

## ANALYTICAL TOOLS FOR VISUALISATION OF INTERACTIONS IN ONLINE E-LEARNING ACTIVITIES ON LMS AND SEMANTIC SIMILARITY MEASURES ON TEXT

<sup>1</sup> YAHYA AL-ASHMOERY , <sup>2</sup> ROCHDI MESSOUSSI , <sup>3</sup> RAJA TOUAHNI

<sup>123</sup> Laboratory Systems of Telecommunications and Engineering of the Decision (LSTED),  
Faculty of Sciences, University Ibn Tofail , Kenitra , Morocco

E-mail: <sup>1</sup> [yahya\\_new@hotmail.com](mailto:yahya_new@hotmail.com), <sup>2</sup> [messoussi@gmail.com](mailto:messoussi@gmail.com), <sup>3</sup> [rtouahni@hotmail.com](mailto:rtouahni@hotmail.com)

### ABSTRACT

This paper aims to build an analytical tools system for Learning Management Systems (LMS) that will aid and support teachers and administrators to understand and analyse interaction patterns and knowledge constructions of the participants involved in ongoing online interactions. Most LMS commercial or open source software does not include any tools for comprehensive access tracking and log analysis capabilities. They also lack the support for many aspects specific to evaluating participation level , analysing interactions , visually representing ongoing interactions , analysing of the content of messages Semantic .It is difficult and time consuming for the teachers and the educationalists to determine the number of participants, non-participants and lurkers in ongoing discussion. Our analytical tools take a novel approach of using Web log data generated by Learning management systems (LMS) to help instructors become aware of what is happening in distance learning classes. The measure of semantic similarity between texts used in our system, to find out the degree of relation between the messages. It gives the value of relevance in numerical format. The semantic similarity of text utilises WordNet as an underlying ontology for calculating semantic similarity. WordNet is a lexical database for English Language. It is organized as a network which consists of concepts or terms called Synsets (list of synonyms terms) and the relationship between them. There are different types of relationship existing in WordNet such as; is-a, part-of, synonym and antonym. The proposed methods can be used in a variety of applications that involve text knowledge representation and discovery.

**Keywords:** *Learning Analytics ,Virtual Learning Environments, Log Analysis ,Semantic Similarity, Wordnet*

### 1. INTRODUCTION

Computer Mediated Communication (CMC) refers to the area of human-computer interaction (HCI) which, in the field of education promotes collaborative online learning. Online learning is a learning experience where the learning content is delivered through technology either through online course materials or online interactions. The rapid growth of online learning has created a drive to study and analyse online participation, which fosters cross-cultural learning communities.

The dynamic nature of online interaction, where new topics are often introduced and new threads and posts are formed to perpetuate the online activity, which is difficult to assess. To deliver a clear understanding of the nature of online

discussions, there is a strong need to study the qualitative and quantitative nature of online interactions. The higher the number of interactions, the more clear the analysis of those interaction patterns become. It seems reasonable to accept that participation and interaction are the enablers of learning and knowledge construction in Virtual Learning Environments (VLE). Communication on a CMC has to be more interactive in order to assess the learning capabilities of an individual. However, sometimes this interaction could be just the formation of a social network. According to Sing and Khine (2006)[1], successful co-construction of knowledge requires active and broad participation, which requires substantially more posted messages. They also state that this data can be used as important background information, to prompt further analysis to assess the quality of interaction.

One possible way in which LMS log data can be used is to visualize learning activity data via a more human-friendly graphical interface. This interface is updated in real-time to reflect all activities, and is integrated into the original LMS. Using this graphical interface, instructors can gain an understanding about the activities of the whole class or can focus on a particular student. Students can use the interface to understand how their progress compares to the whole class. Our hypothesis is that the availability of this graphical interface can provide important insight into how students are accessing online course materials, and this insight will be useful to both the instructor and students.

An important aspect of online teaching and learning is the monitoring of student progress and the utilization of tools on online courses. Educational research shows that monitoring the students' learning is an essential component of high quality education. Using log files of learning management systems can help to determine who has been active in the course, what they did, and when they did it. (Romero, Ventura & Garcia, 2007)[2]. Feedback about the status and the history of the activities in online-courses can be useful to teachers, students, study program managers and administrators. As this helps enable them gain a better understanding of whether the courses have provided a comprehensive learning environment (availability and use of discussion forums, etc.) It also shows to which extent the best practices on online learning are implemented, as students provide regular responses which are visible to teachers to monitor and counter. Learning Management Systems provide some reporting tools, that aim to monitor students' and tool's usage, but these are rarely used mainly because it is difficult to interpret and exploit them. The obstacles to interpretation and exploitation are the following:

Data are not aggregated following a didactical perspective

- Certain types of usage data are not logged .
- The data that are logged may seem incomplete.
- Users are afraid that they could draw unsound inferences from some of the data.

In the attempt of overcoming these difficulties, new reporting functions of LMS have been added, for instance Moodle now provides reporting tools which enable teachers to evaluate the activity patterns of individual students. Moreover, in the last few years researchers have begun to investigate various data mining methods which allow exploring, visualizing, interpreting and analysing

eLearning data thus helping teachers obtain a better understanding and improving their eLearning practice (Romero et al. 2010) (Mazza & Botturi, 2007)[2] [3].

Our log analysis tools system has been developed for the LMS (Moodle,2014), it is superior to the original Moodle log facility in several aspects:

1. Integrated into LMS as a Plugin.
2. Materialization of an intelligent agent to assist teachers in a collaborative work environment.
3. Building a smart student profile.
4. Advising teachers to help them manage distance classes is required to notify them about the students' knowledge status, including both individual and group bases.
5. Interaction visualization in web-based learning using complex graphs.
6. It provides aggregated and meaningful statistical reports.
7. It visualizes the results, which makes comparisons between multiple students much easier.
8. Displaying the activities that a student has performed as well as identifies the materials that have not yet been used by a student.
9. Implementing a visualization tool that will enhance the means by which asynchronous online communications in discussion forums can be analysed.
10. Choose evaluation procedures that can be integrated with the visualization tool that will allow comparisons between groups participating in the online discussion environments.
11. Observe any sort of user, including lurkers, and track any activities and form of communications on both server side and client side.
12. Semantic similarity measure using information content approach with depth for similarity.
13. Sentence similarity based on semantic nets and corpus statistics.
14. Text analysis and content analysis of text messages.
15. Finding the relevance between two messages( the relationship between messages)
16. Searching Engine-Based approach to measure semantic similarity between words.

The next section of this paper gives a brief literature review of an existing web-based learning analytic tools that interoperate with LMS. I have then presented our analytic tools system, which I have indicated by applying some examples. Following

this is semantic similarity measures and WordNet as we focus on several algorithmic approaches for calculating semantic similarity measures on texts. As well as implementations of the Semantic similarity measure tools (system lifecycle). A conclusion would be reached based on the use of our analytics tools, as well as future plans would be considered regarding the tools' usage and functionality.

## 2. COMPARISON OF THE AVAILABLE VISUALISATION TOOLS

One of the important challenges of the concept of visualisation is to present as much information as possible in a given display area without overloading either the display, or the cognitive capability of the person viewing. Having reviewed thoroughly the existing tools, we argue that this challenge has not yet been met with any of those developed so far. Figure 1 below provides an overview of these systems, and their characteristics.

