Concept Mapping to Learn and Innovate
Proc. of Sixth Int. Conference on Concept Mapping
Santos, Brazil, 2014

A METHODOLOGY TO CHANGE STUDENT-WRITTEN TEXTS INTO REPRESENTATIVE CONCEPT MAPS

Marianna M. Junqueira, Patricia A. Silva & Flavio A. Maximiano, Universidade de São Paulo, Brazil
Email: mariannamjunqueira@usp.br

Abstract This paper will present a methodology that uses computer tools to change student-written texts into representative concept maps. This methodology is useful both in educational research and teaching because the final graphic layout obtained provides an illustrative, global and summarized perspective of the main concept relations mentioned in a set of texts. The methodology was developed based on texts collected from undergraduate Chemistry students from the Institute of Chemistry at Universidade de São Paulo (São Paulo University), Brazil, who answered a question related to the Environmental Chemistry II course. Concept maps were built considering the texts as they were written by students, original texts (1), and they were also changed for propositions (2). Concept maps built from texts with different structures were very similar when compared to one another.
Keywords: Conceptual maps, Texts, Methodological tool, Teaching, Educational research.

1 Introduction

Asking students to write texts as an answer to a question within the school environment seems to be the most conventional way of evaluating what students really know, because based on the text produced we are able to infer how much a student knows about a certain subject. When writing a text, the student brainstorms information on the subject, selects the most important pieces of information to answer the question he/she was asked, and then organizes his/her ideas to write a logical and coherent text.

Reading and reviewing students' texts are inevitable tasks, demanding considerable time of teachers, professors and researchers. And when one wants to know the most relevant ideas about a certain subject for a group of students, the task is even more difficult. Reviewing innumerous texts can be exhausting, and this makes it difficult for the teacher/professor or researcher to select and organize students' main ideas.

Based on the foregoing, this paper will present a methodology which has been developed by LAPAQ group - Chemistry Learning Research Lab, Institute of Chemistry, Universidade de São Paulo. This methodology uses computer tools to change student written-texts into concept maps and discusses the influence of text structure on the process. This paper prepared representative concept maps that provide a graphic structure with the major ideas conveyed in the texts written by a group of students.

The development of a methodological tool using written texts and providing a graphic representation of ideas is believed to useful for showing a view of the thoughts related to the subject in question, with major concept relations, all this in an illustrative, global and summarized way. Establishing relations between concepts is fundamental for determining how much a certain subject has been understood (Moreira, 2008).

Changing texts into maps is completely justifiable because although both adequately convey information, graphics are more efficient as they allow people to build more complex meanings and to explicitly integrate knowledge by having a visual resource at hand (Veriki, 2002). Thus, this tool allows people to give a new interpretation or meaning to the contents of the text by looking to a graphic representation (Kowata, Cury, & Boeres, 2012).

The use of this tool is quite broad. It can be used in educational research and teaching. From the visual representation obtained, teachers/professors are able to infer how much knowledge an individual or group of students has. A researcher is able to make innumerous comparisons in the research process, such as conducting longitudinal studies on a certain subject. Visual representations can also broaden the access to materials during research because they show visual aspects that guide the selection and sorting out of information. In teaching, teachers/professors are able to know the relations between the concepts that are the most important ones for a group of students, and are able to organize the concepts of a class and to structure the content of a course, in addition to using the tool to evaluate learning.

2 Methodology

2.1 Data Collection

The methodology was developed based on texts collected from 16 undergraduate Environmental Chemistry students from Universidade de São Paulo, Brazil, who answered a question posed in a final test of the Environmental Chemistry II course. The question was the following: Based on the following set of concepts (Figure 2) and using these concepts (all of them, or just some of them, or words inserted thereon, or other words you may find necessary), write a paragraph to explain their relation in the context of chemical activities and their effects on the environment and human health. Students had to write a text on the subject proposed using the trigger concepts provided in Figure 2, in a way it was possible to identify in the text the presence of relations among these concepts or variations in them. These trigger concepts were pre-established based on their importance in the context of the task, with the objective of promoting students' reflection on the potential existing relations (Peixoto, 2003).

2.2 Changing texts into concept maps

Generally, the process of changing texts into concepts maps is divided into two parts, as shown in Figure 1. The first step is related to the automatic obtainment of a representative concept association matrix of the group of students, i.e., it represents the number of relations among concepts found in all texts. Hamlet® software was used to automatically obtain the association matrix. The second step is related to the semi-automatic preparation of the representative concept map using software CmapTools® and Hamlet®. The entire process is shown below.

