CONCEPTS NETWORK

Text Mining Technology allied with Pedagogical Practices to Qualify the Collective writing Process in Distance Learning

Alexandra Lorandi Macedo, Patricia Alejandra Behar and Eliseo Berni Reategui
PPGEDU/PPGIE - UFRGS, Av. Paulo Gama, 110, 90040-060, Porto Alegre, RS, Brazil

Keywords: Collective Text Editor, Text Mining, Pedagogical Practices.

Abstract: The main goal of this article is to identify possibilities of pedagogical practices from the information generated by a tool called the Concepts Network. Such information is extracted from texts produced by participants / students in a Collective Text Editor - ETC. Thus, this study presents a tool and, from it, alternative teaching practices. The intention is to create conditions so that such practices are developed with quality and minimize the large work load required in digital work spaces in distance learning practices.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

This article presents the development of a study that had as its objective the construction of a tool which allows teachers to monitor the process of collective construction of texts in the Collective Text Editor - ETC (available at: www.nuted.ufrgs.br/etc2). The ETC is an online editor, which was developed with an educational purpose by the Centre of Digital Technology applied to Education at the Federal University of Rio Grande do Sul (NUTED / UFRGS). The first version was built in 2001 and the system has since been applied and perfected, boasting, as a reference, validations and evaluations made by teachers and student users. Some of the principal adjustments and implementations developed during its application can be checked in Macedo et al. (2009). In this period, it was also noted that teachers frequently reported the difficulty they had in following the process of students' written construction, due to the high volume of data generated in this practice.

From this context began the development of this study, which developed a tool to support the monitoring of collective text production in the ETC. Therefore, the next sections describe the theoretical foundation that supported the educational perspective in the use of the tool. The developed tool is also presented along with the validation process, the possible applications from the results obtained and the final considerations.

2 THEORETICAL PERSPECTIVE TO SUPPORT EDUCATIONAL PRACTICE

Advances in science and technology are ongoing and take place in an ever more dynamic form. In this context begins the teacher's challenge, which is to develop strategies that enable the formation of a subject that takes account of this complexity. In order to achieve this, on the basis of studied theory, the following assumptions were elected in this research: the construction of autonomy, critical thinking, the process of learning to learn and knowing how to articulate oneself in collective contexts, cooperating and communicating with all involved in the process.

The process of the student's personal development involves, among other elements, the development of autonomy. It is this that may offer conditions to allow the subject to realize the complexity and the challenges they encounter throughout the learning process.

For Piaget (1973), autonomy is directly related to the type of respect that the subject has for rules or standards. According to the author, there are two kinds of respect: heteronomy and autonomy.

For Piaget (1994), only cooperation leads to autonomy. Thus, it is necessary to favor cooperation in the educational context, not only restricting the social exchange between teacher and student, but also encouraging the exchange among peers as "[...]

Lorandi Macedo A., Alejandra Behar P. and Berni Reategui E..
CONCEPTS NETWORK - Text Mining Technology allied with Pedagogical Practices to Qualify the Collective writing Process in Distance Learning.
DOI: 10.5220/0003801500710076
In *Proceedings of the 4th International Conference on Computer Supported Education* (CSEDU-2012), pages 71-76
ISBN: 978-989-8565-07-5
Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

criticism is born out of discussion and discussion is only possible among equals, therefore, only cooperation will realize what intellectual constraint is incapable of realizing" (Piaget, 1994, p. 298-299).

Based on these conditions, it is understood that to achieve the highest levels of autonomy development, it is necessary that, throughout the educational process, teachers and students take on distinct actions in which the former must create for the latter, ever more challenging situations for resolution and development, offsetting practices which are limited to reproduce or copy. The intention is for the students to be able to support their actions within the principles of critical thinking about their own process. In this study, critical thinking is considered as a sense of experience and need for logical consistency that are placed in the service of an autonomous reasoning, common to all individuals and not depending on any external authority (Piaget, 1998). Thus, social exchange acts as a tool that encourages creativity and critical spirit. The construction of critical spirit is fundamental to the intellectual and social development of the individual (Piaget, 1986). It is from the social relations which follow the cognitive conflict that subjects are enticed to question, doubt and criticize the different points of view and from there are motivated to propose solutions and alternatives that are shown as more viable for conflict resolution.

