# AUTOMATIC ANALYSIS OF ASYNCHRONOUS DISCUSSIONS

Breno Fabrício Terra Azevedo[1], Patricia Alejandra Behar[2] and Eliseo Berni Reategui[2]
[1]*Department of Information Technology, IFF, 273, Dr. Siqueira, Zip Code 28030-130, Campos dos Goytacazes-RJ, Brazil*
[2]*Graduate Program in Information Technology in Education, UFRGS*
*110, Paulo Gama, building 12105, 3th floor, room 332, Zip Code 90040-060, Porto Alegre-RS, Brazil*

Keywords: Qualitative Analysis, Asynchronous Discussions.

Abstract: This paper presents results of an automatic analysis of text contributions made by students in asynchronous discussions. The study was carried with the MineraFórum software. Data collected with the program were compared to appraisals made by teachers. Results show that the average of the analyses of posts made MineraFórum is similar to the average obtained in the analyses made by teachers.

## 1 INTRODUCTION

According to Gilbert and Dabbagh (2005), an important pedagogical benefit of asynchronous communication is its potential to support the co-construction of knowledge through discourse.

The study presented by Garrison, Anderson and Archer (2000) suggests that text-based communication offers time for reflection. The authors' review of the literature indicates that written communication is closely related to careful and critical thinking. Writing can be crucial when the objective is to facilitate thinking over complex issues, and meaningful and deep learning.

Palloff and Pratt (2004) say asynchronous discussions must be stimulated by teachers, as they are the best way to establish interactions among students. According to the authors, student interactions provide time for reflecting over studied educational contents. The ability to reflect is crucial to virtual students, and should be stimulated. Discussion forums are a suitable space to offer this type of action. By participating in the discussion or simply replying to messages, students indicate that they are actually reflecting. The authors also emphasize the importance of the teacher's role in discussion forums. Besides writing messages of support and motivation to students, and answering their questions, teachers should observe the level of participation of each learner. In case the teacher identifies that a student is not participating properly or digressing from the topic of discussion, he/she should try to help learners to overcome their difficulties, and solve problems.

Getting involved in asynchronous discussions, such as in forums, is an important activity for students. By analyzing student interactions in forums, the teacher can diagnose information on learners. However, if the teacher has a significant number of students, he/she will need a great amount of time to do text analysis. A resource that allows the automatic analysis of posts in discussion forums can be of great help to teachers. This resource may allow teachers to identify students who are debating over the topic of discussion, as well as those who are not. By doing so, teachers can have extra time to find out the reasons why some of the learners did not discuss concepts related to the topic. In case the teacher identifies students with learning difficulties, help can be offered.

To perform automatic analysis of texts produced by students in asynchronous discussion, this paper presents a study carried with the software MineraFórum[i].

MineraFórum (Azevedo et al., 2011a; 2011b) uses text mining techniques to analyze posts in threaded discussion. By doing this analysis, it is possible to identify if text contributions produced by learners are relevant or irrelevant in the debate.

Next section presents a brief introduction to text mining. Section 3 informs on some works that use this technique in the analysis of discussion forums. Section 4 explains the software MineraFórum. Section 5 describes the experiments, and section 6 presents the concluding remarks.

Terra Azevedo B., Behar P. and Berni Reategui E..
AUTOMATIC ANALYSIS OF ASYNCHRONOUS DISCUSSIONS.
DOI: 10.5220/0003896900050012
In *Proceedings of the 4th International Conference on Computer Supported Education* (CSEDU-2012), pages 5-12
ISBN: 978-989-8565-06-8

## 2 TEXT MINING

According to Feldman and Sanger (2007), text mining can be defined as an intensive process of knowledge in which the user interacts with a great number of documents by using tools to perform analysis. The objective is to extract useful information from a collection of documents. This information is identified in interesting patterns found in non-structured text data.

Text mining systems are based on pre-processing routines, algorithms for discovering patterns, and elements for presenting results. The system user interacts with the pre-processing stage, with the mining nucleus, and with result output.

Pre-processing operations are based on the identification and extraction of representative features of documents in natural language. These operations are responsible for changing non-structured data, stored in collections of documents, into a structure expressed in an intermediary model (Feldman and Sanger, 2007; Tan, 1999). The intermediary models are based on choice of the minimum text unit: word, concept, sentence, paragraph, or document (Torre et al., 2005).