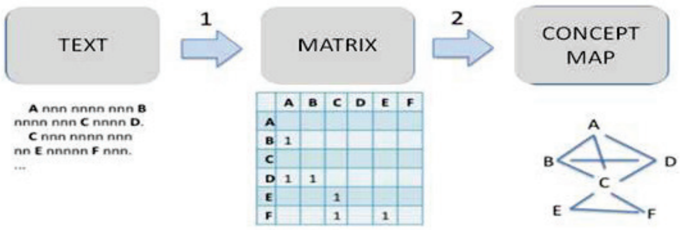

Figure 1: Flowchart of the process of changing texts into concept maps.

2.2.1 From texts to the association matrix

As a first step, all texts were digitalized as they were written and they were called original texts (1). After that, the texts were read by software Hamlet® in a single file so that a concept association matrix was created for the group of students.

Hamlet® searches for words in text files having as reference a list of words previously prepared. The software gives us the frequency in which these words appear throughout the texts, sentences or any text unit specified through an association matrix (Brier & Hopp, 2010). Here, the list of words was prepared based on the trigger concepts indicated in the question and their semantic variations.

The way the software reads texts and quantifies concept relations presented in the word list to create the association matrix is quite simple: for a certain sentence, all concepts present in this fragment and that are also present in the word list are acknowledged by the software as related concepts, that is, concepts that have some connection between them. In case of repeated concepts in one sentence, the software does not relate them to one another, but relates them to all the other concepts. Therefore, if a sentence written by one student has five concepts which are also mentioned in the word list, but one concept is repeated, the software will acknowledge eight different relations in this sentence, similarly to what is shown in Scheme 1. In this example the solvent concept is repeated.

Scheme 1: Example mentioning the relations acknowledged by Hamlet® for a sentence taken from the original text collected in the Environmental Chemistry II course.

SENTENCE	RELATIONS BETWEEN CONCEPTS ACKNOWLEDGED BY HAMLET®
"Decreased **waste** generation and organic **solvent** use can be seen in **atomic efficiency** calculations, which measure how much of **reagent** and auxiliary **solvent** atoms effectively participate in the final product."	1 waste – solvent
	2 waste - atomic efficiency
	3 waste – reagent
	4 waste – solvent
	5 solvent - atomic efficiency
	6 solvent – reagent
	7 atomic efficiency – reagent
	8 atomic efficiency - solvent

Once Hamlet® finished reading the original texts (1), the concept association matrix seen in Figure 2 was obtained. This matrix indicates the number of pair by pair concept relations found in all sentences written by the students.

Catalyzer	1																						
reaction conditions	2	7																					
Carbon dioxide	3	4	2																				
Atomic efficiency	4	11	4	2																			
wastewater	5	2	5	1	2																		
emissions	6	6	4	4	5	5																	
stoichiometric	7	12	6	4	8	3	4																
Hydrogen gas	8	5	2	3	3	2	1	4															
Oxygen gas	9	5	2	3	3	2	1	4	8														
Pharmaceutical industry	10	3	1	0	1	1	1	2	0	0													
Chemical industry	11	2	0	0	0	1	2	0	0	0	8												
Raw materials	12	3	2	2	1	1	1	2	2	2	5	5											
Renewable raw materials	13	4	1	1	2	0	1	2	0	0	0	0	3										
minimize	14	11	5	4	6	2	5	5	2	2	1	1	1	4									
Petrochemistry	15	1	0	1	2	0	2	1	0	0	5	6	10	2	1								
Steam pressure	16	1	0	1	0	3	3	1	1	1	0	0	2	0	1	0							
Reagents	17	7	4	3	6	4	6	8	2	2	2	3	4	0	4	6	0						
Wastes	18	11	6	2	10	3	3	9	4	4	2	1	4	3	9	0	2	6					
inorganic salts	19	8	3	1	3	4	1	5	3	4	0	0	2	0	1	1	1	4	4				
Selectivity	20	11	6	3	9	3	7	10	1	1	1	0	3	2	7	1	1	6	9	3			
organic synthesis	21	4	3	2	2	4	4	5	2	2	1	1	2	1	1	3	2	5	3	4	5		
Solvents	22	8	5	9	5	8	7	10	4	4	2	1	5	0	3	4	6	15	7	7	5	8	

Figure 2: Concept association matrix obtained from the original texts (1) collected in the Environmental Chemistry II course.

