



---

## MONITORING INTERACTION AND COLLECTIVE TEXT PRODUCTION THROUGH TEXT MINING

---

*Alexandra Lorandi Macedo<sup>1</sup>*

*Patricia Alejandra Behar<sup>2</sup>*

*Breno Fabrício Terra<sup>3</sup>*

**ABSTRACT:** This article presents the Concepts Network tool, developed using text mining technology. The main objective of this tool is to extract and relate terms of greatest incidence from a text and exhibit the results in the form of a graph. The Network was implemented in the Collective Text Editor (CTE) which is an online tool that allows the production of texts in synchronized or non-synchronized forms. This article describes the application of the Network both in texts produced collectively and texts produced in a forum. The purpose of the tool is to offer support to the teacher in managing the high volume of data generated in the process of interaction among students and in the construction of the text. Specifically, the aim is to facilitate the teacher's job by allowing him/her to process data in a shorter time than is currently demanded. The results suggest that the Concepts Network can aid the teacher, as it provides indicators of the quality of the text produced. Moreover, messages posted in forums can be analyzed without their content necessarily having to be pre-read.

**KEYWORDS:** Education development. Technology.

---

## ACOMPANHAMENTO DA INTERAÇÃO E PRODUÇÃO TEXTUAL COLETIVA POR MEIO DE MINERAÇÃO DE TEXTOS

---

**RESUMO:** Este artigo apresenta a ferramenta Rede de Conceitos desenvolvida a partir da tecnologia de mineração de texto. O principal objetivo é extrair e relacionar os termos tratados com maior incidência na produção textual e exibir o resultado através de um grafo. A Rede foi implementada no Editor de Texto Coletivo (ETC) que é uma ferramenta online que permite a produção de textos de forma síncrona ou assíncrona pelos autores. Este artigo descreve a aplicação da Rede tanto em produções textuais coletivas, quanto na ferramenta fórum. O intuito é de oferecer suporte para o professor gerenciar o alto volume de dados oriundos das negociações e construção do texto em um menor tempo que o demandado atualmente. Os resultados apontam que a Rede de Conceitos pode auxiliar o professor, pois possibilita a obtenção de indicadores sobre a qualidade da produção textual. Além disso, ele pode analisar as mensagens postadas nos fóruns, sem necessariamente ler todo o conteúdo deste material anteriormente.

**PALAVRAS-CHAVE:** Desenvolvimento educacional. Tecnologia.

---

<sup>1</sup>Mestre em Educação pela Universidade Federal do Rio Grande do Sul. Doutora e Pós-Doutora em Informática na Educação pela UFRGS. Porto Alegre (RS) – Brasil. E-mail: [alorandimacedo@gmail.com](mailto:alorandimacedo@gmail.com)

<sup>2</sup>Mestre e Doutora em Ciência da Computação pela UFRGS. Professora Associada 4 da Faculdade de Educação e dos Cursos de Pós Graduação em Educação (PPGEdu) e em Informática na Educação (PPGIE) da UFRGS. Possui bolsa de Desenvolvimento Tecnológico e Expansão Inovadora (DT), nível I. Porto Alegre (RS) – Brasil. E-mail: [pbehar@terra.com.br](mailto:pbehar@terra.com.br)

<sup>3</sup>Mestre em Informática pela UFES. Doutor em Informática na Educação pela UFRGS. Porto Alegre (RS) – Brasil. E-mail: [brenoter@gmail.com](mailto:brenoter@gmail.com)

**Recebido em:** 24/01/2014 – **Aceito em:** 20/02/2014.

## ACOMPANÑAMIENTO DE LA INTERACCIÓN Y PRODUCCIÓN TEXTUAL COLECTIVA ATRAVÉS DE LA MINERACIÓN DE TEXTOS

**RESUMEN:** Este paper presenta la Herramienta Red de Conceptos (en portugués Rede de Conceitos). Fue desarrollada a partir de la tecnología de mineración de texto. El principal objetivo es extraer y relacionar los conceptos tratados con mayor incidencia en la producción textual y exhibir el resultado a través de un grafo. La Red fue implementada en el Editor de Texto Colectivo (ETC) que es una herramienta online que permite la producción de textos de forma síncrona o asíncrona por los autores. Este artículo describe una aplicación de la Red tanto en producciones textuales colectivas, quanto en un fórum de discusión. La intención es ofrecer soporte para que el profesor gestione el gran volumen de datos originados de las negociaciones y de la construcción del texto en un tiempo menor que lo demandado normalmente. Los resultados apuntan que la Red de conceptos puede auxiliar al profesor, porque posibilita el logro de Indicadores sobre la calidad de la producción textual. Además de eso, él podrá analizar las mensagens postadas em el fórum, sin necesariamente tener que leer todo el material del contenido anteriormente.