Tools Name	Our Tool	MOCLog	TrAVis	GISMO	CourseViz	Authorline Visualisation	DIAS	MTRDS
1 Text Analysis	✓							
2 Relevance between two messages	✓							
3 Search Semantic on Message	✓							
4 Method of Analysis	✓	✓	✓	✓	✓	✓	✓	✓
5 Participation Level	✓	✓	✓	✓	✓	✓	✓	✓
6 Social Interaction	✓	✓	✓	✓	✓	✓	✓	✓
7 Cognitive Presence / Critical Thinking	✓	✓		✓	✓			
8 Lurkers	✓		✓					
9 Integrated into LMS as a Plugin	✓	✓		✓				

Figure 1: A Comparison Of The Available Visualisation Tools Including Our Tools System

All of these tools are different, and have their own advantages and disadvantages. Almost all of them will generate visualizations of a single discussion thread or a single participant interaction. However, when made to show chains of messages, they become unclear and cluttered. Most visualization tools depict crosslinks (cluttering) when describing the message chains. There is an overlap in the nodes, which represent the participants, or a series of arcs which denote the flow of messages, crossing one over the other. Crosslinks and cluttering usually occur when the data to represent the ongoing interaction is vast. A further issue with visualisation tools is their inconsistency in terms of the criteria or 'indicators' they use in order to produce visual representations of data. Because of these issues, it is important to clarify here where such 'indicators' come from, and which were

prioritized in the development of our system Tools. May et al. (2007)[4] states that the traces of the learners' activities are potentially a source of information that reveals not only their activities but also various results of those activities. Donath, Karahalios, and Viegas (1999)[6] showed that the information derived from these interactions helps learners and moderators review their own behavior and that of others. By analyzing traces of activities in collaborative learning environments, moderators could evaluate the social and cognitive behaviors of the learners. Evidence of a similar approach can be found in research of Mazza & Dimitrova (2004)[7], where teachers/moderators could view the three main aspects of online learners' behavior: social, behavioral and cognitive.

Other very commonly used indicators in the visual representation of any CMC are type of analysis, participation level, and presence of social interaction. Not many visual tools give importance to lurkers, participation details, time duration and cognitive presence. Other important indicators are not indicated in any of the visual tools except our system tool:

1. Text Analysis, content analysis of text messages.
2. finding the relevance between two messages, The relationship between messages
3. Search Engine-Based approach to measure semantic similarity between words

Some of the most available visualisation tools:

1)MOCLog (Monitoring Online Courses with log data) – 2012,[8]. MOCLog is a sum of tools that are used for the analysis and presentation of log data within Moodle. The development of the tool was based on GISMO[9]. Thus, some of GISMO's main components for the production of statistical reports for educators and students have been re-used. MOCLog attempts the analysis of interactions occurring in an online course so as to achieve better analysis of both the products and the educational process itself. It distinguishes among users according to their role within the system (course manager – teacher – student) and presents different statistical reports tailored to these roles. Therefore, the system's users have access to summative reports of interactions related to actions on educational resources and educational tools within specific subjects, such as quizzes, assignments, and wikis. This tool also includes some disadvantages: there is no representation of the conversation flow. It also does not include indicators about Text Analysis

(content analysis of text messages) and Lurkers information.

2) TrAVis (Tracking Data Analysis and Visualization Tools), May, M., George (2007) [10], a Web-based system that assists the students in visualizing their communication activities in distance learning situations TrAVis is objectively designed and implemented for both teachers and students. In particular, TrAVis is considered a “reflective tool” that helps students analyze and evaluate their own activities in relation to those of others. May et al. (2007) (2008)[4],[5] has visualised the user communication activities through a web-based Tracking System named TrAVis. There are some disadvantages on this tool; it does not visualize the message flow. It is hard to find the communication pattern from the visual representation. It is also impossible to deduce that the time spent ‘reading’ a message actually relates to ‘reading’. Instead, it could be a page left open.

3) GISMO (Graphical Interactive Student monitoring tool), developed by Mazza & Botturi (2007)[9]. Graphical interactive student monitoring and tracking system tool that extracts tracking data from an on-line course maintained with a Course Management System, and generates graphical representations that can be explored by course instructors to examine various aspects of distance students. It fails to represent many of the important indicators like participant details and post details. This representation also lacks the ability to display the conversation flow and the interaction path. There is no way of identifying if the contribution is relevant, useful, and appropriate.

4) CourseViz –The use of information visualisation to facilitate instructors in distance learning lead to the development of CourseViz. (Mazza & Dimitrova, 2007)[10], a system that takes a novel approach of using Web log data generated by course management systems (CMSs) to help instructors become aware of what is happening in distance learning classes. Specifically, techniques from information visualisation are used to graphically render complex, multidimensional student tracking data. Several graphical representations are generated to help distance learning instructors get a better understanding of social, behavioural, and cognitive aspects related to learners. This representation lacks the ability to visualize the conversation flow. It is difficult to view the 3rd dimension of the graph. The

behavioral graph gives too much information making the graph look cumbersome.

5) Authorline Visualization Howard et al. (2007)[11], in their work describe visual strategies that illustrate patterns of contributions that they propose are a proxy for indicators of social roles (patterned characteristics of communication). They propose that the key behavioral signature of an answer person (individuals whose dominant behavior is to respond to questions posed by other users) is seen in the tendency to reply to discussion threads initiated by others, e.g. the presence of blue circles and the absence of red circles indicating thread initiation.

Although the tool helped to view the dynamics of the interactions at a glance, screen cluttering becomes a problem if the number of authors and the number of posts increase. There would be an overlap of different sized-circles. The inability to read individual messages within the visualization was one of their issues.

6) DIAS (Discussion Interaction Analysis System), 2006 [12] is another software application for monitoring and analyzing the interaction developed by Bratitsis & Dimitracopoulou. The indicators in DIAS along with appropriate interpretation schema helps in drawing various conclusions related to the interactions among the learners. The main purpose of DIAS is to provide a range of indicators for analysis. A lot of quantitative analysis is done, but there is no indicator indicating the analysis of learning. There is no representation of the conversation flow.

7) MTRDS (Mapping Temporal Relations of Discussions Software) Gibbs (2006)[13] produced his software MTRDS to assist analysis of the temporal aspects of online discussions. He represented the interaction in a way that could provide greater awareness of the communication dynamics. MTRDS generates a visual representation of discussions based on time and date. Only one week’s (seven days) worth of messages display at a time. This number can be increased but the clarity of the map may be affected.

### 3. IMPLEMENTATION OF OUR ANALYTICAL TOOLS SYSTEM

The proposed strategy is composed of four major steps in the following:

- 1) Statistical reports for users activity

- 2) Visual Groups Tools
- 3) Monitoring Students Tools
- 4) Semantic Similarity Measures Tools



Figure 2: Main Menu on our analytical Tools system

#### 4. THE ANALYTICAL TOOLS SYSTEM INDICATORS

Some of the indicators in our analytical tools system in the following:

##### A. SUMMARY OF THE COURSE ACTIVITIES

This view of statistics is a list of the overall information that need to be considered about a particular course:

- How many students are enrolled on the course?
- How many students lurkers?
- How many messages (Read, Replied, Initial Thirds)
- How many documents are participated?
- Who are the students lurkers or not active?
- When does the discussion start and finish?