Subsequently, propositions containing explicit concept relations provided in the question were taken from each text written by the students. This was done for the purposes of comparing the influence of text structure upon matrix and concept map creation. It is important to mention propositions are extracted manually and have a subjective nature because they are connected to the interpretation of the person who wrote the text. However, subjectivity is reduced with the steps described below:

1. Digitalization of all texts with their original structure preserved;
2. The pre-defined concepts mentioned in the student's text are highlighted;
3. Effective concept relations mentioned in sentences are established;
4. Finally, the proposition is structured with its respective connecting sentence or expression, preserving the original meaning.

Please find in Scheme 2 an example for better understanding the process of changing original texts (1) into propositional texts (2).

Scheme 2: Example mentioning propositions extracted from a sentence of the original text collected in the Environmental Chemistry II course.

SENTENCE	PROPOSITIONS EXTRACTED
"Decreased **<u>waste</u>** generation and organic **<u>solvent</u>** use can be seen in **<u>atomic efficiency</u>** calculations, which measure how much of **<u>reagent</u>** and auxiliary **<u>solvent</u>** atoms effectively participate in the final product."	**<u>Atomic efficiency</u>** calculations show decreased in production of **<u>waste</u>**.
	<u>Atomic efficiency</u> calculations show decreased in use of **<u>solvent</u>**.
	<u>Reagents</u> how much of its atoms have participated in the final product are shown in the calculation of **<u>atomic efficiency</u>**.
	<u>Solvents</u> how much of its atoms have participated in the final product are shown in the calculation of **<u>atomic efficiency</u>**.

Once all texts have been changed into propositional texts, Hamlet® reads them in a single file, creating the concept association matrix shown in Figure 3. The numbers found in the matrix correspond to the number of times a certain relation was written by the students.

Catalyzer	1																						
reaction conditions	2	2																					
Carbon dioxide	3	0	0																				
Atomic efficiency	4	8	1	0																			
wastewater	5	0	1	0	0																		
emissions	6	1	1	2	0	0																	
stoichiometric	7	7	0	0	2	0	1																
Hydrogen gas	8	1	0	1	1	0	0	1															
Oxygen gas	9	1	0	1	1	0	0	1	0														
Pharmaceutical industry	10	4	0	0	2	2	2	1	0	0													
Chemical industry	11	2	0	0	1	1	1	0	0	0	0												
Raw materials	12	2	1	2	0	0	1	0	0	0	0	0											
Renewable raw materials	13	0	0	1	0	1	1	0	0	0	1	1	1										
minimize	14	6	0	1	0	0	0	2	0	0	0	0	0	1									
Petrochemistry	15	0	0	0	1	0	1	0	0	0	3	3	6	3	0								
Steam pressure	16	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0							
Reagents	17	5	1	0	0	0	1	2	5	5	0	0	1	0	2	3	0						
Wastes	18	2	1	0	2	1	0	5	1	1	1	0	0	0	3	0	1	0					
inorganic salts	19	5	0	0	0	2	0	1	0	1	0	0	1	0	0	0	0	1	1				
Selectivity	20	9	0	0	1	0	0	2	0	0	0	0	1	0	0	0	0	0	3	0			
organic synthesis	21	1	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	1	0	1		
Solvents	22	3	2	7	2	2	0	3	0	0	0	0	1	0	2	2	5	0	2	0	0	2	

Figure 3: Concept association matrix obtained from the reading of propositional texts (2).

2.2.2 From the association matrix to the representative concept map

A percentage limit representing the relations mentioned more often in the text was chosen to create the concept maps with the help of *CmapTools* (Cañas et al., 2004). In this sense, the percentage limit is established as follows: first, for a certain representative association matrix, the numeric values appearing thereon are organized in a sequence. Then, the number of times each number appears in this matrix is counted. As an example, please find below in Figure 4 how the numeric values in Figure 2 matrix are organized in a sequence and counted to create the percentage limit. The percentage limit is 25%. Thus, the concept map was created considering only those concept relations appearing more than 6 times.