Operations in the mining nucleus, also called knowledge distillation processes, represent the core of a text mining system, and involve: pattern discovery, trend analysis, and incremental algorithms for knowledge discovery. The most used mechanisms are distributions and proportions, sets of frequent concepts, and associations. Activities can also be related to comparisons, and to the identification of levels of interest with some patterns (Feldman and Sanger, 2007).

Elements involved in the presentation of results represent the system interface, with navigation function, and access to the language used in the search (Feldman and Sanger, 2007; Puretskiy et al., 2010; Tan, 1999).

Text mining explores techniques and methodologies from areas such as information retrieval, information extraction, and corpus linguistics. To extract useful information, one must discover relevant characteristics in the documents, the most usual being: characters, words, terms, and concepts. Characters are individual letters, numbers, special characters, and spaces. Words are represented by clusters of characters. Terms are unique words or sets of words selected straight from the text. Concepts are features generated for a document by using different methodologies. Hybrid approaches can be used to generate document representation based on features. For example, one can first extract terms from a text, and then adapt them by comparing them to a list of relevant topics (concepts) obtained by categorization (Feldman and Sanger, 2007).

Considering the four features described above (characters, words, terms, and concepts), terms and concepts are those that possess the highest semantic level. There are many advantages in using those features to represent documents in text mining. Representations using terms can be more easily generated from the original text if compared to concepts. However, representations with concepts are better than any other. They can also be processed in order to support very sophisticated hierarchies by using knowledge of the domain given by ontologies and knowledge bases (Feldman and Sanger, 2007).

Text mining using graph technique discovers words with greater occurrence in texts, and identifies if they are near one another. The graph obtained in the mining process presents the most frequent words in its nodes. Associations between nodes indicate the proximity between words. Figure 1 represents the graph generated from the text “There are several techniques in text mining. Some techniques used in text mining include: information extraction, topic tracking, summary production, text categorization, text clustering, conceptual links, information visualization, analysis of questions and answers”.

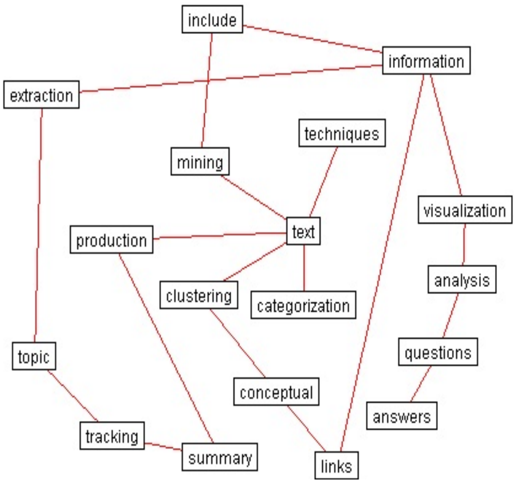


Figure 1: Graph generated from a text.

## 3 ANALYSIS OF DISCUSSION FORUMS WITH TEXT MINING

Rebedea et al. (2008) present an analysis of chats that may be used in threaded discussion. Their study proposes extraction of socio-semantic data from conversations produced by participants using text mining techniques based on ontologies. This method uses a combination of a social-cultural and dialogic perspective with text processing techniques. The study also presents the software developed to discover the most relevant topics in a debate, the contribution of each participant in the conversation, and how it can offer a representation of multiple voices in the conversation. Text mining was used to: analyze if the content of chat messages is related to the discussion theme, and determine the moment a new topic is introduced in the discussion. WordNet[ii] was used to identify synonyms in the texts selected for the study.

Ravi and Kim's work (2007) presents an approach to make automatic identification of features in students' posts in threaded discussions. The authors used word sequence resources and SVM algorithms (Support Vector Machine) to develop "speech acts" classifiers to identify purposes in individual messages, such as: questions, answers, formulations, corrections. Classifiers were used to find messages containing questions or answers. Authors used a set of rules for topic analysis in order to find out those that could contain unanswered questions and need the teacher's attention.