This set of elements shows the complexity of students' reflection, the complexity that is found in a critical thought (Parrat-Dayan, 2007). Thus, Piaget's perspective (1986) indicates that pedagogical practice must insist on the exchange of different points of view in order to provide mutual enrichment between subjects. Such a perspective consists of leading each individual to think for themselves and position themselves relative to another.

Different challenges can make the subject feel the need to seek elements, information that will help him/her to realize the given situation. The measure to which the student can realize their limitations, procure and articulate new information to enable him/her to overcome the challenges posed is that we can talk, beyond autonomy and critical thinking, of learning to learn.

Piaget's perspective (1973) on learning focuses on the reflexive action of each individual with the world and exchanges between individuals. For the author, real learning is that which generates knowledge.

It is noteworthy that, on a permanent basis, inter-individual exchanges touch on the basic assumptions of pedagogical practice highlighted in this study. In this sense, each subject is responsible for their production and learning and may ultimately contribute to the learning and production of others, since the interaction can enable favorable conditions for this process. Hence the need for pedagogical practice to consider and contemplate actions that allow the articulation of the subject in collective contexts.

This study understands the articulation of the subject in collective contexts from the social exchanges dealt with by Piaget (1973).

3 THE CONCEPTS NETWORK TOOL

It is noteworthy that the focus of the problem identified by teachers who utilize collective text practices is the amount of time required to monitor the high volume of data generated in the writing construction process. It was at this point, from numerous analyses and investigations that we chose Text Mining technology for the development of the tool, here called the Concepts Network, in order to meet the demands posed.

When based on statistical methods, Text Mining (Feldman and Sanger, 2006) depends on the frequency with which the terms appear in the texts.

To generate the Concepts Network, the first processing step comprises the lexical analysis, where the produced text is broken down word by word. Following this, all extracted concepts are subjected to statistical analysis. At this point, based on the statistical data, a base of concepts is created, which will then assist in building the Network. In the next step, the system removes words that do not add meaning to the text, such as articles, conjunctions of the verb to be and have, as well as pronouns. This process was based on the method used by Schenker (2003).

The representation of information extracted from the texts requires specific data structures such as Vector Space Model (VSM). The VSM-based representation is basically a list of keywords commonly used in Information Retrieval - IR – (Greengrass, 2001). In the process of indexing, each document is represented by a list of keywords. When the user submits a query to the system, it is converted also into a list of terms. Both lists subsequently undergo a process of comparison, using the scaling method (Russel and Norving, 2003). Thus, the search may return documents in order of relevance and of similarity. However, the VSM has, in its own model, some undesirable

characteristics. One such characteristic relates to the way in which words are stored, as it prevents one from knowing the order in which they appear in the text and their relation to the context (Greengrass, 2001). Given these considerations, an alternative approach is highlighted that permits the organization of words extracted from the text and the relationship between them – graphs.

Graphs form part of a line of research known as 'Graph Theory'. One of the most famous problems involving the use of graphs is the problem of the bridges of Königsberg, in the XVIII century, and formulated by Swiss mathematician Leonhard Euler (Berry and Linoff, 1997). In short, graphs are abstractions created to represent relationships. In essence, the graphs are comprised of two distinct parts, as described below:

Vertices (singular Vertex) – contain information which generally represent the inter-related points;

Edges – represent the relationship between two vertices, so two vertices that are in some way related will be connected by an edge.

The process of implementing and testing the tool in the Collective Text Editor was initiated using the results generated by the Concepts Network. It was then that, with the objective of meeting the demands identified in the study, the first contact between the results of the Network and the original texts were delineated. The next step involved the application of the tool.

At the start of the analysis, without the original texts having been read, the Networks which showed a similarity in the presentation of terms were grouped. They were separated into two groups: one with a higher incidence of loose terms and the other with a higher incidence of related /connected terms. The analysis showed that the first group is made up of texts that need improvement, the absence of linkage coming from the productions lacking in logical sequence. Meanwhile, the second group, with an incidence of terms related to each other, is made up of texts presenting coherence, sequence and logic in the development.