Relationships established from 6 times shall be considered in the map without in 25%

Value in the matrix (1)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of times that value appear in the matrix	43	40	26	29	19	12	7	8	5	4	4	1	0	0	1

<u>Sum up to this point</u>: 42 – Value nearest 49.75 (25% of the 199 relationships between

Figure 4: 25% percentage limit example applied to Figure 2 matrix data.

First, we obtain graphic structures showing the concepts and the existing connections according to the numeric values appearing in the matrix. Once the pair by pair relations to be shown in the concept map are defined, we searched for connecting sentences as a way to build the propositions part of the concept map with the help of Hamlet® KWIC (Key Word In Context) and human resources. Figure 5 shows the concept map created after the original texts (1) had been read and Figure 6 shows the concept map created after the propositional texts (2) had been read.

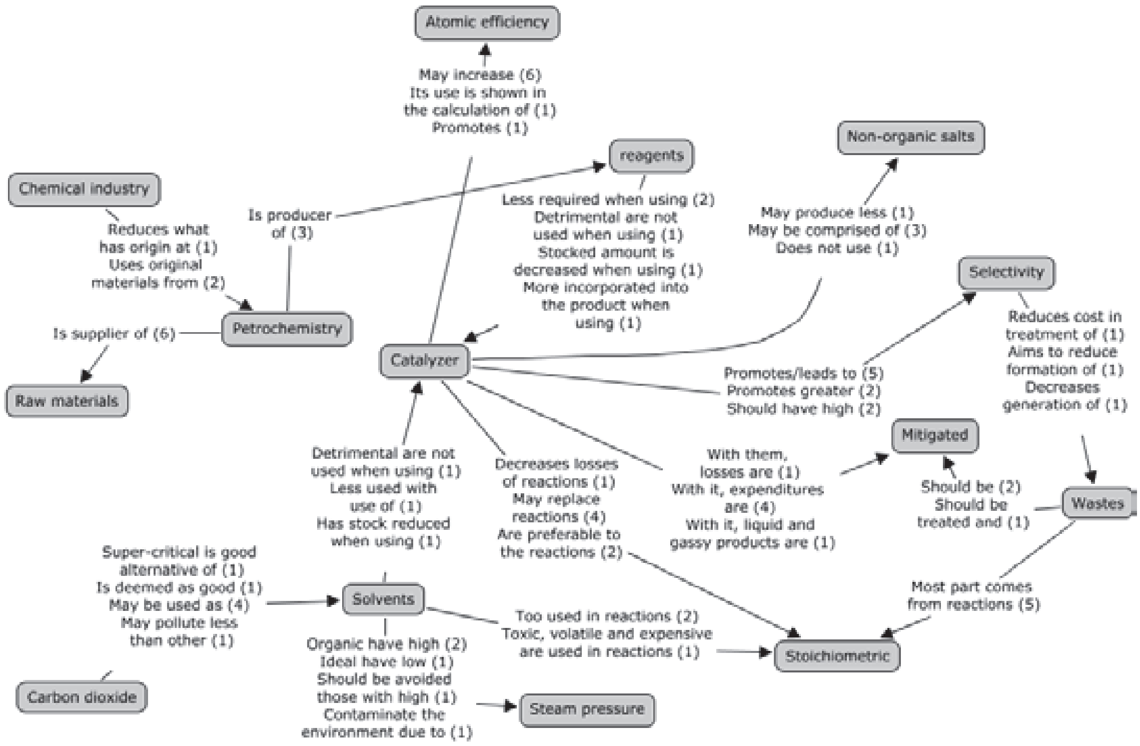

Figure 5: Concept map created from original texts (1).

