methodology combined the two approaches of creating conceptual maps, starting with text-analysis of Wikipedia articles to end up with a comprehensive domain map of the analyzed domains of knowledge. The support of experts is essential in evaluating the produced results and they specially helped in comparing between the initial Wikipedia category tree map and the resulting DMT map.

The resulting DMT maps showed unarguable results in describing domains of knowledge in a way that Wikipedia couldn't achieve before. This leads to better understanding of a domain and a better navigation across different concepts within the same domain. Users initially get lost by the amount of articles within each category those are not necessarily related to the domain and limited by the navigational approaches offered by Wikipedia which are hyperlinks and categories. By representing a domain in a conceptual map, the whole domain is plotted in a single map that a user can zoom in and out and choose which concepts to view and which not.

This research can be extended in the future by placing the produced conceptual DMT maps as a front end for Wikipedia users, these maps won't replace the original text but they will act as a navigational gate for the domain underlying these maps. Users can choose at any point to navigate further down and read the details that exist in each article. Users will also be able to add concepts to the map and edit it the way Wikipedia offers now, so users themselves will be able to see the domain map and decide what is missing or what can be added.

The focus of this paper was on scientific domains, this research can be experimented on other domains to measure how effective the proposed method can be with other domains. The proposed method can also be used to classify and determine the nature of wikipedia different domains.

REFERENCES

[1] CoMPASS, "Concept mapped project-based activity scaffolding system", http://www.compassproject.net/info/index.html, Wisconsin Center for Education Research. 2008. Web. 24 May 2011,

[2] J. A. Hampton, H. E. Moss, "Concepts and meaning: Introduction to the special issue on conceptual representation". *Language and Cognitive Processes*, 18(5-6), pp. 505-512, 2003

[3] A. El Sayed," Conceptual Representations a content modeling perspective", unpublished, 2011.

[4] J.D. Novak, D.B .Gowin, "Learning how to learn", Cambridge: Cambridge University Press, 1984
[5] S.O.Tergan, "Concept Maps for Managing Individual Knowledge", Proceeding of the first joint meeting of theEARLI, pp. 229- 238, 2004
[6] J.Basque, B.Pudelko, "Using a concept mapping software as a knowledge construction tool in a graduate online course", Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications,
pp. 2268-2274, 2003
[7] W. Gräber, A. Neumann, S.O. Tergan , "Mind Mapping for Fostering Self-Regulated Resource-Based Learning in Science Classes", International Workshop on Visual Artifacts for the Organization of Information and Knowledge, May 13th - 14th, 2004.
[8] K.Börner, C.Chen, K. W. Boyack, , "Visualizing knowledge domains", Annual Review of Information Science and Technology , Vol. 37 , Nr. 1 Wiley, S. 179-255, 2003.
[9] K.Toru , K.Atsushi , T-U.Kaku , M.Tetsuyuki , "A Proposed Net-space Service Using InfoLead Cruising Navigation Technology", IEEE Computer Society. pp. 31-32, 2002.
[10] S.Puntambekar, A.Stylianou, R.Hübscher," Improving navigation and learning in hypertext environments with navigable concept maps". Human Computer Interaction, 18 (4), pp.395-426, 2003.
[11] C-C. Liu, S-H.F.Chiang, C-Y.Chou, S.Y. Chen, "Knowledge exploration with concept association techniques", Online Information Review, Vol. 34 Iss: 5, pp.786 - 805, 2010.
[12] A.Elsayed. "A text analysis methodology for text-based conceptual modeling of knowledge domains", Technical Report, Learning Systems International, Leeds, UK, 2012.
[13] S.O.Tergan, "The Use of Digital Concept Maps as Cognitive Tools for Managing Knowledge and Knowledge Resources", AAAI, symposia, SS-05-06, pp.73-76, 2005.
[14] Wikipedia,"Categories, lists, and navigation templates", http://en.wikipedia.org/wiki/Wikipedia:Categories,_lists,_and_navigation_templates, 2001.web. 6 Feb. 2013.
[15] H.Roger," Conceptual Representations and the Dimensions of Textual Content" , M3C International Workshop taking place at (ICALT), University of Bolton, UK, July 2009
[16] A.Elsayed ,"Relations for Graphical Conceptual Representations (Presented Conference Paper style)", M3C International Workshop taking place at (ICALT), University of Bolton, UK, July 2009
[17] Higher Education Statistics Agency," JACS3 Classification", http://www.hesa.ac.uk/content/view/1787/281/, 1993. Web. 5 Sept. 2012.
[18] De. Antonella , M. Dario, M.Alberto, "Subject Classifications in the Scientific and Overall Digital World", High Energy Physics Libraries Webzine, Draft, 2001.
[19] EduTechWiki,"Cognitive tools", http://edutechwiki.unige.ch/en/Cognitive_tool
, 2001. Web.5 Mar. 2013.
[20] EduTechWiki, "Main Page", http://edutechwiki.unige.ch/en/Main_Page
, 2001. Web.5 Mar. 2013.
[21] J.Novak, C.J. Alberto, "The Theory Underlying Concept Maps and How to Construct and Use Them ", Technical Report IHMC CmapTools 2006-01.
[22] Wikipedia, "Wikipedia", http://en.wikipedia.org/wiki/Wikipedia
, 2001.web. 6 May 2013,
[23] A.Elsayed, "A domain representation approach for the use of encyclopaedic knowledge", Technical Note, M3C Lab, University of Bolton, UK, 2009.
[24] School of Mathematical & Computer Sciences, "F21DF1 Database and Information Systems", http://www.macs.hw.ac.uk/~trinder/DbInfSystems/l4RelModel2up.pdf, 2010.web. Nov.2010.