In parallel to the applications described, one in particular (Klemann, et al, 2011) was developed in order to ascertain the degree of precision in the correspondence between the text and the result presented by the Concepts Network. In this experiment, students were challenged to construct a production based on a text made available on the curricular subject proposed in class. Subsequently, the text base was mined, and from there, a comparative analysis was done between the concepts extracted by the Concepts Network and the concepts from the reference text used by the students in their own productions. In this process, it was noted that the texts produced by students contained 61.6% of the words highlighted by the Concepts Network. This statistic shows that the tool was able to emphasize a considerable number of relevant terms from the base text.

Once in possession of the collected data, pedagogical practice notes were created with the objective of offering support to the teacher who uses the tool.

4 EDUCATIONAL APPLICATIONS BASED ON THE RESULTS OF THE CONCEPTS NETWORK

This section aims to establish relationships between the results obtained with the Concepts Network tool and the theoretical perspective chosen in this study.

4.1 Focus on the Subject Studied

The results returned by the Concepts Network showed that, depending on the written structure of the text, if well written or not (presence or lack of cohesion and coherence), it is possible to identify the theme of production even without having read the text before.

It is worth highlighting that based on the interpretation of data, it was noted that the result of the Network is closely related to the structure and content developed in the text. Thus, if the Network does not give indications of the subject, this may be an indication that the text produced did not focus on the proposed subject to be developed. Faced with this observation, care should be given in relation to the exchanges. It is known that the Concepts Network is the result of social changes in the Collective Text Editor during the construction of a text.

When there is an imbalance in this exchange, it can also be directly related to the type of respect that students have regarding the rules constructed in the group. Thus, if in the construction of the text, the exchange between participants is supported by a type of heteronomous respect, where obedience governs the development of the writing, the chances that the group cannot adjust to the perspectives of the participants regarding the clarity and objectivity of the production increase and may compromise the quality.

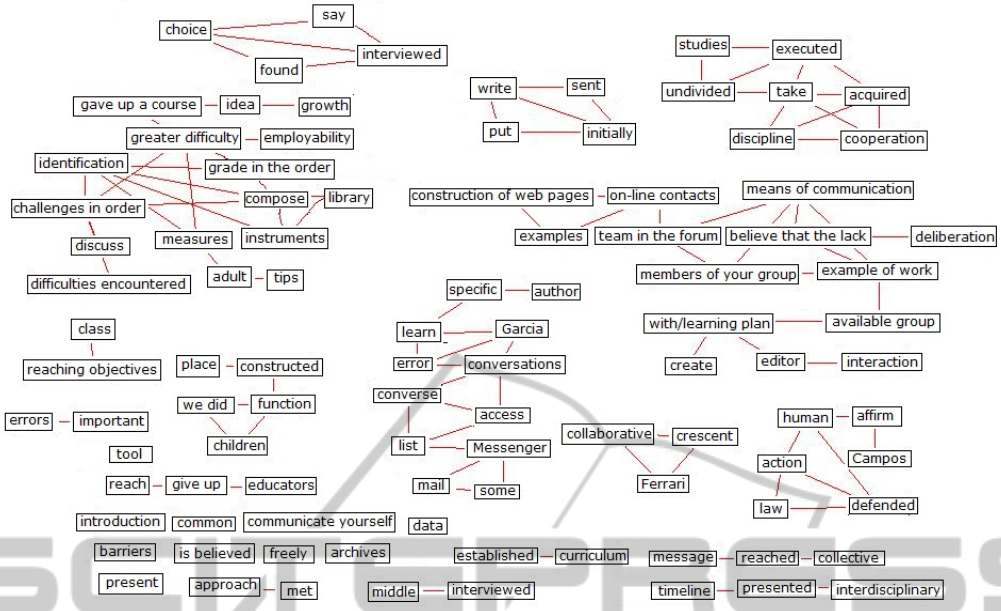

Figure 1: Concepts Network with incidence of loose terms.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