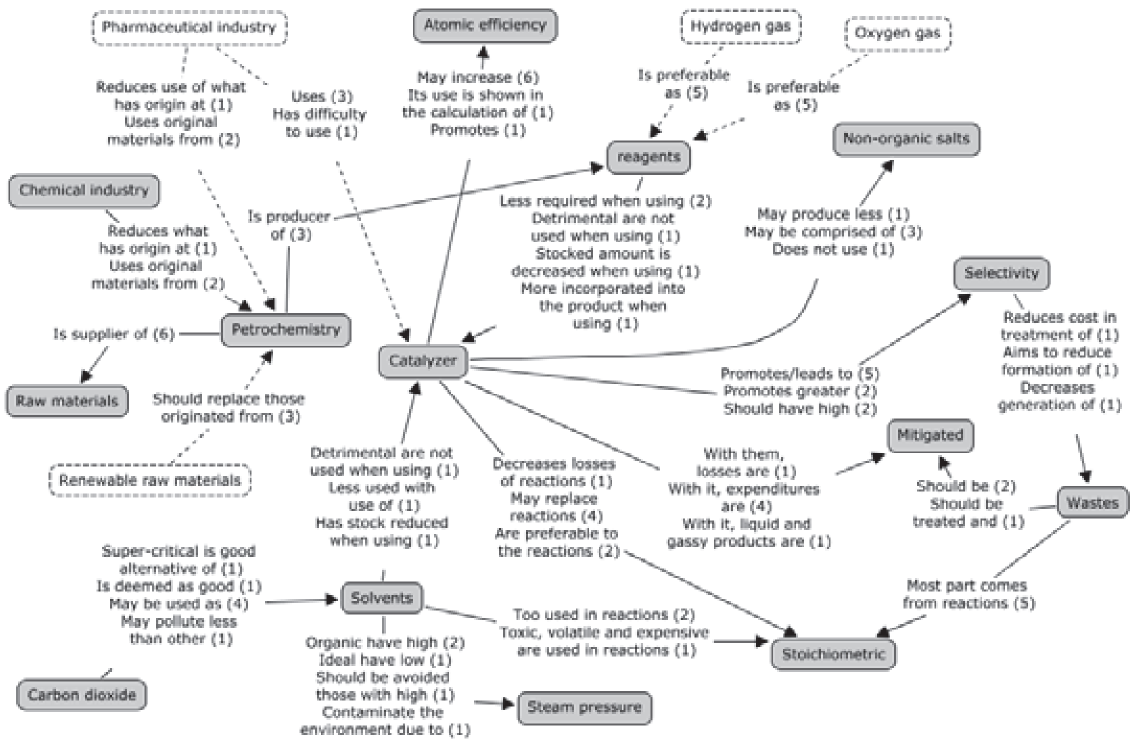

Figure 6: Concept map created from propositional texts (2).

3 Results

Original text (1) reading and propositional text (2) reading create different concept association matrixes because the number of concept relations counted is different. Scheme 3 below shows a sentence and the relations counted after the original texts and the proposition-modified texts had been read.

Scheme 3: Relations counted after original texts had been read and propositions extracted for the sentence.

SENTENCE	RELATIONSHIPS BETWEEN WORDS PRESENT IN THE ORIGINAL TEXTS (1)	RELATIONSHIPS BETWEEN WORDS PRESENT IN THE PROPOSITIONAL TEXTS (2)
"Decreased **waste** generation and organic **solvent** use can be seen in **atomic efficiency** calculations, which measure how much of **reagent** and auxiliary **solvent** atoms effectively participate in the final product."	waste – solvent	
	waste - atomic efficiency	waste - atomic efficiency
	waste – reagent	
	solvent - atomic efficiency	solvent - atomic efficiency
	solvent – reagent	
	atomic efficiency - reagent	atomic efficiency - reagent

The example given shows that concept relations to be identified by Hamlet® essentially depend on the structure of text in question. Therefore, the matrix created by the software, for the purposes of quantifying concept relations, will also be different according to the structure of the text being analyzed. It is possible to say the matrix created by Hamlet® based on the analysis of students' original texts quantifies direct and indirect concepts relations. And this means that part of the relations identified coincide with those identified in the modified texts; however, others do not. For sentence 01 (Scheme 3), for example, relations between three pairs of concepts *waste* - *atomic efficiency; solvents* - *atomic efficiency;* and *atomic efficiency* - *reagents* are identified by analyzing two types of text structure. Relations between concept pairs *waste* - *solvents*; *waste* - *reagents*; and *solvents* - *reagents* cannot, therefore, be considered direct or propositional relations, and they have been acknowledge by Hamlet® after original text reading because the concepts mentioned were present in the same sentence.

The differences presented in the matrixes will influence the creation of concept maps. Specifically, the concept map based on propositional texts included concept relations above 3 times, with a number of concepts and connections adequate for the creation of an organized and easy-to-understand graphic structure.

Similarly, the concept map based on original texts also considered the 25% limit, but included relations above 6 times, because the original text matrix counted all relations between concepts and the values marked are higher. However, when searching for the connecting sentences between the pair of concepts for the creation of the map, it was possible to see, once again, that some concept pairs were not related in the students' texts in a propositional way, as shown in Figure 7.