A discussion forum with advanced technological features is presented by Li et al. (2008). This project uses domain ontology and text mining techniques. In this study, transcriptions of discussion forums are automatically changed into a structural modelling in three stages: topic acknowledgment, identification of the type of transcription, and the semantic association among them. The first step clusters messages from a set of discussion into a document. Each document is represented by a vector of weighted terms. Cosine method is used to calculate similarity between the vector in the document and the vector of concepts in the domain ontology. The second step identifies six types of messages: question, opinion, suggestion, recommendation, request, and reference. The third step uses the SLN model (Semantic Link Network) to organize texts with semantic association. The forum used in the study offers three functions to teachers: search of information considered useful to their needs, thematic navigation through messages, and recommendation to students that might be interested in communicating and collaborating. An experiment was carried to demonstrate the effectiveness of the approach to find learning peers with the same interests, and message search with thematic navigation.

## 4 MINERAFÓRUM

MineraFórum is a program that makes qualitative analysis of posts in discussion forums. It was developed at NUTED[iii]/PGIE/UFRGS. At present, version 3.0 of the software is being used. This program is capable of calculating the relevance of each post within a particular discussion. To analyze the content of text contributions, the program uses the text mining using graph technique. Figure 2 presents the main interface of MineraFórum.

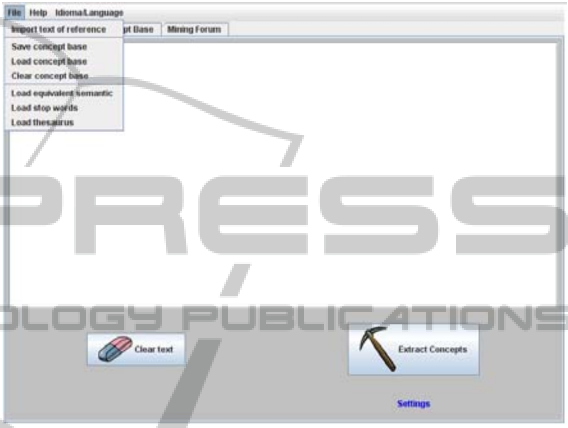


Figure 2: Main Interface of MineraFórum showing selection of the "File" menu.

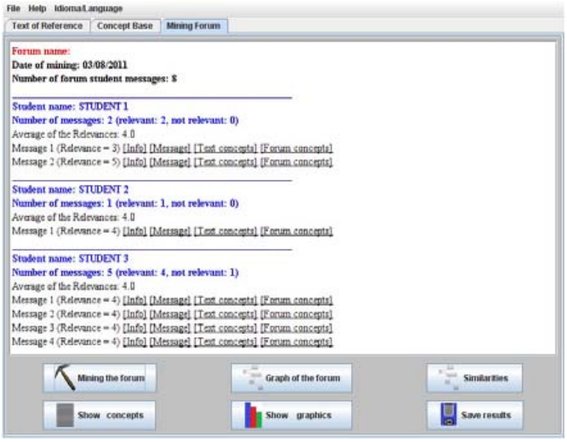
Some resources offered by version 3.0 of the MineraFórum are listed below:

- It allows the user to load or type reference text on the discussion topic.
- If the user wishes so, instead of informing a text of reference, it is possible to type concepts considered to be relevant in the discussion, and make associations among them.
- It uses a thesaurus in the mining process. This type of dictionary was previously defined in the software. Nevertheless, if the user finds it necessary, another synonyms dictionary can be informed. Synonyms are important when MineraFórum compares words typed in the posts with the concepts considered as relevant in the reference text.
- In addition to the thesaurus, the user can inform words that are semantically equivalent.
- It calculates the relevance of each post.
- It shows a graphic with the mean relevance level of messages posted by each author.
- It identifies similar messages written in the discussion forum.
- It allows results of the mining process to be stored in html files.

- It shows a report with information on the analysis of posts: total number of messages written by each student, the amount of relevant contribution made by individual learners, concepts used in the relevant posts, relevance of each message, information if the message is similar (or not) in the forum, the relevance average in posts written by each student, the number of times each message was cited in the debate.