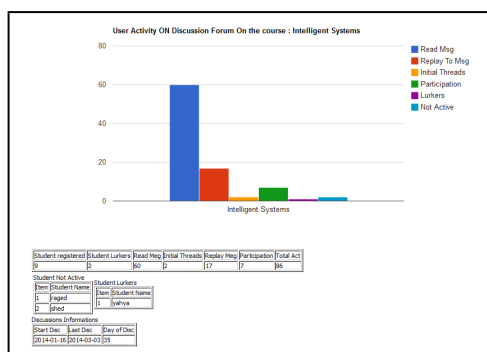


Figure 3: summary of the Course activities

##### B. USER ACTIVITIES

It provides a rich set of information about how often and how intensely students interacted with Moodle. Figure 4 shows the student activity

statistics for the course. The statistics show seven fields of data: Last access, total number of views, total number of sessions, total online time, number of viewed resources, number of initial threads

Fuzzy Logic   Building Profile   Advice Generation							
List of User Activity on the course : All							
ID	Name	Last Access	Total View	Sessions	viewed resources	Online time	Initial Threads
Average			60.71	11.14	4		
3	yahya	2014-03-03 12:41:29	38	21	9	23:29:28	1
4	omer	2014-02-27 23:36:08	36	10	6	09:52:22	2
5	hoda	2014-02-27 23:32:04	30	8	1	38:00:02	1
6	nazir	2014-03-03 16:46:19	52	10	3	29:07:24	0
7	ahmed	2014-02-28 16:30:37	69	9	2	18:36:30	2
8	safa	2014-02-27 23:30:37	65	12	5	06:56:50	2
9	abr	2014-02-28 23:43:45	55	8	2	18:54:36	1

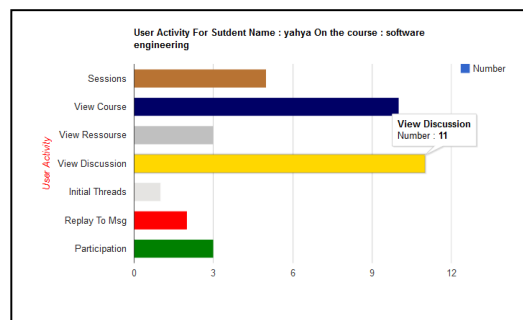
posted by the user, and the number of follow-up messages posted.

Figure 4: user activities

##### C. SUMMARY OF USER ACTIVITIES

The indicator that features the statistical data related to different activities of a user on a discussion forum. The main objective of such indicator is to provide an overview of the following activities:

- Reading messages.
- Viewing course materials or assignments that are posted in the forum.
- Posting new messages (or starting new discussion threads).



- Replying to messages.
  - Documents participated
- Figure 5: Data indicators for user activity

##### D. VISUALIZATION OF DISCUSSION

Figure 6 shows data on discussions through the course. In this, visual instructors can have an overview of all the discussions performed by students. For each student of the course it gives indications on the number of messages posted, the number of messages read and the number of threads started by the student in the discussions.





FIGURE 6: REPRESENTATION OF THE DISCUSSION

## E. MAPPING ASYNCHRONOUS DISCUSSIONS SYSTEM

This help them to understand the flow of conversation on forum discussions, we can get all information about node from mouse move on it

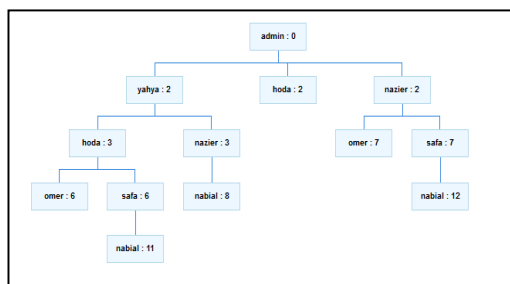


FIGURE 7: MAPPING ASYNCHRONOUS DISCUSSIONS SYSTEM

It displays the message along with the message details. Placing the mouse over the node, which represents the boundary of the discussion thread, reveals a tool-tip, which details the thread name along with the time period, the thread was active. Figure 8 displays all the above mentioned features.

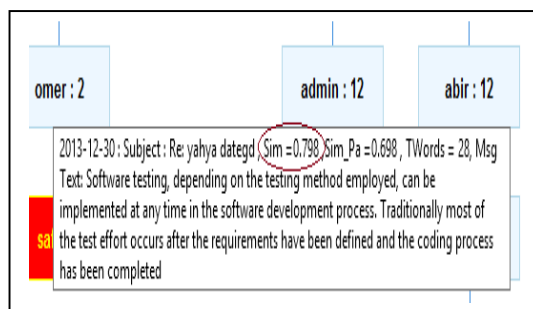


Figure 8: Details-on-demand on map

## F. GROUP OF DISCUSSION

The indicators that features the statistical data related to different activities for each group on a discussion forum.

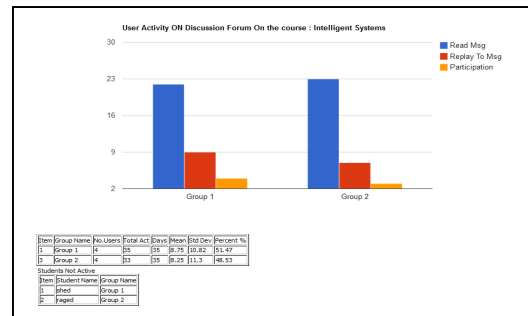


Figure 9: Data indicators for Group activity

## G. MONITORING ASYNCHRONOUS DISCUSSIONS SYSTEM

Monitoring student activity on discussion forum. This features the statistical data related to different activities for each group on a discussion forum.



Figure 10: Monitoring Asynchronous Discussions System

## 5. CHANGE IN TOPIC TITLE

According to Herring (1999)[14],online interactions often tend to change the topic of conversation. This happens when the person responding to the earlier message does not take the earlier contribution into consideration. A slow change in topic direction may occur at various points in the time of discussion. Although each note was written as a meaningful extension of the previous conversation, the thrust of the conversation may tend to change. As a result, the threaded discussion can change the whole area of conversation from one idea to another. Both the messages are related to one another but a change in subject line indicates that there is no direct implication of the previous

message. This might indicate drawing a different conclusion altogether.

A change in the subject also lets the participants reading a post think about whether they have to reply to the original post or the last post that was posted with the changed subject. Van der Pol (2009)[15] states that such situations suggest participants are trying to develop a new understanding of the subject as the respondents not only have to respond about the subject but also guess the meaning of the discussion statement.

Subramani & Hahn (2000)[16] contest that a change in subject has no great impact on how the conversation is perceived by the participants. However, we disagree with their statement as there is clear evidence in the conversations that we studied that a change in subject will affect the flow of conversations. The attention goes from one issue to another within a thread. We could also see that a change in subject line not only affects the flow of conversation, but also could stop the conversation from continuing. On the whole, we can say that a change in area of conversation leads to negligence and lack of awareness to the key area of conversation and this lack of awareness may lead to several educational concerns.