**PALABRAS-CLAVE:** Desarrollo educacional. Tecnología

### **1 COLLECTIVE WRITING: THE RELATIONSHIP BETWEEN THE PROCESS OF INTERACTION AND THE VOLUME OF DATA**

Recent and significant advances in technology have seen the development of activities which favor exchange, communication and negotiation between individuals. Within this context, the process of collective writing in distance learning offers tools that support the exchange of ideas and debate between different points of view. With this in mind, it is understood that students learn in an interactive, collective process which can favor learning which can be considered qualitatively better than that which is developed in individual practice. Ascott (1997, p. 338) suggests that “[...] without interaction, nothing new happens. Without interaction no meaning is generated. Without interaction no experience is created [...]”. In this way, the collective process of creation, construction and exchange is formed. This process calls all participants to act, to interact, to expose themselves and to modify themselves. In this context, the collective is understood as being all individuals in action, including the teacher who monitors and mediates the collective productions.

Group conflicts in digital contexts are usually more complex to resolve than those which occur in a face to face context. The interpretation that can be given for one or other written statement/argument on a certain subject sometimes creates complications. To favor the continuity of the process, it is important to guarantee freedom among the participants. Such freedom includes moments to question, to position oneself and to counter argument new themes and new solutions that can emerge from debates. This favors the continuity of exchanges and the construction of new knowledge. The interaction space offered by communication tools and the editing of text creates a high and permanent volume of data and



information hindering the job of the teacher in monitoring the students' progress. It is in this context that the Collective Text Editor (CTE) is used.

The management of data on the part of the teacher requires precious time which could otherwise be spent on pedagogical interventions in the collective process with a view to advancing and improving students' work. Thus, the implementation of the Concepts Network in the CTE seeks to support the analysis and management of data in a shorter period of time than was previously demanded. In order to achieve this aim, Text Mining resources were employed as a means of offering conditions to meet the needs identified in this study.

## **2 TEXT MINING**

According to Feldman and Sanger (2007), text mining can be defined as an intensive process in which a user interacts with a large quantity of documents using tools to analyze them. The objective of the process is to extract useful information from collections of documents. The useful information is identified in interesting patterns in data taken from unstructured texts.

The systems of text mining are based on preprocessing routines, algorithms to discover the patterns and elements to present the results. In these systems, the preprocessing operations are based on the identification and extraction of representative characteristics from documents in natural language. These operations are responsible for transforming unstructured data, stored in collections of documents, into a structure expressed in an intermediate format (FELDMAN, SANGER, 2007; HEARST, 1999; TAN, 1999).

Text mining exploits techniques and methods taken from areas such as the recuperation of information, the extraction of information and computational corpus linguistics. In order to extract useful information, relevant characteristics are found in documents, the most used being characters, words, terms and concepts. Hybrid approaches can be performed to generate representations of documents based on characteristics (FELDMAN, SANGER, 2007). For example, it is possible to extract the terms from documents, then adapt or normalize them, comparing them to a list of relevant topics (concepts) extracted via a categorization.

Technology employed in text mining includes: information extraction, topic tracking, summary creation, text categorization, text grouping, concept linking, information visualization and question and answer analyses (FAN et al, 2006).

A very common technique employed is the representation of characteristics of a document through a vector space model. In this technique, every term of the document becomes a dimensional characteristic. The value of every dimension might indicate the number of times that a specific term appears in the text, or may indicate the importance of a word being considered. This could be the quantity of documents in which the term appears for instance. However, this technique discards useful information, such as the order in which the terms appear, where they appear and the proximity among them (SCHENKER, 2003).

Graphs are important mathematical constructions in representing the modeling of relations and structural information. Graphs are used in many kinds of problems including: ordination, compression, traffic analysis and resource allocation, etc. As graphs retain more information than vectors of simple atomic characteristics, they represent precious modeling and can be used in the representation of texts (SCHENKER, 2003)

The technique of text mining using graphs highlights words with the greatest incidence in a text and identifies whether they are related. The graph obtained by mining presents the words which occur most in the form of knots. The associations between the knots of the graph indicate the proximity between the words.

This software, Sobek (LORENZATTI, 2007) allows the construction of a graph using the statistical information obtained from a text, where vertices and edges present information related to the absolute and relative occurrences of terms (vertices) and associations (edges) in a specific document. The graph obtained from the document represents a concepts network of words generated from the text.

The first step of processing in the Sobek software involves lexical analysis, where the text is dismantled word by word. The words are subsequently submitted for statistical analysis. At this point, a base of concepts is created using the statistical data. Concepts may be composed of one or more words. The next step consists of removing words which do not add meaning to the text such as: articles, conjugations of certain verbs and pronouns.

After removing the words and completing the statistical analysis, a graph is created. At this stage, parameters such as absolute frequency and the size of the concept (if simple or compound) are analyzed. The Sobek interface allows the selection of the size of the graph desired, the options being small, medium, large and very large. The size of the graphs corresponds to the configuration profiles, as these graphs contain the parameters used in the text mining. These parameters determine the minimum number of occurrences that a term must present in order to be integrated into the graph.

It is worth noting that in Sobek, the vertices of the graph can have more than one word. In this way, a rule was developed to select the concepts and, consequently, the vertices. At the moment of selection, the similarities between the concepts are evaluated as well as the size of them (number of words) and the absolute frequency. For this reason, the calculation is worked out between similar concepts (composed of the same word), prioritizing the biggest, which express more comprehensive ideas (LORENZATTI, 2007). One should note that Sobek was used in this study as the basis for the development of the Concepts Network.