Figure 3 shows the MineraFórum interface after the user selected the button "Mining Forum". Informations about the mining are presented: name of the forum, data of mining, and the total number of messages posted by all the students. For each learner, one can see: full name, total of messages, relevance average of posts. For each message, the software shows the relevance value and four links ([Info] [Message] [Text concepts] [Forum concepts]).

- Info: provides information on the number of times the message was cited in the forum, and if it is similar to any other. Figure 4 shows informations about the first post of student 3.
- Message: shows the first characters of the message. Figure 5 shows the second message of student 3.
- Text concepts: presents the concepts indicated by the reference text found in the post.
- Forum concepts: shows the main concepts used in the forum and found in the posts. Figure 6 shows the relevant concepts cited in the forum that were found in third post of student 3.



Forum name:
Date of mining: 03/08/2011
Number of forum student messages: 8

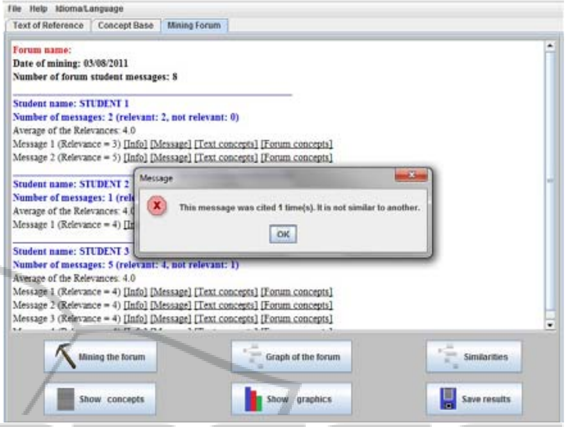
Student name: STUDENT 1
Number of messages: 2 (relevant: 2, not relevant: 0)
Average of the Relevances: 4.0
Message 1 (Relevance = 3) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 5) [Info] [Message] [Text concepts] [Forum concepts]

Student name: STUDENT 2
Number of messages: 1 (relevant: 1, not relevant: 0)
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]

Student name: STUDENT 3
Number of messages: 5 (relevant: 4, not relevant: 1)
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 3 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 4 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]

Figure 3: Interface of the "Mining Forum" window.

During the process of analyzing posts, MineraFórum organizes and clusters the student's messages. The software calculates the relevance value of each message. To do this procedure, three criteria are considered: the thematic relevance of the message (TR), relevance of message reference (MR), and similarity of the message (MS).



Forum name:
Date of mining: 03/08/2011
Number of forum student messages: 8

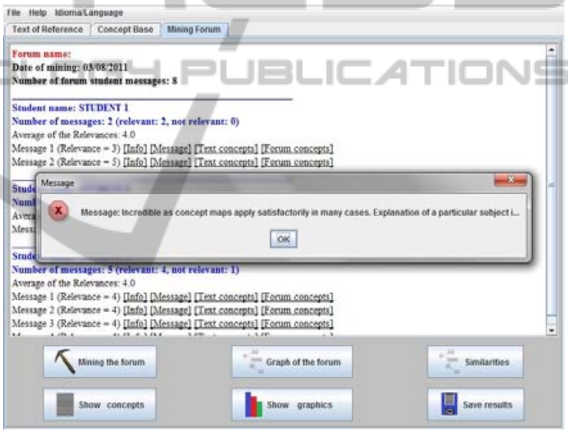
Student name: STUDENT 1
Number of messages: 2 (relevant: 2, not relevant: 0)
Average of the Relevances: 4.0
Message 1 (Relevance = 3) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 5) [Info] [Message] [Text concepts] [Forum concepts]

Student name: STUDENT 2
Number of messages: 1 (rel
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [In

This message was cited 1 time(s). It is not similar to another.

Student name: STUDENT 3
Number of messages: 5 (relevant: 4, not relevant: 1)
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 3 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]

Figure 4: Informations about the first post of student 3.