Semantic Similarity Measure has been used in our system tools to find out the degree of relation between the messages. It gives the value of relevance in numerical format

## 6. SEMANTIC SIMILARITY MEASURE

In many research fields such as Psychology, Linguistics, Cognitive Science and Artificial Intelligence, computing semantic similarity between words is an important issue.

Semantic similarity or semantic relatedness is a metric defined over a set of documents or terms, where the idea of distance between them is based on the likeness of their meaning or semantic content as opposed to similarity which can be estimated regarding their syntactical representation. These are mathematical tools used to estimate the strength of the semantic relationship between units of language, concepts or instances, through a numerical description obtained according to the comparison of information supporting their meaning or describing their nature.

Concretely, Semantic similarity can be estimated by defining a topological similarity, by using ontologies to define the distance between terms/concepts. For example, a naive metric for the comparison of concepts ordered in a partially ordered set and represented as nodes of a directed

acyclic graph (e.g., a taxonomy), would be the shortest-path linking the two concept nodes. Based on text analyses, semantic relatedness between units of language (e.g., words, sentences) can also be estimated using statistical means such as a vector space model to correlate words and textual contexts from a suitable text corpus.

## TAXONOMY

The concept of semantic similarity is more specific than semantic relatedness, as the latter includes concepts as antonymy and meronymy, while similarity does not. However, much of the literature uses these terms interchangeably, along with terms like semantic distance. In essence, semantic similarity, semantic distance, and semantic relatedness all mean, "How much does term A have to do with term B?" The answer to this question is usually a number between -1 and 1, or between 0 and 1, where 1 signifies extremely high similarity

## TOPOLOGICAL SIMILARITY MEASURES

There are essentially two types of approaches that calculate topological similarity between ontological concepts:

- Edge-based: which use the edges and their types as the data source
- Information Content : which the main data sources are the nodes and their properties.

## SEMANTIC SIMILARITY

Semantic Similarity or semantic relatedness is a concept of measuring closeness between set of terms or document in context of their meaning. We have two different methodologies for calculating semantic similarity, one by defining a topological similarity, using ontology to define a distance between words or using statistical means such as vector space model to correlate words and textual contexts from a suitable text corpus.

We focus on the former approach using WordNet ontology for semantic similarity calculation. Similarity calculation in this approach relies on the fact that similarity is dependent on both common and distinct features of the objects.

Another approach introduced by Information content based approach capture the informative part of the concept high information content implies more relevance and specific to the subject than the lower one. Path length based approach measure similarity as a function of distance between

concepts in the ontology. Hybrid approach is a combination of the similarity measure mentioned above. Parameters Length, depth and local density forms a part of nonlinear function, which measures similarity between concepts

### WORDNET:

WordNet is lexical ontology for English language. Is a semantic network database for English language, which developed by University Princeton. It models the lexical Knowledge into a taxonomic hierarchy. WordNet contains three databases: one for nouns, one for verbs and one for adverb and adjectives. Terms and concepts are organized into synsets (list of synonyms terms or concepts).

Have been developed in many languages. Basic part of WordNet is synset which a set of synonym of a concept. Synsets are related in some model, such as hypernymy (type of), meronymy (part of) and antonymy (opposite word).

Semantic similarity in WordNet can be divided into two methods, which called path length and information content method. The path length method is to calculate number of node or relation between node in taxonomy. Shorter distance between two concepts have higher similarity. The advantage of path length is not depended to corpus statical and word distribution. The weakness is for taxonomy which has uniform distance. We have considered only is-a relationship and noun concepts in the WordNet hierarchy for similarity calculation. We use WordNet 2.1 which contains nine separate noun hierarchies containing, path between two concepts may not exists in the WordNet. So we create a root node that subsumes all the nine given hierarchies in the WordNet

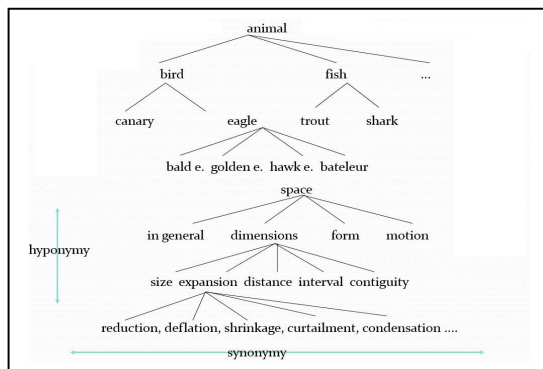


Figure 11: WordNet hierarchy

### Standard Measures of Similarity:

There are several algorithmic approaches for calculating semantic similarity measures of texts

#### 1) Path Based

- Rada et al., 1989 (path),[7]

#### 2) Path + Depth

- Wu & Palmer, 1994 (wup),[3]

#### 3) Path + Information Content

- Resnik, 1995 (res),[4]

- Jiang & Conrath, 1997 (jcn),[4]

- Lin, 1998 (lin).[5]

#### 1) PATH LENGTH (SHORTEST PATH) ALGORITHM

When concepts are organized in a hierarchy, it is convenient to measure similarity according to structural measures that find path lengths between concepts. In fact, there have been a variety of such approaches proposed in general English. Rada, et al. (1989) [17], developed a measure based on path lengths between concepts in the WordNet hierarchy. Shortest path measure focuses on the closeness of two concepts in the hierarchy.

The shorter path between two words/senses in a thesaurus hierarchy graph, the more similar they are: Words are quite similar to parents and siblings. Less similar to words far away in the network.

Pathlen(c1, c2) = Number of edges in shortest path. Path-based similarity often involves a log transform. path-length based similarity:

$$\text{Sim path}(c1, c2) = -\text{Log pathlen}(c1, c2)$$

#### 2) RESNIK SIMILARITY ALGORITHM

(Philip Resnik), 1995, [18] Sun Microsystems Laboratories, presents an alternative to path finding via the notion of information content. This is a measure of specificity assigned to each concept in a hierarchy based on evidence found in a corpus. A concept with high information content is very specific, while concepts with lower information content are associated with more general concepts. The information content of a concept is estimated by counting the frequency of that concept in a large corpus, along with the frequency of all the concepts that are subordinate to it in the hierarchy. The probability of a concept is determined via a maximum likelihood estimate, and the information content is the negative log of this probability.



Resnik defines a measure of similarity that holds that two concepts are semantically related proportional to the amount of information they share. The quantity of shared information is determined by the information content of the lowest concept in the hierarchy that subsumes both the given concepts.

- information-content word-similarity:
- Still relies on structure of thesaurus
- Refines path-based approach using Normalizations based on hierarchy depth
- Represents distance associated with each edge
- Adds probabilistic information derived from a corpus
- Probability of random word being an instance of concept :

$$p(c) = \frac{\sum_{w \in W(c)} \text{count}(w)}{N}$$

Where  $W(c)$  is set of words subsumed by concept  $c$ ,  $N$  is the number of words in corpus and also in thesaurus,  $P(\text{root}) = 1$  since all words are subsumed by root concept.