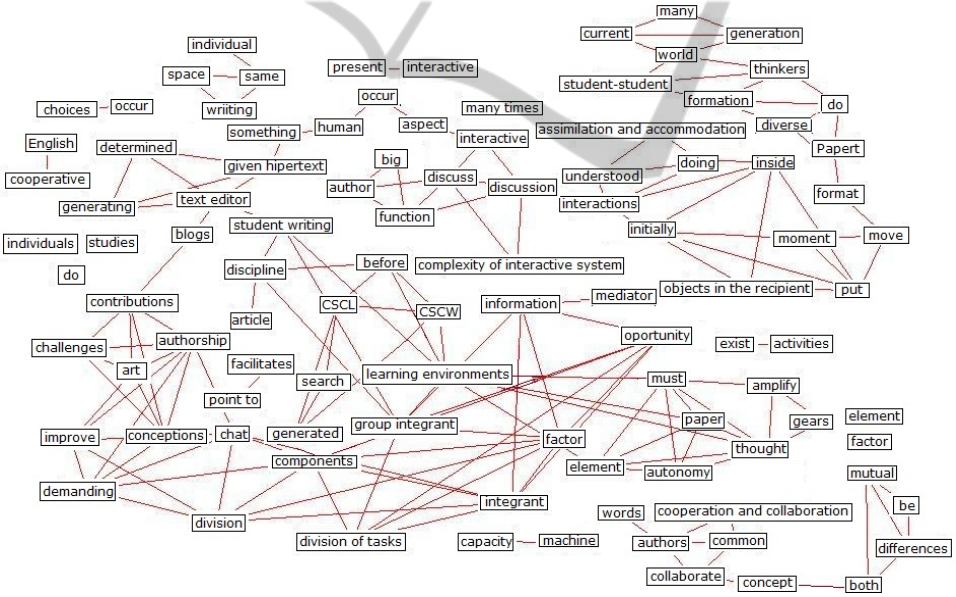

Figure 2: Concepts Network with incidence of interrelated terms.

Only cooperation leads to autonomy and the scope for mutual respect, where individuals can understand the result of agreements between different points of view and in this way integrate, a group that transcends the perspective of a single individual (Piaget, 1973).

Freire (2003) points out that autonomy is founded upon the experience of many decisions. In this sense, the pedagogical practice should focus on stimulating experiences of decision and responsibility (Freire, 2003). So, when the Concepts Network does not give evidence of the approached subject, indicating that the production did not focus on the proposed theme, it is suggested that the interaction between the teacher and student group adjusts to that focus.

The challenges and questions proposed, in addition to promoting the development of students, may also serve as a reference to the teacher who needs to know the potential and limitations of each

individual.

It is further highlighted that, in a context where, from the Concepts Network it is possible identify the developed theme, indicating that the written production gives focus to the proposed issue, the teacher may use this information as a basis for investigating students' level of knowledge on the topic. Based on this, the teacher can provide materials and discussions that allow theoretical deepening about the subject. Thus, such conditions can leverage the knowledge and relationing of the theme in question with other aspects, encouraging new developments, increasing knowledge and improving social exchanges that support the collective writing of text.

A pedagogical practice which encourages and creates favorable conditions for research, reading, analysis, inter-individual discussion and the production of new knowledge, can provide the development of autonomy, forming a subject with critical spirit, with initiative go in search of elements that meet their needs (learning to learn), articulating and confronting their perspectives with others, who think and argue from different points of view. Thus, in this study it is believed that, having the Concepts Network as support for detecting if the text production focused on the proposed theme, can qualify the teaching practice in order to point out the need for new strategies which meet the needs posed.

4.2 Focus on Quality of Text

In addition to identifying the theme dealt with in the text, the Concepts Network differentiates productions which require perfecting and those textual productions well developed. The Networks in which isolated terms or small groups of terms prevailed, were drawn from texts that needed perfecting. Figure 1 shows a Network with these characteristics.

In this case, the studied texts showed a greater need, in regard to clarity in development and objectivity, as well as in the sequence and consistency in the development of the writing. On the other hand, Networks in which related terms prevailed were also identified. Here, in contrast to the previous situation, the studied texts showed clarity, objectivity, sequence and consistency. Figure 2 shows a Network with an incidence of interrelated terms.

Subjects build their knowledge from their interaction with the physical and social environment (Piaget, 1973). In this sense, reading, reflecting, building a perspective on the selected theme and discussing it with other subjects is a learning exercise. It is this direction that it is believed that the pedagogical practice which is concerned with creating conditions to foster interactions should be supported, buoyed in conditions of equilibrium as predicted by Piaget (1973).

In addition to the conditions already considered, another important factor to investigate associated with Concepts Networks which show evidence of the need for improvement in the text, is the presence or absence of relationships based on self-centeredness, coercion and cooperation.