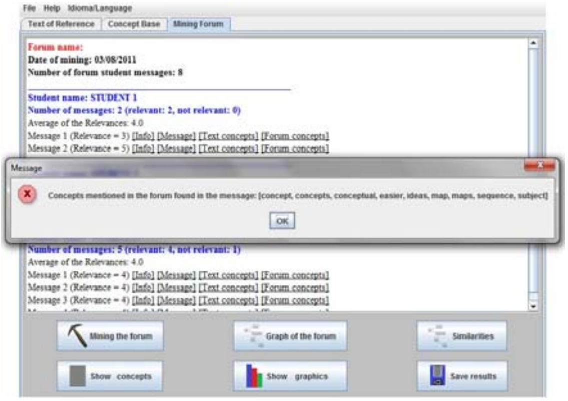
Forum name:
Date of mining: 03/08/2011
Number of forum student messages: 8

Student name: STUDENT 1
Number of messages: 2 (relevant: 2, not relevant: 0)
Average of the Relevances: 4.0
Message 1 (Relevance = 3) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 5) [Info] [Message] [Text concepts] [Forum concepts]

Message: Incredible as concept maps apply satisfactorily in many cases. Explanation of a particular subject i...

Number of messages: 5 (relevant: 4, not relevant: 1)
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 3 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]

Figure 5: Second message of student 3.



Forum name:
Date of mining: 03/08/2011
Number of forum student messages: 8

Student name: STUDENT 1
Number of messages: 2 (relevant: 2, not relevant: 0)
Average of the Relevances: 4.0
Message 1 (Relevance = 3) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 5) [Info] [Message] [Text concepts] [Forum concepts]

Concepts mentioned in the forum found in the message: [concept, concepts, conceptual, easier, ideas, map, maps, sequence, subject]

Number of messages: 5 (relevant: 4, not relevant: 1)
Average of the Relevances: 4.0
Message 1 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 2 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]
Message 3 (Relevance = 4) [Info] [Message] [Text concepts] [Forum concepts]

Figure 6: Relevant concepts cited in the forum that were found in third post of student 3.

To calculate the thematic relevance of messages (TR), MineraFórum performs the following actions:

a) From the reference text, it builds a graph of the discussion topic indicated by the user. In this process, stopwords (words that can be removed in the mining stage, such as adverbs, articles, and prepositions) can be deleted, and the most recurring words in the text are identified. The most frequent words represent the most relevant concepts in the mined text, and correspond to the vertices in the graph. The edges between the vertices are created according to the proximity between words. If the user decides to insert important concepts related to the discussion topic, instead of indicating the reference text, MineraFórum builds the graph from those concepts.

b) Automatic loading of all posts. The software interacts with the Virtual Learning Environment where the forum took place and get all the messages.

c) Generation of a graph by mining each post.

d) To calculate thematic relevance of a post in relation to the reference text, MineraFórum analyzes the correspondence between the generated graph from the text and graph built from the message. In this case, one can identify which vertex in the first graph are equivalent to those in the second graph. MineraFórum considers that two vertices are equivalent if they present similar content, that is, (i) if they have the same words, (ii) if the words can be reduced to the same stem, (iii) if they have synonyms, and (iv) if they are semantically equivalent. In the second stage of analysis, the program uses a formula that takes into consideration three aspects of the equivalent vertices: the amount of such vertices in the two graphs, the distance between them in the respective graph, and their weight in their own graphs. The resulting value corresponds to the thematic relevance of the post in relation to the reference text ($TR_{TX}$).

e) Generation of a text graph built from the whole set of posts, named “forum graph”. The procedure used to find $TR_{TX}$ is used to calculate the thematic relevance of the post in relation to the forum graph. The resulting value is named $TR_{TF}$.

f) Thematic relevance (TR) of a post is calculated from the average between $TR_{TX}$ and $TR_{TF}$.

To calculate the MR value of a post, the software divides the number of times the message has been cited by the total number of posts in a forum. The computation of the similarity of a message (MS) with other was held with text mining using graphs. The graphs of the messages that have similar values of TR are compared to verify if the posts are similar. If the post is similar to another in the forum, the MS value will be equal to TR, with a negative sign.

The relevance value of a post (PR) is obtained from the weighted mean between TR and MR. If the message is similar to another, the MS value will be subtracted from PR. If the text contribution is not similar to another in the forum, then the calculated value for PR will be maintained.

The final value for PR is converted into a whole value, in a scale from 0 (zero) to 5 (five). Zero value means the message is not relevant in the debate. Value five indicates that the post has maximum relevance.