- The lower a concept in the hierarchy, the lower the probability

Two more definitions are needed:

- 1) Information content of a concept:

$$c: IC(c) = -\log P(c)$$

- basic information theory

- 2) Lowest common subsumer:  $LCS(c1, c2)$

= lowest node in hierarchy that is a hypernym of  $c1$  &  $c2$

The Resnik similarity measure:

$$\text{Sim Resnik}(c1, c2) = -\log P(LCS(c1, c2))$$

estimates common amount of information between words by information content of lowest common subsume.

### 3) LIN SIMILARITY ALGORITHM

Dekang Lin (University of Manitoba - Canada), 1998[19], based on Resnik's similarity. Considers the information content of lowest common subsumer ( $lcs$ ) and the two compared concepts. For example, animal and mammal both are the subsumers of cat and dog, but mammal is lower subsumer than animal for them.

Similarity is about more than just common information the more differences between  $A$  &  $B$ , the less similar they are

Commonality:  $IC(\text{common}(A, B))$

Difference:  $IC(\text{description}(A, B)) - IC(\text{common}(A, B))$

where  $\text{description}(A, B)$  "describes"  $A$  and  $B$

Similarity Theorem:

The Similarity between  $A$  and  $B$  is measured by the ratio between the amount of information needed to state the commonality of  $A$  and  $B$  and the information needed to fully describe what  $A$  and  $B$  are.

$$\text{simLin}(A, B) = \text{common}(A, B) / \text{description}(A, B)$$

the information in common between two concepts is twice the information in the lowest common subsume. Final Lin similarity function for concepts in a thesaurus:

$$\text{Sim Lin}(c1, c2) = \frac{2 \times \log P(LCS(c1, c2))}{\log P(c1) + \log P(c2)}$$

For example:

Figure 13 is a fragment of the WordNet. The number attached to each node  $C$  is  $P(C)$ . The similarity between the concepts of Hill and Coast is:

$$\text{Sim Lin}(\text{Hill\_Coast}) = \frac{2 \times \log P(\text{Geological - Formation})}{\log P(\text{Hill}) + \log P(\text{Coast})}$$

which is equal to 0.59.

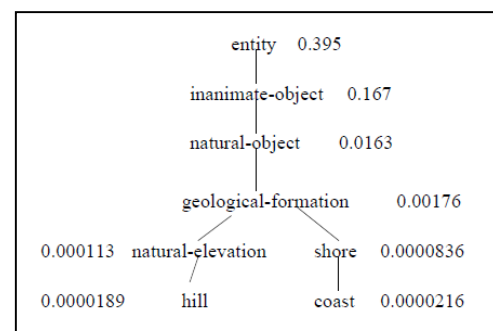


FIGURE 12: IS A FRAGMENT OF THE WORDNET

### 4) JIANG-CONRATH DISTANCE ALGORITHM

Jay J. Jiang (University of Waterloo - Canada), (1997)[20]. Related to  $\text{simLin}$  - expressed as distance instead of similarity considers the

information content of lowest common subsumer (lcs) and the two compared concepts to calculate the distance between the two concepts. The distance is later used in computing the similarity measure.

$$\text{distJC}(c1, c2) = 2 \times \log P(\text{LCS}(c1, c2)) - (\log P(c1) + \log P(c2))$$

Transform into a similarity measure by taking the reciprocal:

$$\text{sim JC}(c1, c2) =$$

$$\frac{1}{2 \times \log P(\text{LCS}(c1, c2)) - (\log P(c1) + \log P(c2))}$$

The Resnik measure may not be able to make fine grained distinctions since many concepts may share the same least common subsumer, and would therefore have identical values of similarity. Jiang and Conrath (1997)[20] and Lin (1998)[19] developed measures that scale the information content of the subsuming concept by the information content of the individual concepts. Lin does this via a ratio, and Jiang and Conrath with a difference

##### 5) WU AND PALMER MEASURE ALGORITHM

Wu and Palmer (1994)[21] present a measure of similarity for general English that relies on finding the most general concept that subsumes both of the concepts being measured. The path length from this shared concept to the root of the ontology is scaled by the sum of the distances of the concepts to the subsuming concept.

Wu and Palmer similarity measure calculate the most specific common ancestor of the two Concepts, with minimum number of is-a Link in the path of the common subsume.

$$\text{Sim Wu} = \frac{2 * h}{h1+h2+h}$$

Here h is depth of the subsume from the root of the hierarchy, h1 and h2 is the minimum number of is-a link from concept c1 and c2 to the most specific common subsumer.

$$\text{depth}(x) = \text{shortest is-a path}(\text{root}, x)$$

The Shortest Path Measure:

The Shortest path measure focuses on the closeness of two concepts in the hierarchy.

$$\text{Sim} = 2 * \text{MAX} - 1$$

Where MAX is the maximum path length between two concepts in the taxonomy and L is the minimum number of is-a link between concepts c1 and c2.

## 7. IMPLEMENTATION OF THE SEMANTIC SIMILARITY MEASURES TOOLS

We developed project in Java Application on Eclipse for calculating semantic similarity Measures between Texts on messages.

This tools design for:

- Calculate semantic similarity Measures between the messages on LMS(Moodle)
- Analysis the content of the messages
- How many correct words in the message?
- Using the different algorithms for calculate semantic Similarity Measures between the text on messages with approach Information content and Edge-based
- Finding the relevance between two messages.
- Search Engine-Based approach to measure semantic similarity between Words in text
- What is the ratio of messages outside the discussion topic or not as close
- How do we know the students are outside (or not as close as) from the discussion topic?
- Where did the direction of discussion change from the topic?
- Return the students to right direction in discussion
- Using the WordNet 2.1 lexical ontology database

### SEMANTIC SIMILARITY COMPUTATION DIAGRAM

The Semantic similarity computation diagram show in Figure 12 (system lifecycle). We proposed for Semantic Similarity Measure Tools strategy is composed of four Phases:

**Phase 1 :Text analysis module**

**Phase 2 :Term construction module**

**Phase 3 :Calculate Semantic module**

**Phase 4:Procedures Semantic Similarity Measure**

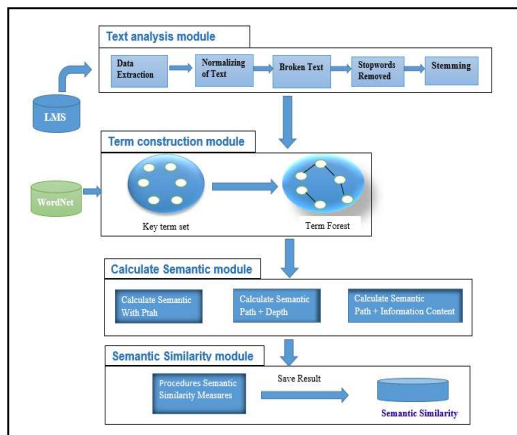


Figure 13: Semantic similarity computation diagram-system lifecycle

### Phase 1 :Text analysis module

This phase is consist of five major setups:

1. Data Extraction
2. Text Normalizing.
3. The Text Breaking
4. Stop words Removing.
5. Stemming and lemmatization

#### 1) DATA EXTRACTION

The system Collect all the messages post in different discussion forums in the data base system Server (LMS).