In this sense, by searching for connecting sentences between concept pairs for the creation of the concept map, it was possible to see that some concept relations included in the graphic structure were not propositional. Thus, for some concept pairs, the number of times the relation has been established can be decreased when the initial graphic structure is compared to the final concept map, and some concept relations can even be excluded from the created concept map.

Therefore, out of 42 different connections between concepts included in the graphic structure initially created from the original text representative matrix (not shown here), 13 were excluded during the creation of the corresponding concept map because they were not propositional. Additionally, by searching for connecting sentences, it was possible to see that other 13 different related concept pairs actually had a propositional relation by less than 3 times. In this context, considering the purpose of this paper is converging maps created from original and modified texts as to check for the similarity between these two structures, it does make sense to create a concept map from original texts with concept relations established by less than 3 times, which are also not included in the modified text map. Thus, the original text concept map was completed with a total of 16 different connections, where the total number of relations established for each concept pair is below the number found in the graphic structure of origin concepts.

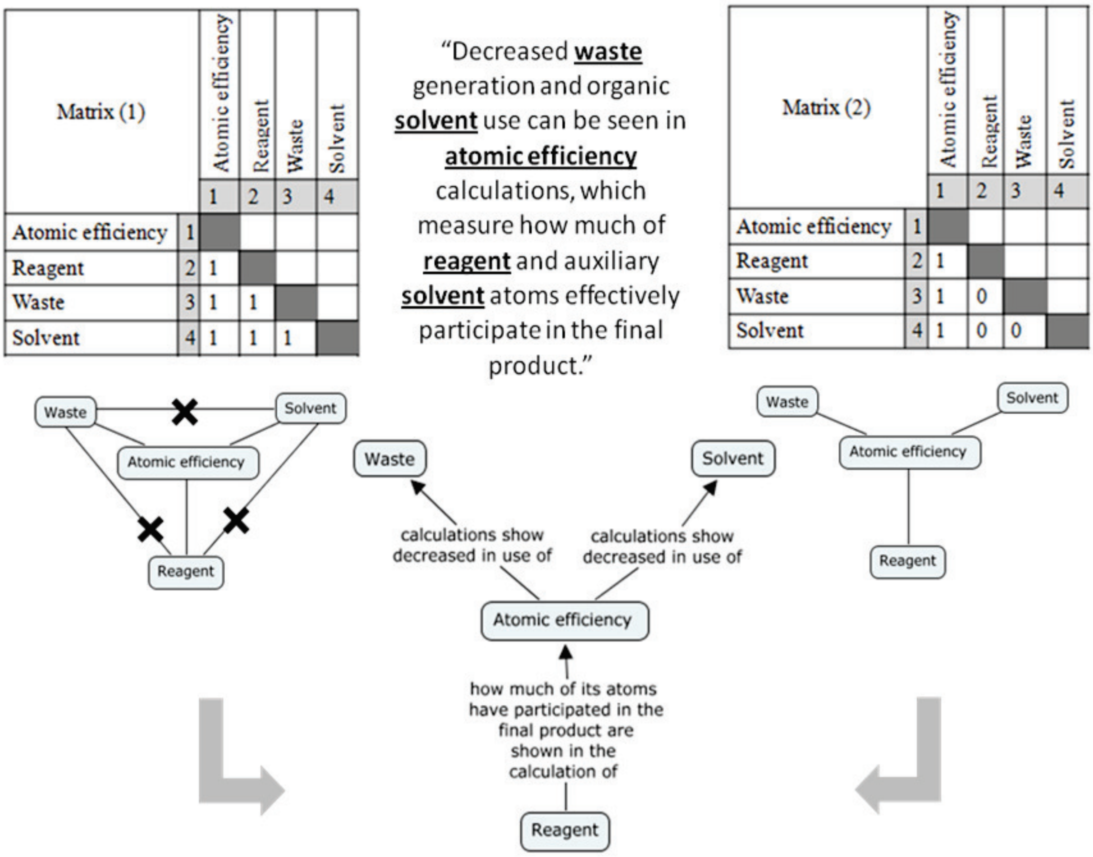

Figure 7: Illustration for the concept map creation process from two possible text structures (1 - text as written by the student; and 2 - text modified for propositions).