Table 1 presents a comparison between MineraFórum and the correlational studies presented in section 3. The comparison shows an analysis of the similarities and differences between MineraFórum and other works.

Table 1: Comparison between MineraFórum and correlational studies.

| Authors | Similarities between MineraFórum and correlational work | Differences between MineraFórum and correlational work |
|---|---|---|
| Rebedea et al. (2008) | Tool that uses text mining techniques to analyze relevance of messages in relation to the discussion topic. | Analyzes chats messages to provide indicators related to Bakhtin’s theory of poliphony.<br>The tool is not integrated into a VLE. |
| Ravi and Kim (2007) | Classifiers to analyze content of forum messages as an aid for teachers. | Classifiers analyze texts by identifying “speech acts”.<br>Indicators given by classifiers are different from those presented by MineraFórum. |
| Li et al.(2008) | Tool that uses text mining techniques to analyze relevance of posts in relation to the discussion topic. | The tool represents texts with vectorial space model, and identifies if they are similar by using the cosine similarity measure.<br>The system developed by the authors offers help to students, while MineraFórum presents information to help teachers. |

MineraFórum is a resource that can be used by teachers in a Virtual Learning Environment (VLE). It is integrated into ROODA[iv] (Behar, 2007), ETC[v] (Macedo et al., 2010) and MOODLE[vi] platforms. The teacher can choose the discussion forums to mine the messages.

## 5 EXPERIMENTS CARRIED WITH MINERAFÓRUM

To validate results using MineraFórum, five experiments were made. In each experiment, a discussion forum was analyzed by both the software and two teachers. The forum topics were distinct, as well as school level, and course modality. Forums selected for the experiments were extracted from ROODA, ETC and MOODLE platforms.

The purpose of the experiments was to compare the average of message relevance calculated by MineraFórum with the average obtained in the assessments made by teachers. It is worthy observing that the program calculates a relevance value between zero and five for each text contribution. For that reason, teachers were requested to use the same values in their evaluations. Evaluation of the messages was made by the following group of teachers: two Ph.Ds and four Ms teachers. Two teachers have long experience in distance learning, and four have little experience.

Tables 2 and 3 describes the characteristics of the forums used in the experiment: the VLE in which they were offered, the discussion theme, the course, the school level, and the modality of each course. Tables 4 and 5 present values obtained in the analysis made by both MineraFórum and teachers. Information in this table includes: the teacher who assessed the messages, the number of students who participated in each forum, the number of posts, the average obtained from the software analysis, the average obtained from the teachers' appraisals, and the degree of similarity between the averages. The average of analysis of the MineraFórum and the average of the teachers' assessments were obtained by summing the values assigned to each post, dividing by the number of messages. The degree of similarity was obtained dividing the two averages.

Table 2: VLE and theme of analyzed forums.

| Forum | VLE | Theme |
|---|---|---|
| 1 | Rooda | Learning as transformation |
| 2 | Rooda | Learning as transformation |
| 3 | ETC | Team work |
| 4 | ETC | Competence Development |
| 5 | Moodle | Digital Ceritification |

Tables 4 and 5 present the degree of similarity between the average of the analysis performed by MineraFórum and the one obtained in the analysis made by teachers. This result reveals that the average of the analysis calculated by the software is similar to the average of the assessments made by teachers.

Table 3: Characteristics of analyzed forums.

| Forum | Course | School Level | Modality |
|---|---|---|---|
| 1 | Education (group 1) | Undergraduate | Distance Education |
| 2 | Education (group 2) | Undergraduate | Distance Education |
| 3 | Extension | Extension | Face to face education |
| 4 | Extension | Extension | Face to face education |
| 5 | Informatio n Systems | Extension | Face to face education |

Table 4: Analysis by MineraFórum and by Teacher 1.

| Forum | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Teacher | A | A | C | C | E |
| Number of students | 28 | 31 | 18 | 11 | 12 |
| Number of messages | 48 | 73 | 76 | 42 | 12 |
| Average of analysis by MineraFórum | 2,92 | 2,88 | 3,00 | 3,21 | 3,42 |
| Average of analysis by teacher | 2,79 | 2,32 | 3,61 | 3,90 | 2,67 |
| Degree of similarity between analyses by MineraFórum and teacher | 95,71% | 80,48% | 83,21% | 82,32% | 78,05% |

In Table 5, the degree of similarity in forum 3 presented the least value, 76,00%. Analysis of forum 2, as shown in Table 5, obtained the highest value, 96,77%.