#### 2) TEXT NORMALIZING

Text normalization is the process of transforming text into a single canonical form that it might not have had before. Normalizing text before storing or processing it allows for separation of concerns, since input is guaranteed to be consistent before operations are performed on it. Text normalization requires being aware of what type of text is to be normalized and how it is to be processed afterwards. There is no all-purpose normalization procedure.

Normalization is the process by which you can perform certain transformations of text to make it reconcilable in a way, which it may not have been before. Let's say, you would like searching or sorting text, in this case you need to normalize that text to account for code points that should be represented as the same text.

The normalization is applicable when you need to convert characters with diacritical marks, change all letters case, decompose ligatures, or convert half-width katakana characters to full-width characters and so on. Normalize transforms Unicode text into an equivalent composed or decomposed form, allowing for easier sorting and searching of text. Normalize supports the standard normalization forms described in Unicode.

Character normalization is a process that can improve recall. Improving recall by character normalization means that more documents are retrieved even if the documents do not exactly match the query. The normalizations are based on Unicode character properties and are not language-dependent.

For Example :

String s1= "Thĩs ỉs â fũñķŷ Šťřĩng" → = " This is funky string "

#### - 3) THE TEXT BREAKING

First of all, each text is broken into sentences. Then, terms in each sentence are extracted as features. A term is regarded as the stem of a single word.

#### - 4) STOP WORDS REMOVING

Stop words are words, which are filtered out before or after processing of natural language data (text). All the stop words on the Text are removed.

#### - 5) STEMMING AND LEMMATIZATION

Lemmatization in linguistics is the process of grouping together the different inflected forms of a word so they can be analysed as a single item. In computational linguistics, lemmatization is the algorithmic process of determining the lemma for a given word. Since the process may involve complex tasks such as understanding context and determining the part of speech of a word in a sentence (requiring, for example, knowledge of the grammar of a language) it can be a hard task to implement a lemmatize for a new language. In many languages, words appear in several inflected forms. For example, in English, the verb 'to walk' may appear as 'walk', 'walked', 'walks', 'walking'. The base form, 'walk', that one might look up in a dictionary, is called the lemma for the word. The

combination of the base form with the part of speech is often called the lexeme of the word.

Lemmatization is closely related to stemming. The difference is that a stemmer operates on a single word without knowledge of the context, and therefore cannot discriminate between words which have different meanings depending on part of speech. However, stemmers are typically easier to implement and run faster, and the reduced accuracy may not matter for some applications.

In linguistic morphology and information retrieval, stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form—generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root. Algorithms for stemming have been studied in computer science since the 1960s. Many search engines treat words with the same stem as synonyms as a kind of query expansion, a process called conflation.

#### **Phase 2 :Term construction module**

The objective of the second module has selected all the words on the Text that exists on WordNet database and get the relation between words. We use WordNet for generating a richer Text representation of the given Text set. As the relationships of relevant terms have been predefined in WordNet, in this module, we intend to use the hypernyms provided by WordNet as useful features for Text

#### **Phase 3 :Calculate Semantic module**

We use the different algorithms above for calculate semantic Similarity Measures between texts on messages with three different methods:

1. Path(Short Path)
2. Approach Information Content
3. Edge-based

#### **Phase 4: Procedures Semantic Similarity Measure**

Semantic vectors for T1 and T2 can be formed from T and corpus statistics. The process of deriving semantic vectors for T1

word  $w$ , define

$$\text{sim}(w_1, w_2) = \max_{c_1, c_2} [\text{sim}(c_1, c_2)]$$

$$\text{Sim}(T_1, T_2) = \sum_{i=1}^n \text{Sim}(W_i, W_{i+1}) / n$$

We get Semantic Similarity Measures value for all algorithm above between Text1, Text12:

- 1) Sim path (T1, T2)=value1
- 2) Sim Resnik (T1, T2)=value2
- 3) Sim Lin (T1, T2)=value3
- 4) Sim JC (T1, T2)=value4
- 5) Sim Wu (T1, T2)=value5

$\text{Sim}(T_1, T_2) = \text{Max}(\text{value1}, \text{value2}, \text{value3}, \text{value4}, \text{value5})$

Save all the results about Semantic Similarity information in database Server.

### **8. SEMANTIC SIMILARITY BETWEEN SENTENCES**

Sentences are made up of words, so it is reasonable to represent a sentence using the words in the sentence. Unlike classical methods that use a precompiled word list containing hundreds of thousands of words, our method dynamically forms the semantic vectors solely based on the compared sentences. Recent research achievements in semantic analysis are also adapted to derive an efficient semantic vector for a sentence. Given two sentences, T1 and T2, a joint word set is formed:

$$T = T_1 \cup T_2 \\ = \{W_1, W_2, \dots, W_n\}$$

The joint word set T contains all the distinct words from T1 and T2. Since inflectional morphology may cause a word to appear in a sentence with different forms that convey a specific meaning for a specific context, we use word form as it appears in the sentence. For example, boy and boys, woman and women are considered as four distinct words and all included in the joint word set. Thus, the joint word set for two sentences

For Example

Text1=" Software engineers apply the principles of engineering to the design, development, maintenance, testing, and evaluation of the software and systems that make computers "

```
Text2=" Software developers are the creative minds
behind computer programs. Some develop the
applications that allow people to do specific tasks
on a computer or other device. "
```

Until Semantic vector14

Max Semantic vector =

$$\{1, 0.84, 0.66, 0.4, 0.1, 0.82, 0.4, 0.73, 0.5, 0.46, 1, 0.58, 0.73, 0.88\}$$

Matrix semantic vectors between text1, text2:

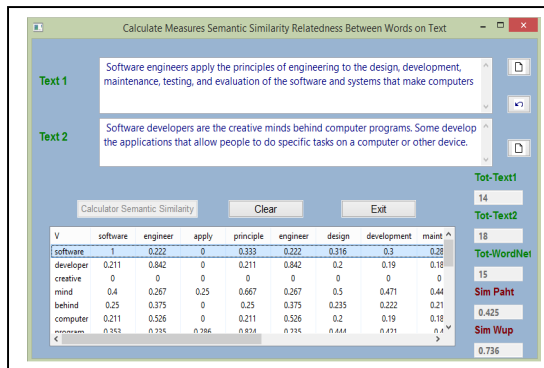


Figure 14: Calculate Semantic Similarity Between Two Text

First, our system will be done the Text analysis module (Phase 1 above) on the two texts in The Following:

1. Text Normalizing
2. The Text Breaking
3. Stop Words Removing.
4. Stemming and lemmatization

Second, we calculate semantic similarity measures between text1, text2, which results as the following:

Total number of words on Text1=14

Total number of words on Text2=18

Total number of words existing in WordNet=15

```
vector Text1 = { Software , engineer, apply,  
principle, design, development, maintenance, test,  
evaluation , system,make,computer}
```

```
vector Text1 = { Software ,developer, mind,behind  
computer, program, develop, ,application, allow  
people ,do ,specific, task, device}
```

The System generators 14 Semantic vectors:

Semantic vector1 =

$$\{1, 0.22, 0, 0.33, 0.31, 0.3, 0.28, 0.28, 0.35, 0.23, 0.3, 0.21\}$$

Semantic vector2 =

$$\{0.21, 0.84, 0, 0.21, 0.20, 0.19, 0.18, 0.18, 0.22, 0.55, 0.2, 0.5\}$$

Semantic vector3 =

{0.4.,0.26,0.3,0.66,0.5,0.47,0.44,0.60,0.57,0.28,0.6,  
0.25

.....