By comparing the two representative concept maps obtained, we have an APA - Average Percent of Agreement - (Clariana & Koul, 2008) value of 0.88. This index is used to compare similarities among concept maps. The calculation is as follows:

$$APA = \left(\frac{\mathbf{NL}\ MC1MC2}{\mathbf{NL}\ MC1} + \frac{\mathbf{NL}\ MC1MC2}{\mathbf{NL}\ MC2}\right) \cdot \frac{1}{2}$$

NL $_{MC1\ MC2}$ = number of common connections between map 1 and map 2.
NL $_{MC1}$ or NL $_{MC2}$ = total number of connections in map 1 and map 2.

APA values near one indicate high similarity between maps. On the other hand, values near zero indicate few similar connections. The value found in the comparison indicates concept maps are 88% similar, which can also be seen in the comparison made and organized in Scheme 4. Scheme 4 shows that the number and nature of core concepts, concepts connected to three or more concepts, is different in only one concept. Regarding end concepts, those having only one connection, we have only three different concepts in the propositional text concept map.

For the subject Environmental Chemistry II the conceptual maps represented in figures 5 and 6, for example, have too much information in relation to the 5 concepts deemed as central in both graphic structures. Acknowledgment of these concepts by the students as central and important within the green chemistry context is proved to be interesting from the point of view of the understanding of the students on the subject, mainly regarding the so-called twelve green chemistry principles (Prado, 2003).

Scheme 4: Comparison among concept maps regarding concepts and existing connections.

	Concept Map – Original Texts	Concept Map – Propositional Texts
Number of concepts	14	18
Number of core concepts	5	6
Nature of core concepts	Catalyzer stoichiometric Petrochemistry Wastes Solvents	Catalyzer stoichiometric Petrochemistry Wastes Solvents *Reagents*
Number of end concepts	6	9
Nature of end concepts	Carbon dioxide Atomic efficiency Chemical industry Raw materials Steam pressure inorganic salts	Carbon dioxide Atomic efficiency Chemical industry Raw materials Steam pressure inorganic salts *Oxygen gas* *Hydrogen gas* *Renewable raw materials*
Total number of connections	16	21

4 Conclusions

The process of creating a concept map from two potential text structure matrixes is a dynamic and interactive process. The purpose is having an accurate and coherent representation of what was written by the author of the text and for this, the text and the graphic structure are constantly taken into consideration in the process. By searching connecting sentences for creating the original text concept map (1), it was possible to obtain a graphic structure which was considerably similar to the propositional text concept map (2). Additionally, the similarity between the maps suggests Hamlet® is a great software for obtaining matrixes that quantify concept relations in a text as it was written, when compared to the manual and time-demanding process of extracting propositions from a text for the creation of a matrix.

5 Acknowledgements

The Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, Brazil, for financing the work.

References

Brier, A., & Hopp, B. (2010). Hamlet II, Software for computer-assisted text analysis. Tutorial on the operation of the program. Available in: <http://apb.newmdsx.com/hamlet2.html> Acessed: 12 may 2014.

Cañas, A. J. et al. (2004). Cmaptools: A Knowledge Modeling and Sharing Environment. In: CAÑAS, A. J.; NOVAK, J. D.; GONZÁLEZ, F. M. (eds.), Concept maps: theory, methodology, technology. Proceedings of the First International Conference on Concept Mapping, p.125 – 133. Pamplona, Espanha: CMC.

Clariana, R.B., & Koul, R. (2008). The effects of learner prior knowledge when creating concept maps from a text passage. International Journal of Instructional Media, 35 (2), 229-236.

Kowata, J. H., Cury, D., & Boeres, M. C. S. (2010). Concept maps core elements candidates recognition from text. In: SÁNCHEZ, J.; CAÑAS, A. J.; NOVAK, J. D. (eds.), Concept maps: making learning meaningful, Proceedings of the Fourth International Conference on Concept Mapping, p.120 – 127, Viña del Mar, Chile: CMC.

Moreira, M. A. (2008). Conceptos en la educación científica: ignorados y subestimados. Qurriculum. v. 21, 9-26.

Peixoto, H. R. C. (2003). Natureza da Ciência e Formação de Professores de química: uma experiência de sala de aula. Dissertation (Master's in science education) - Universidade de São Paulo.

Prado, A. G. S. (2003). Química Verde, os desafios da química do novo milênio. Química Nova, 26(5), 738 – 744.

Veriki, I. (2002). What is the value of graphical displays in learning? Educational Psychological Review, v.14, 261–312.