Table 5: Analysis by MineraFórum and by Teacher 2.

| Forum | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Teacher | B | B | D | D | I |
| Number of students | 28 | 31 | 18 | 11 | 12 |
| Number of messages | 48 | 73 | 76 | 42 | 12 |
| Average of analysis by MineraFórum | 2,92 | 2,88 | 3,00 | 3,21 | 3,42 |
| Average of analysis by teacher | 3,29 | 2,97 | 3,95 | 4,12 | 2,92 |
| Degree of similarity (analyses by MineraFórum and teacher) | 88,61% | 96,77% | 76,00% | 78,03% | 85,37% |

It is relevant to remind that the software uses three criteria to analyze messages: thematic relevance of the text contribution, relevance of message reference, and similarity of the post. When assessing the posts, each teacher used their own criteria.

Figure 7 presents the variation of the degree of similarity found in the analysis of the five discussion forums. Graphic "Similarity1" corresponds to the degree between the averages obtained from Teacher 1 and from MineraFórum. Graphic "Similarity2" refers to the comparison with Teacher 2.

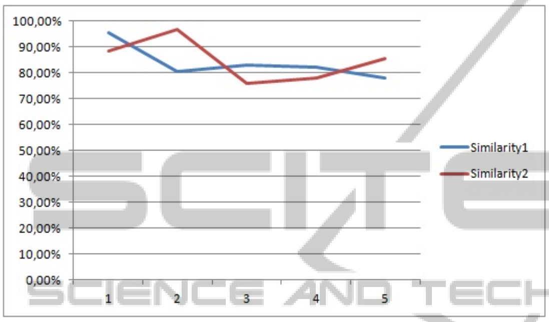


Figure 7: Variation in the degree of similarity in the experiments.

It was found that in some situations, the values of the relevance of each post calculated by MineraFórum were different of more than 2 points regarding to the analysis of teachers. This occurred in the following situations:

a) The message cited relevant concepts to the topic, but not had value in the debate.

These cases occurred in posts where there was no coherence and cohesion in the typed text. The MineraFórum not analyses these parameters. The software calculated the relevance of messages according to the important concepts mentioned. The teacher assigned a low value for these posts.

b) The message mentioned relevant concepts, which were not cited in the reference text or in the forum.

In this situation, the teacher assigned a high value for the post. As there were no conditions for the software to identify these concepts, the calculated relevance was low.

c) The post did not mentioned relevant concepts to discussion, but had an important example of a personal experience.
The tool assigned low importance to these messages. The teacher analysed these posts with high relevance.

d) The post did not cited relevant concepts to discussion, but indicated an important bibliographic reference or site.

The MineraFórum not analyzes the importance of references or sites. Thus, the software calculated low relevance to the post and the teacher indicated a high value.

e) The message did not mentioned relevant concepts to discussion, but attached an important file or image.

The MineraFórum not analyzes the content of files or images. Thus, the software calculated low relevance to the post and the teacher indicated a high value.

## 6 CONCLUSIONS

MineraFórum is a resource aimed at helping teachers in qualitative analyses of forum posts. The software performs the automatic execution of the aforementioned activities. In such situations, the teacher's role in the debates is seen as crucial.

Considering the results presented in section 5, the objective of the experiments was reached. It was possible to verify that the average of the analysis of the messages calculated by the software is similar to the average obtained with appraisals made by teachers.

MineraFórum is able to present the teacher with a picture of the contributions made by learners, by organizing and clustering the posts of the students. This means that the program can provide information that may be helpful to teachers in their tutorial activities.

With information given by MineraFórum, the teacher can guide his/her support to students who posted few relevant contributions in a forum. Teachers can also stimulate interactions between learners who posted more relevant messages and those who contributed with few relevant ones.

## REFERENCES

Azevedo, B. F. T., Behar, P. A., Reategui, E. B., 2011a. Automatic analysis of messages in discussion forums. In *Proceedings of the 14th International Conference on Interactive Collaborative Learning*. Piešťany, IEEE, pp. 76-81.