	Software	engineer	apply	principle	design	development	maintenance	test	evaluation	system	make	computer	Max	Sum
Software	1	0.22	0	0.33	0.31	0.3	0.28	0.28	0.35	0.23	0.3	0.21	1	1
developer	0.21	0.84	0	0.21	0.2	0.19	0.18	0.18	0.22	0.55	0.2	0.5	0.84	0.84
mini	0.4	0.26	0.3	0.66	0.5	0.47	0.44	0.50	0.57	0.28	0.6	0.25	0.66	0.66
behind	0.25	0.37	0	0.26	0.23	0.22	0.21	0.21	0.26	0.4	0.2	0.35	0.4	0.4
computer	0.21	0.52	0	0.21	0.21	0.19	0.18	0.18	0.22	0.77	0.19	1	1	1
program	0.35	0.23	0.3	0.82	0.44	0.42	0.41	0.50	0.5	0.35	0.3	0.73	0.82	0.82
develop	0	0.25	0.4	0	0.26	0	0	0.28	0	0	0.4	0	0.4	0.4
application	0.33	0.22	0	0.44	0.73	0.6	0.57	0.38	0.7	0.23	0.4	0.22	0.73	0.73
allow	0	0.28	0.5	0	0.33	0	0	0.33	0	0	0.5	0	0.5	0.5
people	0.42	0.29	0.28	0.42	0.4	0.37	0.36	0.35	0.46	0.3	0.37	0.26	0.46	0.46
do	0.33	0.28	0.5	0.44	0.56	0.5	0.47	0.38	0.58	0.23	0.1	0.21	1	1
specific	0.35	0.23	0	0.58	0.44	0.42	0.4	0.50	0.5	0.25	0.5	0.22	0.58	0.58
task	0.33	0.22	0.3	0.44	0.73	0.6	0.57	0.38	0.7	0.23	0.4	0.21	0.73	0.73
device	0.23	0.58	0	0.23	0.22	0.21	0.2	0.20	0.25	0.87	0.2	0.88	0.88	0.88

Figure 15: Matrix Semantic Vectors Values

Finally, we get the Semantic Similarity between Text1, Text2 is equal 0.74 by Wu and 0.43 by path

## 9. START CLASSIFICATION MESSAGES TEXT ON LMS

The system tools start the following Phases:

### *Phase 1: TEXT analysis module*

### *Phase 2: Term construction module*

### **Phase 3: Calculate Semantic module**

#### **Phase 4: Procedures Semantic Similarity Measure**

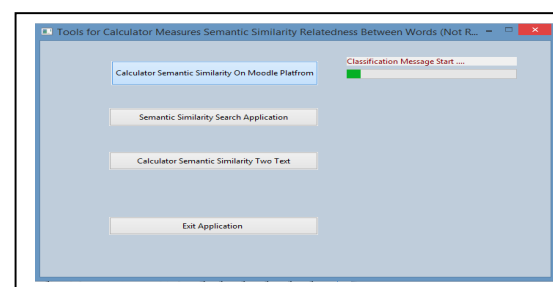


Figure 16: Calculate Semantic Similarity On LMS

Content analysis of the text messages start and get all the information about the text through:

- Calculate semantic similarity measures between the messages.
- What is relevance between the messages and the first post message (the discussion topic)?
- What is relevance between the messages and the parent post message?



- What is the ratio of messages outside of the discussion topic?
- Where did the direction of discussion change from the topic? and who is the student?
- The correct words on the message and size
- Students Lurkers?

Looking at the discussion map on Figure 7, which tells us how to identify the students who post a message outside or (not as close as) of the discussion topic.

Figure 17 : Matrix Semantic Vectors Values

We can easily see this through the discussion map on Figure 18. We can also note the following points:

- The messages indicated by a different colour (red colour), are not as close as the previous post (Parent message)
- Four messages are not as close as the previous post (Parent message)
- Five messages are not as close for the discussion topic
- The Semantic similarity value (sim=0.68) between the message post with student (nabial) and parent (nazier)

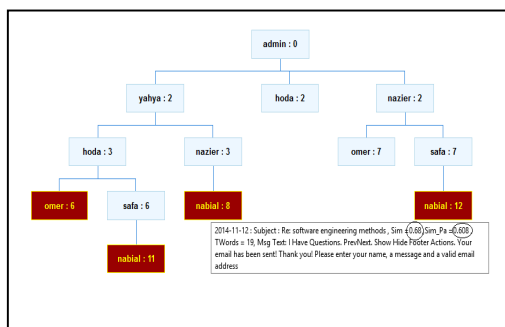


Figure 18 : Example For Relevance Two Messages

- The Semantic similarity value (sim\_pa = 0.60) between the messages post with student nabial and discussion topic

- 38% percent rate of the messages outside of the discussion topic
- 31% percent rate of the messages not as close as the previous post (parent message).
- One student Lurker (abier).
- The messages indicated by a different colour

red is not as close as from the discussion topic  
Figure 19 : Data Indicators For Forum Discussion

## 8) SEMANTIC SIMILARITY SEARCH ON MESSAGES

Search Engine-Based approach to measure semantic similarity between words in Figure 20. We can determine the Sentences Similarity between texts on the messages and topic you wants to research it in forum discussion for any course on LMS. The system Search on all messages post and show the result with all information about this messages and semantic similarity value and who is sent, date, others

Figure 20 : Semantic Similarity Search On Messages

## 10 CONCLUSION

This paper to design and implement analytical tools system to track students' online learning activities based on LMS logs and semantic similarity

between Texts on messages post based on semantic and word order information.

Calculate semantic similarity measures between the messages on LMS, Analysis the content of the messages. Semantic similarity is derived from a lexical knowledge base and a corpus. The lexical knowledge base models common human knowledge about words in a natural language this knowledge is usually stable across a wide range of language application areas A corpus reflects the actual usage of language and words.