### **3 CONCEPTS NETWORK**

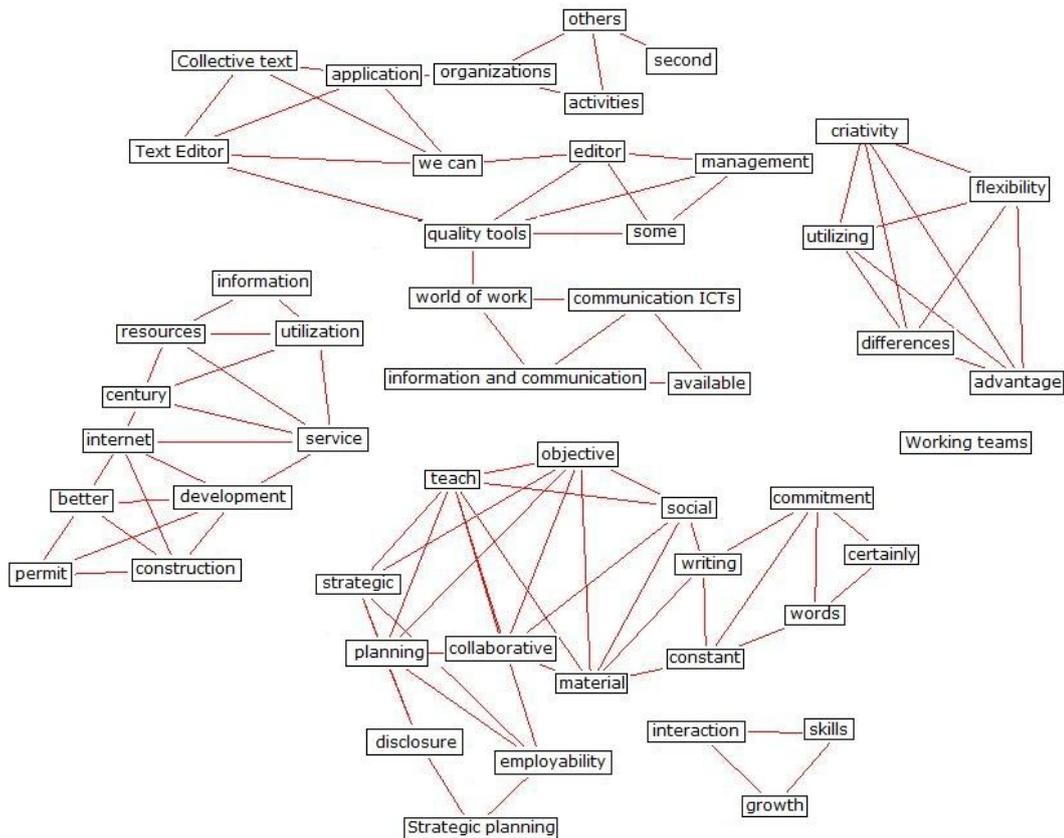
The Concepts Network – a tool developed based on text mining technology – aims to extract and interlink the main terms present in a text. The intention is for the result obtained to be sufficient to allow the user to have an idea of the theme and the developments in production, without a prior reading of the text.

The Network was implemented using Sobek software, presented in the previous section, and was integrated with the CTE, allowing the developed texts to be easily mined. Its application is recommended for all those who wish to use the Editor and want to visualize and understand the perspective of a text, without necessarily having to read it first in its entirety.

### **4 THE APPLICATION OF THE CONCEPTS NETWORK IN COLLECTIVE TEXTS**

Macedo *et al* (2009) and Macedo (2010), propose the use of the Concepts Network in the Collective Text Editor (CTE) available at <http://www.nuted.ufrgs.br/etc2> in order to support the work of the teacher. The use of the Concepts Network allows problems related to cohesion and coherence (among others) to be identified via the viewing of graphs.

The Concepts Network highlighted in Figure 1 was extracted from a text produced in an extension course and presents the concepts that appear with the highest frequency.



**FIGURE 1** – Concepts Network Text 3 - Group: Teamwork  
Source: Macedo, 2010, p. 114.

Even without having read the original text, one is able to see that the Concepts Network presents concepts such as: teamwork, collaborative, competence and digital technologies. This observation can be made by highlighting merely ten of the terms displayed in the Network of Figure 1: teamwork, collaborative, social, interaction, competences, organizations, flexibility, ICT communication, resources and internet.

Extracts from the text are presented in the table below in an attempt to relate the result of this Concepts Network with the text from which it originated. The extracts presented support of the argument made in this article. These extracts correspond, at least in part, to the terms highlighted above. The aim is to show the issue that was addressed. The extracts in Table 1 are arranged in paragraphs, which are not in all cases complete. Each row in the box below corresponds to a paragraph. The order in which the paragraphs are presented below is not necessarily the order in which they are found in the original text.