Azevedo, B. F. T., Behar, P. A., Reategui, E. B., 2011b. Um software para análise das mensagens de fóruns de discussão. In *Proceedings of the IADIS Ibero-Americana WWW/Internet (CIAWI) Conference*. Rio de Janeiro, IADIS Press, pp. 195-202.

Behar, P. A., Bernardi, M., Souza, A. P. F. de Castro, Kellen, K., 2007. ROODA: desenvolvimento, implementação e validação de um AVA para UFRGS. In *XII Taller Internacional de Software Educativo.* Santiago, Chile, LOM Ediciones S.A., v. 1, pp. 321-338.

Feldman, R., Sanger, J., 2007. *The Text Mining Handbook*: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press.

Garrison, D. R., Anderson, T., Archer, W., 2000. Critical inquiry in a text-based environment: computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3), pp. 87-105.

Gilbert, P. K., Dabbagh, N., 2005. How to structure online discussions for meaningful discourse: a case study. *British Journal of Educational Technology*, v. 36, n. 1, pp. 5-18.

Li, Y., Dong, M., Huang, R., 2008. Semantic Organization of Online Discussion Transcripts for Active Collaborative Learning. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies*. IEEE Computer Society Press, pp. 756-760.

Macedo, A. L. et al., 2010. ETC: o que mudou e por quê?. *Revista Novas Tecnologias na Educação*, v. 8, pp. 2-12.

Palloff, R. M., Pratt, K., 2002. *Construindo comunidades de aprendizagem no ciberespaço*, Artmed. Porto Alegre.

Palloff, R. M., Pratt, K., 2004. *O aluno virtual*: um guia para trabalhar com estudantes on-line, Artmed. Porto Alegre.

Puretskiy, A. A., Shutt, G. L., Berry, M. W., 2010. Survey of text visualization techniques. In Berry, M. W., Kogan, J. *Text mining*: applications and theory. John Wiley & Sons Ltd, pp. 107-127.

Ravi, S., Kim, J., 2007. Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers. In *Proceedings of the AI in Education Conference (AIED)*, Los Angeles.

Rebedea, T., Trausan-Matu, S., Chiru, C., 2008. Extraction of Socio-semantic Data from Chat Conversations in Collaborative Learning Communities. Dillenbourg, P., Specht, M. (eds.). *EC-TEL 2008*. LNCS, Springer-Verlag Berlin Heidelberg, n. 5192, pp. 366-377.

Tan, A., 1999. Text Mining: The State of the Art and the Challenges. In *Proceedings of the PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases*, pp.71-76.

Torre, C. J. de la, Martín-Bautista, M. J., Sánchez, D., Vila, M. A., 2005. Text mining: Intermediate forms for knowledge representation. In *Proceedings of the Joint Eusflat/Lfa Conference*. Montseny, E., Sobrevilla, P. (eds.), European Society for Fuzzy Logic and Technology, pp. 1082-1087.

[i] MineraFórum is part of the research projects: "MineraROODA: Ferramentas de mineração de conteúdo cognitivo e de subjetividade afetiva no Ambiente Virtual de Aprendizagem ROODA", Announcement MCT/CNPq 014/2010 - Universal; "ROODA: novas ferramentas para incorporação no ambiente virtual de aprendizagem", Researcher Gaúcho Program, Announcement FAPERGS 006/2010. "Ampliando possibilidades pedagógicas através da tecnologia de mineração de textos integrada à escrita coletiva a distância", Announcement MEC/CAPES 029/2010.

[ii] WordNet is a large lexical database of English; available at: http://wordnet.princeton.edu

[iii] Núcleo de Tecnologia Digital aplicada à Educação / Pós-graduação em Informática na Educação.

[iv] ROODA is one of the platforms used for Distance Education at UFRGS. ROODA is available at: https://www.ead.ufrgs.br/rooda/

[v] ETC is a collective text editor develped by NUTED, and used at UFRGS as well as in extension courses, available at http://www.nuted.ufrgs.br/etc2/

[vi] MOODLE, with MineraFórum integrated into it, is available at http://www.nie.iff.edu.br/moodle/