Our piloting and evaluation of this tool suggests that it offers an appropriate and effective means of managing discussion forum messages in a visual format, which then facilitates decision-making for additional in-depth analysis. The initial visual pattern suggests broad categories into which discussion threads may be falling, and offers directions for further analysis. While our examples are drawn from a limited pilot study, they have nonetheless yielded indications about how asynchronous Forums in Moodle (specifically) are used by teachers and students in a campus-based university setting. We then visualize the results with a simple graphical interface. Is also able to automatically send feedback and reminder emails based on students' activity histories. We have integrated the module into the Moodle LMS and made some small interface changes to Moodle. Intuitively, the presence of the system popularity bars should encourage students to check the course materials more frequently and promptly if he or she sees most of their classmates have already done so. Our hypothesis is that the availability of system statistics will positively affect both how an instructor adapts the course and how students learn. Monitoring student learning activity is an essential component of high quality education, and is one of the major predictors of effective teaching (Cotton, 1998)[34]. Although the proposed approach system tools has recorded very good performance It is important to acknowledge the limitations of the system tools: it is in essence a support for coding and management of the data, rather than offering in and of itself a new method for analysing that data. For such analysis, we need to consider the wider model used by Schrire [17] or indeed to pursue existing qualitative methods. Tools does not yet allow us away to analyse the multi-modal nature of the student discourse in unmoderated Forums, and the inclusion of images, sounds and other media which students are now accustomed to using. This is a further area of work we need to:

- Upgrade our analytical tools system to take the data from social media like Facebook, Twitter, Google+, and others.
- Allow to analyse the different Multimedia data Processing like (Image, video, audio, Ext)

## REFERENCES:

- [1] Sing, C., & Khine, M. "An analysis of interaction and participation patterns in online community". Educational Technology & Society, vol. 9(1), pp. 250-261,2006
- [2] Romero, C., Ventura, S. & Garcia, E., Data mining in course management systems : Moodle case study and tutorial. Computers & Education. Volume: 51, Issue: 1. Elsevier, 368-384 ,2008
- [3] Romero, C., Ventura S., Pechenizkiy M. and Baker R. Handbook of Educational Data Mining. CRC Press, 9-26,2010
- [4] May, M., George, S., & Prevot, P. Tracking, analyzing and visualizing learners' activities on discussion forums. In Proceedings of the 6th IASTED International Conference on Web-based Education (WBE) (pp. 649–656),2007
- [5] May, M., George, S., & Prevot, P. A closer look at tracking human and computer interactions in web-based communications. Interactive Technology and Smart Education,5(3), 170–188,2008
- [6] Donath, J., Karahalios, K., & Viegas, F. B. Visualizing conversation. In. Proceedings of the 32nd Annual Hawaii International Conference on System Science, Vol. 4 (4),1999
- [7] Mazza, R. & Dimitrova, V. "Visualizing Student Tracking Data to Support Instructors in Web-Based Distance Education". In Proceedings of 13th International Conference on World Wide Web, pp. 154-161,2004
- [8] Riccardo Mazza, Marco Bettoni, Marco Faré, Luca Mazzola. MOCLog – Monitoring Online Courses with log data. In: In Retalis, S., and Dougiamas, M."1st Moodle Research Conference Proceedings", pp. 14-15, 2012.
- [9] Mazza, R. & Botturi, L. Monitoring an Online Course with the GISMO Tool: A Case Study. Journal of Interactive Learning Research. Vol. 18 N. 2, 251-265. ,2007
- [10] Mazza, R. & Dimitrova, V. CourseVis: A Graphical Student Monitoring Tool for Facilitating Instructors in Web-Based Distance Courses. International Journal in Human-

- Computer Studies (IJHCS). Vol. 65, N. 2 , 125-139,2007
- [11] Howard T.W., Eric G., Danyel F. & Marc S. "Visualizing the Signatures of Social roles in online Discussion Groups". Journal of social structure, vol.8 (2),2007
- [12] Mohamed, R., Ferguson, J. D., MacCormick, A., & Elswailer, D. Developing bulletin board visualizations. In Proceedings of the IADIS International Conference on Web Based Communities (pp. 11–18),2004
- [13] Gibbs W.J., Olexa V. & Bernas R.S. "A Visualisation Tool for Managing and Studying Online Communications". Educational Technology & Society. Vol. 9(3), pp. 232-243,2006
- [14] Herring, S. "Interactional coherence in CMC". Proceedings of the 32nd Hawaii International Conference on System Sciences. Los Alamitos, CA: IEEE Computer,1999
- [15] Van der Pol, J. "Online Learning Conversations: potential, challenges, and facilitation". In Payne. C. R. (ed.) Technology and Constructivism in Higher Education: Progressive Learning Frameworks. Hershey, PA: IGI-Global,2009
- [16] Subramani, M. R., & Hahn, J. "Examining the Effectiveness of Electronic Group Communication Technologies: The Role of the Conversation Interface". MISRC Working Paper 00-03, 2000
- [17] R. Rada, H. Mili, E. Bichnell, and M. Blettner, "Development and Application of a Metric on Semantic Nets," IEEE Trans. System, Man, and Cybernetics, vol. 9, no. 1, pp. 17-30, 1989.
- [18] Resnik, Philip. Using information content to evaluate semantic similarity. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Canada, pages 448–453,1995
- [19] Lin, Dekang. Automatic retrieval and clustering of similar words. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING–ACL '98), Montreal, Canada, pages 768–774, ,1998
- [20] J.J.Jiang and D.W Conrath, Semantic Similarity Based on Corpus Statistics and Lexical Taxanomy, Procs.ROCKLING X, 1997.
- [21] Z.Wu and M.Palmer. Verb semantic and Lexical Selection .In Proceedings of 32 Annual Meeting of the association of computer Linguistics(ACL 994),Las Cruces,New Mexico,1994.
- [22] Abik, M. & Ajhoun, R.. "Normalization and Personalization of Learning Situation: NPLS", International Journal of Emerging Technologies in Learning (iJET), vol. 4(2), pp. 4-10, 2009
- [23] Sánchez, D.. "A methodology to learn ontological attributes from the Web." Data & Knowledge Engineering 69(6): 573-597,2010
- [24] Harispe, S., et al. "A framework for unifying ontology-based semantic similarity measures: a study in the biomedical domain." J Biomed Inform 48: 38-53, 2014
- [25] McInnes, B. T. and T. Pedersen. "Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text." J Biomed Inform 46(6): 1116-1124, 2013
- [26] Chen, C.-L., et al.. "An integration of WordNet and fuzzy association rule mining for multi-label document clustering." Data & Knowledge Engineering 69(11): 1208-1226., 2010
- [27] Bratitsis, T., & Dimitracopoulou, A.. Interpretation issues in monitoring and analyzing group interactions in asynchronous discussions. International Journal of e-Collaboration (IJeC), 4(1), 20–40, 2008
- [28] Schrire, S. "Interaction and cognition in asynchronous computer conferencing". Instructional Science. vol. 32. pp. 475–502,2004
- [29] Schrire, S."Knowledge building in asynchronous discussion groups: going beyond quantitative analysis". Computers & Education, vol.46, pp.49-70,2006
- [30] Sujana Jyothi, Claire McAvinia, John Keating (2012) ."A visualisation tool to aid exploration of students' interactions in asynchronous online communication" Computers & Education, Volume 58, Issue 1, Pages 30-42, 2012
- [31] Jong, B., Chan, T. & Wu, Y." Learning Log Explorer in E-Learning Diagnosis". IEEE Transactions on education. Volume: 50, Issue: 3, IEEE, 216-228,2007
- [32] Chen, C.-L., et al. "An integration of WordNet and fuzzy association rule mining for multi-label document clustering." Data & Knowledge Engineering 69(11): 1208-1226,2010



- 
- [33] Riccardo Mazza, Marco Bettoni, Marco Faré, Luca Mazzola. MOCLog – Monitoring Online Courses with log data. In: In Retalis, S., and Dougiamas, M."1st Moodle Research Conference Proceedings", pp. 14-15, 2012.
- [34] Cotton, K., "Monitoring student learning in the classroom", in School Improvement Research Series Close-Up #4, May 1988.