**TABLE 1** – Excerpts from Texts

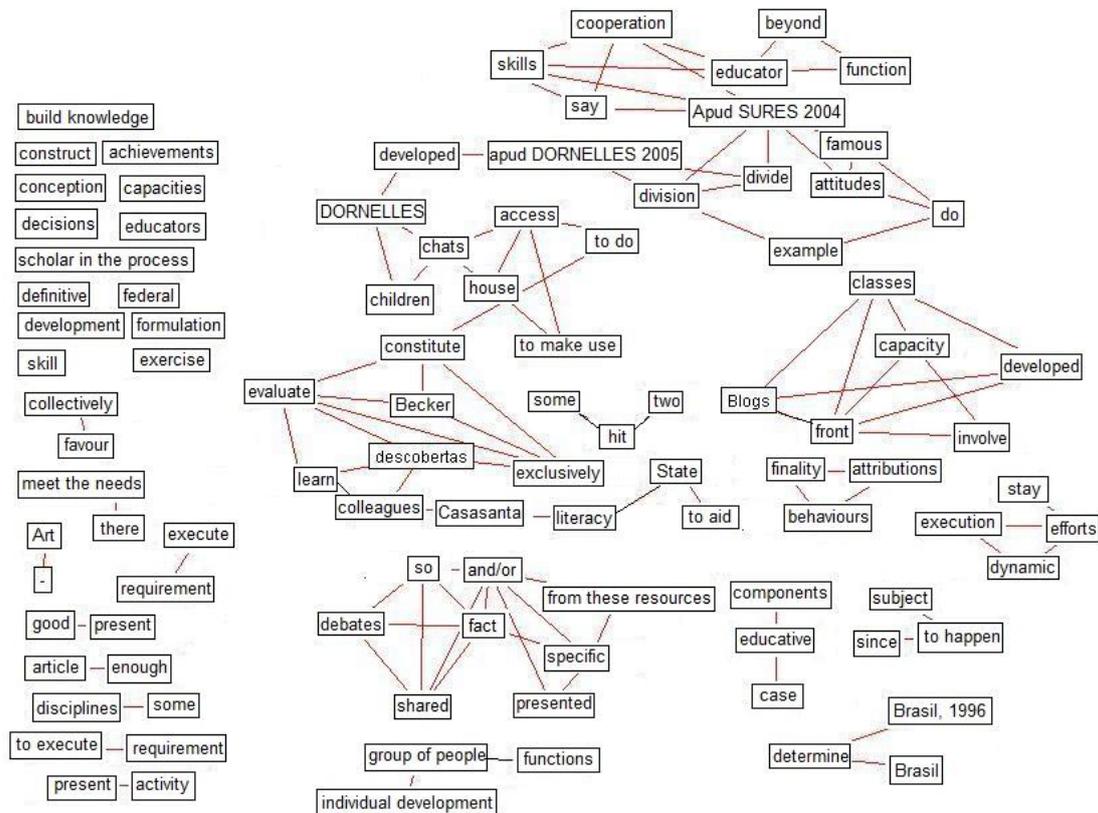
<p>Companies need employees who can work in teams, using creativity, flexibility and cooperation in search of common goals. We focus, in the first item, on the differences between groups and teams and how to succeed by working together, respecting the individuality of each group member.</p>
<p>In today's world where information is produced and consumed in a virtual form, in large quantities and with global coverage, the intense use of resources that bring people together, disseminate ideas and save time and energy are fundamental. New Information and Communication Technology (ICT), management tools and new concepts of collaborative work - in real or virtual teams, have already become a reality in this new century.</p>
<p>In the new digital era, the increasingly widespread usage of Information and Communication Technology (ICT) has meant that asynchronous production of group work, that is work realized in different times and places, can be carried out more quickly and with more objectivity. Flexibility now exists in the production of collaborative texts without compromising involvement, interaction and creativity. This flexibility and speed provides a foundation in the search for greater competitiveness.</p>

Source: Macedo, 2010, p. 115.

The extracts of the text allow us to understand the context surrounding the written production and refer to the idea of the terms initially highlighted as indicators of the central thematic axis.

It should be emphasized that the analysis in full, both of the Concepts Network and of the text, provides better tools for analysis and conclusions. In this study, for the purpose of organization and objectivity, the fragments sought were those indicative of the possibilities found from the reading and general interpretation of the collected elements.

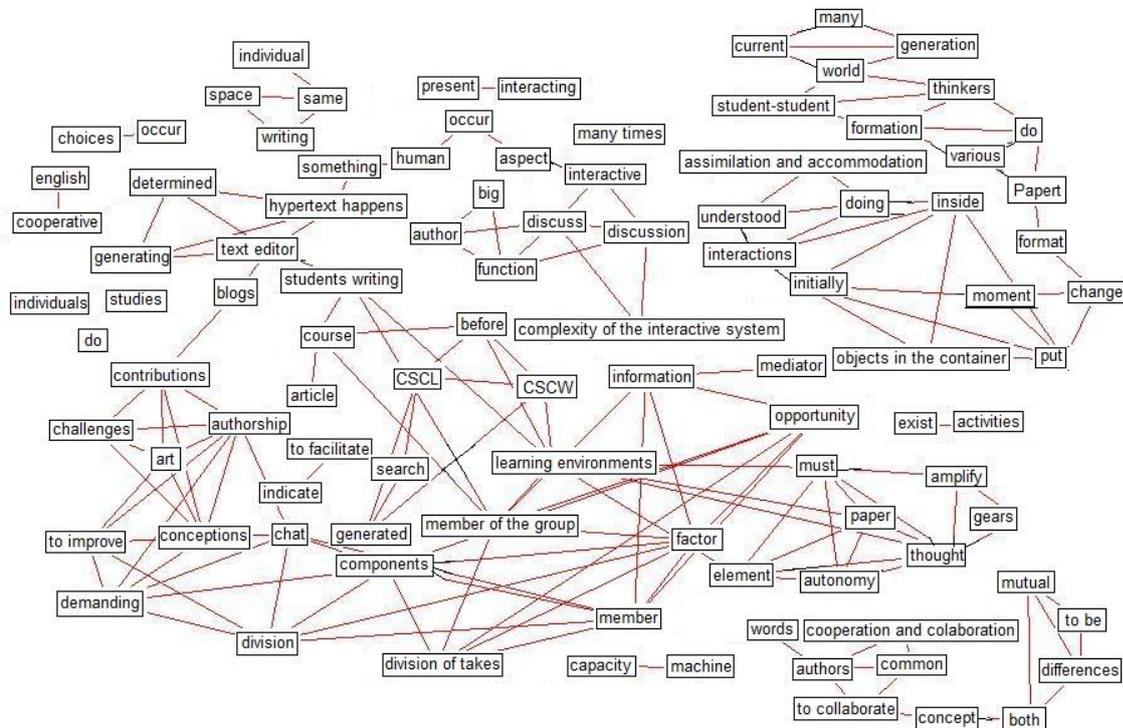
In addition to identifying or not the theme, it was noted that the structure of the Network is directly related to the quality of the produced text. That is, the Networks “expose” texts in need of further theoretical deepening or with a juxtaposition of paragraphs. An example of the need for improvement can be seen in the Network shown in Figure 2.



**FIGURE 2 – Concepts Network Text 4 – Group: Teamwork**  
**Source:** Macedo, 2010, p. 122.

In this Concepts Network, it is possible to note the incidence of several loose terms. Loose terms are those not related to any other in the Network and which alone do not allow the context in which the term was used to be understood. Furthermore, the incidence of several small groups of terms was identified. In this case, it is clear that the identification of both the central theme and its possible developments have been compromised. The isolated concepts are rather vague indicators because they don't allow one to know the context of the text. A reading of the original text was carried out following the analysis of the Network in question. Here it was verified that the innumerable concepts that are loose in the Network do in fact reveal a text that can also be considered 'loose' in essence. The text can be considered loose as it cites several different topics but doesn't deal with any with real ownership, nor does it relate the topics in the writing. It is a text that is in need of improvement in relation to coherence and cohesion. It is a text in need of enhancement.

On the other hand, the analysis of the Networks also covered those Networks which presented an incidence of interconnected terms, as represented in Figure 3.



**FIGURE 3** – Concepts Network text 5 – Virtual Learning Workshops  
**Source:** Macedo, 2010, p. 133.

In this case, it is noteworthy that sets of inter-related concepts are prevalent, in contrast to the incidence of isolated terms in the previous example. The interpretation of the Concepts Network in this situation indicates a text that deals with collective distance writing. It is significant that, in general, the concepts complement each other, forming a thematic unit. For instance: text editor, learning environments, blog, chat, division of labor, cooperation and collaboration, complexity of the interactive system, discussion, contributions, authorship and challenges, are some of the concepts which, while not being entirely interconnected by edges, complement each other in the indication of the central theme of the text produced.

Following the completion of the analysis of the Concepts Network, our attention turned to the reading of the text, at which stage it was noted that the terms which appeared in the Concepts Network also appeared in the text. The first section of the text dealt with the conceptualization of the terms CSCW, CSCL, cooperation and collaboration; together with the term interaction in the context of learning and hypertext. Next the concept of computational collective writing was introduced and was followed by a focus on the use of

the CTE (Collective Text Editor) in relation to the actions of the group. The text under analysis ended with considerations on the use of technology as a means to support learning. Significantly, this text differs from the text in the previous example in terms of structure. Here, the flow of the text presents linking, sound development, a logical approach and good argument.

Based on the application and analysis of the Network, generated from the texts produced in the CTE, it is clear that the Network was sufficiently effective in presenting the theme discussed in the text written on the given theme, since the aim was to understand what the text deals with, without a pre-reading before generating the Network. Moreover, it can be said that the Concepts Network provides indicators which relate to the text, allowing the teacher to focus on the needs and potential of the students, without having to read the text produced by the student. This was the principal motivation behind this study.

## **5 THE APPLICATION OF THE CONCEPTS NETWORK IN DISCUSSION FORUMS**

Azevedo *et al* (2009, 2010) present another experience in which the Concepts Network has been applied. Azevedo *et al* (2009, 2010) proposed the use of the Network to mine discussion forums in virtual learning environments. The objective was to identify whether the discussions in the forum were relevant and if there was connection or integration among the concepts related to the subject of debate.

In any one discussion forum there can be hundreds of contributions, depending on the proposed theme, the profile of the participants and the interest in the subject. Some written texts are relevant to the theme, and others are not. Significant contributions address important concepts related to the subject under discussion.

Considerable time is needed for the teacher to be able to monitor carefully all contributions made by students in a discussion forum. Thus, it is important to provide a way of visualizing the relevant contributions recorded in the forum, so as to assist the teacher in interventions with the students involved.

The use of the Concepts Network to analyze discussion forums is currently under development; therefore, this article presents the results of a pilot study. The same principle of extraction, employed in the Concept Networks, was used in mining the forum texts. Thus, a manual study of the networks generated was carried out so as to analyze which contributions were relevant to the proposed theme.

In order to make this activity viable, the thematic relevance coefficient - TRC (Azevedo *et al*, 2009) - of a written contribution was defined. The objective of the TRC is to determine how relevant a written contribution is to a particular topic of discussion. The relevance coefficient indicates the degree of relevance of the written contribution according to the theme of the forum and is calculated with the use of the following formula:

- NC: number of relevant concepts used in the contribution
- NA: number of associations between the relevant concepts used in the contribution

$$TRC = NC + NA$$

For the calculation of NC, the terms regarded as semantic equivalents are also considered relevant concepts. Prefixes, suffixes, and plurals of the relevant concepts cited in the written contribution are converted so as to also be considered as relevant concepts. With the TRC, the networks that have more important concepts related to the topic, with more associations between the concepts, have a higher coefficient of thematic relevance.

The methodology employed involved:

- a) the choice of a discussion forum.
- b) the definition of the important concepts related to the theme of the forum, as well as the definition of the association between these concepts. The concepts can be entered and related by the teacher, or a reference text can be indicated by the teacher. The concepts and their relationships are extracted from the text by the tool.
- c) the definition of possible terms that can be considered semantic equivalents to the concepts involved.
- d) the definition of the minimum value of the thematic relevance coefficient to be considered for analysis.
- e) the collection of written contributions produced by students in the forum.
- f) the generation of a Concepts Network for each written contribution.
- g) the generation of the thematic relevance coefficient of the written contribution using each Concepts Network.
- h) the relevant written contributions were arranged according to the networks generated.
- i) an analysis of the number of relevant contributions made by each student in the forum using the generated Networks.

All messages posted in the forum are considered. Each message is analyzed individually and a Concepts Network is built from the written text. In each Network, a calculation of the thematic relevance coefficient of the written contribution is made. For example, in a forum involving 20 students where each student writes 10 posts, 200 Concepts Networks will be generated and evaluated.

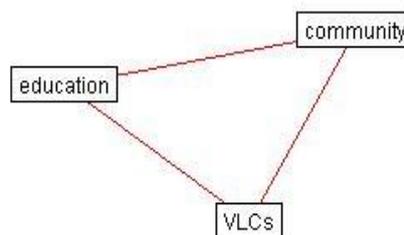
The first study was conducted in a discussion forum in the ROODA Environment, in the *Special Topics Z1* course of the Computing in Education PhD, during the first semester of 2008. The theme of the forum was *Virtual Communities*. Five students participated in the forum with a total of 25 texts being contributed.

The teacher defined the following concepts as important for this particular forum: *communities, virtual, virtual communities, learning, virtual learning communities, education* and *social relations*. The relevant associations between the concepts were established as follows: the concept of *virtual learning communities* is related to the concepts *communities, virtual communities, learning* and *education* respectively. The terms *VCs* and *VLCs* were defined as semantic equivalents to the terms *Virtual Communities* and *Virtual Learning Communities*.

A minimum value of 1 was defined as the thematic relevance coefficient (TRC) to be considered in the analysis. The thematic relevance coefficient was calculated using the Concepts Network generated from each written contribution. An example of the calculation of the TRC can be seen in the message below, extracted from the forum and originally posted by student C.

Message: *I believe that the VLCs can provide a valuable resource for education, as they provide a collective knowledge construction mechanism, through discussions between community members.*

The network generated from the above message can be seen in Figure 4.



**FIGURE 4** – Network generated from an experimental forum message  
**Source:** Azevedo, 2009, p. 80.

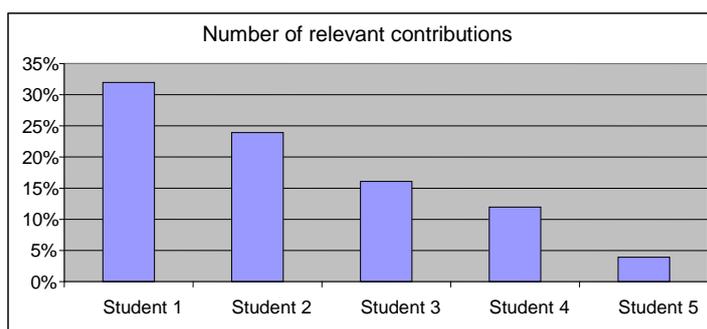
For the above Network, the TRC is calculated as follows:

- $NC = 3$

- $NA = 3$

$$CRT = NC + NA = 6$$

An analysis of the amount of relevant contributions provided by each student in the forum was carried out, the results being presented in the graph in Figure 5.



**FIGURE 5** – Number of relevant contributions in the first experiment  
**Source:** the author.

By using the proposed methodology, it was noted that in Figure 6, it was possible to verify that most of the contributions made by students were relevant to the topic under discussion. This study considers relevant contributions as being those which cite concepts related to the topic of discussion and whose TRC is higher than the minimum established by the teacher.

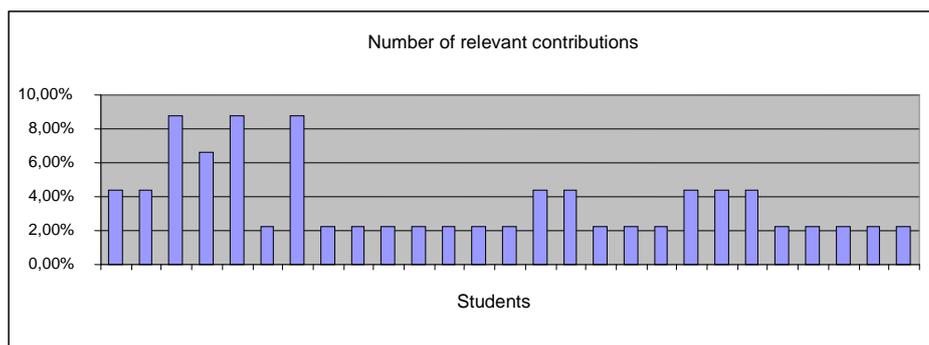
At the time that the experiments were performed, a qualitative analysis of the messages posted in the forums was guided by the calculation of the TRC of each message. Software for text mining is currently being developed to automatically analyze messages in discussion forums. This software enhances the formula of the TRC to also consider the weight of each concept. In the experiments presented in this article, the calculation of the contribution of the TRC of each textual contribution was performed manually. In the software that is under development, the calculation is performed automatically by the software.

The second and third experiments were conducted in a discussion forum of the ROODA Environment, in the *Integrator Seminar VII* discipline – B in an undergraduate class of Education, which occurred in the second semester of 2009. The theme of the forum was *Learning with others*. The class was divided into two groups and each group contributed to a

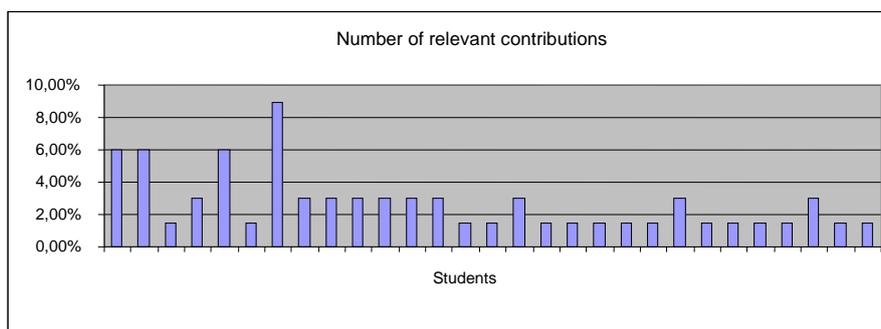
discussion forum. The first forum was utilized by 27 pupils, who provided 45 written contributions. The second forum was utilized by 29 students, who provided 67 contributions.

In these forums, the teacher defined the following concepts as important: *learn, learning, knowledge, knowledge construction, Piaget, affection, affectivity, cognitive structures, decentralization, affective relationships, affective experiences, affective bonding, cooperation and collaboration*. Meanwhile, the relevant associations between these concepts were also established.

A minimum value of 1 was defined as the thematic relevance coefficient (TRC) to be considered in the analysis. The thematic relevance coefficient was calculated using the Concepts Network generated from each written contribution. An analysis was carried out of the amount of relevant contributions provided by each student in the first and second groups. The results are shown in Figures 6 and 7, respectively.



**FIGURE 6** – Number of relevant contributions in the first group  
**Source:** the author.



**FIGURE 7** - Number of relevant contributions in the second group.  
**Source:** the author

The number of contributions relevant to the theme of the forum produced by the students can be verified by using the results presented in Figures 5, 6 and 7. It is interesting to



note that the method used helps the teacher to identify the quantity of messages posted by each student, in which concepts related to the topic under discussion were cited.

It was observed that the current version of the Concepts Network, used in pilot experiments, carried out perfectly the mining of a text and the generation of the graph representing the concepts found. However, the other activities in the experiments were performed manually.

To enable the Concepts Network to offer the appropriate resources to the teacher, it is necessary to improve it and create new features to allow it to be able to:

- a) perform the *stemming* of words.
- b) allow the reading of a set of concepts that have semantic equivalence.
- c) allow the identification of questions asked by students.
- d) allow the automatic reading of texts provided by a discussion forum.
- e) allow the definition of the minimum value of the thematic relevance coefficient subjects to be considered for analysis.
- f) perform the calculation of the thematic relevance coefficient for each graph generated using the texts produced by students.
- g) perform an analysis of the amount of relevant contributions made by each student in the forum using the graph generated.
- h) generate a visual report for the teacher containing information about the relevant contributions from the topic of discussion. This information, made available to the teacher, will indicate the number of relevant and non-relevant contributions made by each student

Based on the results of this study, it was observed that the presented methodology can help the teacher to carry out diverse activities related to the forum, such as:

- analyzing textual contributions which need intervention
- identifying the students who provide only a few relevant contributions in the forum to offer them more help
- encouraging students who have provided many relevant contributions to interact with students who have provided few

From the activities cited, the teacher can direct his/her attention to the students who made few relevant contributions. The teacher can also stimulate a collaborative learning

process, facilitating interaction between those students who produced the most number of relevant texts and the students who wrote few.

## **6 FINAL CONSIDERATION**

With regards to the application of text mining resources in collective text production, it is significant that the Concepts Network provides information about the theme of the textual production and qualitative indicators of production. The qualitative indicators differentiate texts requiring improvement from those which were developed with clarity and objectivity. These situations can be identified through the Network, without a prior reading of the original text. Under these conditions, the time that was once devoted to reading this material can now be spent on intervention with students using various teaching practices, with the aim of enhancing the learning process and the written production itself.

In relation to the discussions performed in discussion forums, it was observed that the Concepts Network can be used in the creation of indicators of the relevant posts of a forum, thus establishing links to the texts developed by the students and, in turn, verifying the relevance and the development and construction process of each production. Significantly all this can be done in a shorter time than is currently demanded because of the resources made available by the technique of text mining.

It is important to note that the Concepts Network is being improved to be adapted to different Virtual Learning Environments and even to work with a variety of formats of text editors and discussion forums. Another interesting step which is in the process of analysis is the implementation of the Concepts Network as an API that can be adapted to VLEs with various technologies such as PHP and Python.

## **REFERENCES**

ASCOTT, Roy. Cultivando o hipercórtex. In: DOMINGUES, D. (Org.). **A arte no século XXI: a humanização das tecnologias**. São Paulo: Fundação Ed. da UNESP. p. 336-344, 1997.

AZEVEDO, Breno Fabrício Terra, et al. (2009). Estudo de análise qualitativa em fórum de discussão. **RENOTE: Revista Novas Tecnologias na Educação**, Porto Alegre, v. 7, n. 2, p. 76-86, 2009.

AZEVEDO, Breno Fabrício Terra; BEHAR, Patricia Alejandra; REATEGUI, Eliseo Berni. Qualitative Analysis of Discussion Forums. In: IADIS MULTI CONFERENCE ON

COMPUTER SCIENCE AND INFORMATION SYSTEMS, 2010, Freiburg. **Proceedings of...** Freiburg: IADIS, 2010. p. 251-258. v.1.

FAN, Weigo, et al. Tapping the power of text mining. **Communications of ACM**. Nova York, v. 9, n. 49, p. 76-82, 2006.

FELDMAN, Ronen; SANGER, James. **Text Mining Handbook**. Cambridge: University of Cambridge, 2007.

HEARST, Marti. Untangling Text Data. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 37., Maryland. **Proceedings of...** Stroudsburg, PA: Association for Computational Linguistics, 1999.

LORENZATTI, Alexandre. **SOBEK**: uma ferramenta de mineração de textos. 2007. Trabalho de Conclusão – Departamento de Informática, Universidade de Caxias do Sul, Caxias do Sul.

MACEDO, Alexandra Lorandi. **Rede de conceitos**: uma ferramenta para contribuir com a prática pedagógica no acompanhamento da produção textual coletiva. 2010. Tese (Doutorado) – Programa de Pós-Graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre.

MACEDO, Alexandra Lorandi, et al. **Using text-mining to support the evaluation of texts produced collaboratively**. In: TATNALL, Arthur.; JONES, Anthony. (Org.). Education and technology for a better world. In: IFIP TC3 WORLD CONFERENCE ON COMPUTERS IN EDUCATION, 9., Berlin. **Selected papers of the...** Berlin: Springer, 2009, p. 368-377.

SCHENKER, Adam **Graph-theoretic techniques for web content mining**. 2003. Tesis (PhD in Computation Science) – University of South Florida, Florida, United States.

TAN, Ah-Hwee. Text mining: the state of the art and the challenges. In: PACIFIC-ASIA WORKSHOP ON KNOWLEDGE DISCOVERY FROM ADVANCED DATABASES, 1999, , Beijing . **Proceedings of...**[S.l.]: Springer, 1999, p. 65-70.

### **Como citar este documento:**

MACEDO, Alexandra Lorandi; BEHAR, Patricia Alejandra; AZEVEDO, Breno Fabrício Terra. Monitoring interaction and collective text production through text mining. **ETD – Educ. temat. digit.**, Campinas, SP, v. 16, n. 1, p. 67-83, jan./abr. 2014. ISSN 1676-2592. Disponível em: <http://www.fae.unicamp.br/revista/index.php/etd/article/view/5730>. Acesso em: 30 Abr. 2014.