When in the presence of self-centeredness, the subjects fail to coordinate their points of view, once they understand the things and the other individuals from their actions. The movement that this writing requires depends fundamentally on the articulation and coordination of the propositions of the subjects involved. Otherwise this relationship is governed by an unbalanced situation.

In this perspective, it is understood that the teacher must recognize when to intervene to foster the relationship the student must make between the object of knowledge and the level of development that the student finds himself/herself in that moment. This does not imply a single method of work. On the contrary, it is understood that the pedagogical practice should adopt different ways of working for different necessities, always with the goal of building the student's knowledge.

It is worth noting that this study does not intend to reduce the teaching strategies that were contemplated in this writing, as it is aware of the diversity of situations and variables that a learning process may involve.

5 FINAL CONSIDERATIONS

This study shows that the Concepts Network can both indicate the theme developed in the text and provide qualitative indicators of this production. Qualitative indicators differentiate the texts requiring improvement from those which were developed with clarity and objectivity. All these situations can be identified through the Network, without the prior reading of the original text. It is understood that this option only makes sense if it serves to support a teaching practice committed to the process of student learning and equally committed to its own qualification. From this perspective, four assumptions were taken as minimum to account for the complex relationship that involves the process of collective text

production, they are: autonomy, critical thinking, learning to learn and knowing how to articulate oneself in collective contexts. It is noteworthy also that this research brings, as a support perspective for these assumptions, the reflective practice of the teacher who must deal with diversity, with differing needs and different demands resulting from the process of collective construction and in this sense, a "single only" or an "ideal model" does not make sense.

Bearing in mind the above, the principal contributions that result from this study are presented in an objective form. First, there is the significant reduction of reading time required of the teacher to monitor students' collective text production. As a result, there is an increase in the length of direct interaction time between students and teacher, which can provide significant qualification in the teaching-learning process. In addition, the indicators shown by the Concepts Network can help the teacher to focus on his/her actions, acting directly on the needs and potentiality of students. Such conditions can extend the possibilities of knowledge building and skills of written production. Finally, it is emphasized that the possible strategies of pedagogical practices may qualify the practice of the teacher who cares and always aims to achieve ever higher and significant levels of excellence in what he/she does.

ACKNOWLEDGEMENTS

This study has been partially supported by CAPES, through grant No 2737/2010.

REFERENCES

Berry, J. A., and Linoff, G., 1997. *Data Mining Techniques for Marketing*, Sales and Customer Support.: Wiley.

Feldman, R. and Sanger, J., 2006. *Text Mining Handbook. Inglaterra*: Universidade de Cambridge.

Freire, P., 2003. *Pedagogia da Autonomia: saberes necessários à prática educativa*. Paz e Terra, São Paulo.

Greengras, E., 2001. *Information retrieval*, A survey. Available:http://citeseer.ist.psu.edu/greengrass00information.html. Accessed: March 2011.

Klemann, M., Reategui, E., et al., 2011. *Sobek: a Text Mining Tool for Educational Applications*. International Conference on Data Mining, Las Vegas, Estados Unidos.

Macedo, A. L., Behar, P., et al. 2009. Collective Text Editor: a new interface focused on interaction design. In: Arthur Tatnall; Anthony Jones. (Org.). "*Education and Technology for a better world*". 1 ed. Berlin / Germany: Springer, v. 1, p. 331-339.

Parrat-Dayan, S., 2007. A discussão como ferramenta para o processo de socialização e para a construção do pensamento. In: *Educação em Revista*, nº 45, Belo Horizonte. Available in: http://www.scielo.br/scielo.php?pid=S0102-46982007000100002&script=sci_arttext&tlng=es. Accessed: março 2011.

Piaget, J., 1973. *Estudos Sociológicos*. Forense, Rio de Janeiro.

Piaget, J., 1986. *A linguagem e o pensamento da criança*. Martins Fontes, São Paulo.

Piaget, J., 1994. *O juízo moral na criança*. Summus, São Paulo.

Piaget, J, 1998. *Sobre a Pedagogia*. Casa do Psicólogo, São Paulo.

Schenker, A., 2003. *Graph-Theoretic Techniques for Web Content Mining*. Tese de Doutorado em Ciência da Computação, University of South Florida.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS