

Concept Maps as Cognitive Visualizations of Writing Assignments

Jorge Villalon^{1,2} and Rafael A. Calvo¹

¹School of Electrical & Information Engineering, University of Sydney, Australia // ²Facultad de Ingenieria y Ciencias, Universidad Adolfo Ibanez, Chile // jorge.villalon@uai.cl // villalon@ee.usyd.edu.au // rafa@ee.usyd.edu.au

ABSTRACT

Writing assignments are ubiquitous in higher education. Writing develops not only communication skills, but also higher-level cognitive processes that facilitate deep learning. Cognitive visualizations, such as concept maps, can also be used as part of learning activities including as a form of scaffolding, or to trigger reflection by making conceptual understanding visible at different stages of the learning process. We present Concept Map Miner (CMM), a tool that automatically generates Concept Maps from students' compositions, and discuss its design and implementation, its integration to a writing support environment and its evaluation on a manually annotated corpora of university essays (N=43). Results show that complete CM, with concepts and labeled relationships, are possible and its precision depends the level of summarization (number of concepts) chosen.

Keywords

Concept Map Mining, Concept Map, Automatic generation, Text Mining, Writing

Introduction

University students are expected to develop higher-order cognitive skills such as analysis and synthesis, and also meta-cognitive skills. The first goal is often tackled through writing activities, arguably the task where higher cognitive functions are best developed (Emig, 1977). The second goal of developing meta-cognitive skills can be facilitated by cognitive visualizations (also known as knowledge visualizations) (Jacobson, 2004). Cognitive visualizations are tools that make “thinking” visible, reifying learners’ mental model about domain knowledge onto an explicit graphical device. They make possible the application of “cognitive apprenticeship” approaches like reflection and scaffolding (Collins, Brown, & Holum, 1991). Reflection can be supported by confronting the learner with a different visualization of her own knowledge, while scaffolding can be implemented by contrasting the learners’ visualization to that of an expert.

One cognitive visualization technique is Concept Mapping. Concept Maps (CM) represent a person's understanding of a topic by mapping concepts and their relationships in a hierarchical way, where more general concepts are placed higher in the map and concepts at the same level of generalization are grouped together. There is extensive evidence that drawing a CM requires students to engage in higher cognitive functions (Novak & Gowin, 1984). CMs have typically been used in reading activities to aid students’ comprehension of texts. For instance, ready made CMs may be presented as semantic summaries of texts that students need to comprehend (Hauser, Nuckles, & Renkl, 2006), or students may be asked to construct their own CMs to address specific questions (Chang, Sung, & Chen, 2002).

Our own work presents a novel approach in the educational application of CMs. Here, CMs are embedded in writing (as opposed to reading) activities and are used to summarize the students’ own writing. Unlike in the more typical scenarios of using CMs to support reading, in our work the CMs become approximate representations of students’ current state of knowledge. From the students’ perspective, such CMs can be used to reflect on their own knowledge and also to help students see their writing from a different perspective. From the instructor’s point of view, such CMs can be used as a rapid assessment of students’ conceptual understanding.

One challenge to using CM as a way to support students’ writing, is that it is time consuming and expensive. Automated tools for generating CMs aim to reduce the workload (of the instructor) or redirect it to other learning activities that support the expected outcomes. Attempts to automatically generate CMs have so far been limited to approaches that generate pseudo-CMs where the relationships between concepts are not labeled, and with evaluations that cannot be replicated in a new experimental setting. Our research program has been to use linguistic analysis and statistical techniques to automatically generate complete CMs. We have incrementally stepped towards the automatic extraction of full CMs. First we proposed the idea and the evaluation framework (Villalon & Calvo, 2008), then the first results for a concept extraction algorithm (that did not include relationships) (Villalon & Calvo, 2009). This work required a linguistic and morphological analysis of the benchmarking corpus that was reported (Villalon,

Calvo, & Montenegro 2010). Despite this progress many challenges remain, the first set being of a technical nature. The automated CMs described in previous work were still significantly less accurate and complete than those built by humans, they do not even have the required relationships between the concepts. The second set of challenges is of a pedagogical nature. Concept maps have so far always been prepared by human readers. The research literature has focused on how they are useful for students to describe (and develop) their conceptual schemas on a certain topic (e.g. on something they have recently read). Since automated CM are only now becoming possible, there is no clear understanding of how they can be used in real learning situations. This new technical opportunity brings the challenge of figuring out the new pedagogical opportunities. This analysis requires a review of learning activities where they can be used, particularly reading and writing where CM can provide a means for developing cognitive and metacognitive skills.

The first contribution of this paper addresses one of the key technical challenges, a technique for automatically producing labeled relationships. The second contribution of the paper is a review of possible learning activities using automated CMs. In particular, next section reviews aspects of the literature on academic writing and concept mapping as learning activities, supporting the pedagogical motivation of the framework. We also review current progress on automated techniques for concept map mining. We show how CMs can be useful to support writing activities, as they are for reading. We follow with a description of a new CMM tool and its integration into Glosser (Calvo & Ellis, 2010) a tool for automated feedback in writing activities. This new tool provides a means for the meta-cognitive skills described earlier by showing the student a different perspective of their own writing. Furthermore we provide an evaluation using a collection of student essays. We finalize with a discussion on the implications of our work.

Background and Related Work

Essay writing

Essay writing is considered a unique way of learning because it involves en-active (learning "by doing"), iconic (learning "by depiction in an image") and symbolic learning (learning "by restatement in words"). As Emig (1977) explains, writing is "the symbolic transformation of experience through the specific symbol system of verbal language is shaped into an icon (the graphic product) by the en-active hand". This symbolic transformation involves the recall and synthesis of ideas.

A study by Paltridge (2004) found that universities require graduate students from all disciplines to write academically. Another study by Moore and Morton (1999) found that the academic essay represented almost 60% of writing tasks at both the undergraduate and graduate levels. The increased emphasis on writing may reflect educational leaders' recognition of the connection between writing and knowledge production in an information era (Warschauer & Ware, 2006). However, writing academic essays is challenging. It requires much more than good surface writing skills such as producing grammatically correct sentences. As Paltridge (2004) observed, writing academic essays demands deep understanding of "the roles played by the people involved in the production of the texts, and the contexts in which the texts are produced and assessed".

Automatic feedback in the writing process

One way to help students learn to write is by providing feedback on their writing process. In considering writing as a process, Keh (1990) explains that "feedback is the drive which steers the writer through the process of writing on to the product". It is from feedback that writers learn if she "has misled or confused the reader by not supplying enough information, illogical organization, lack of development of ideas". Quality feedback makes the author's thinking visible and allows the author to reflect on her own work (Lin, Hmelo, Kinzer, & Secules, 1999). Regrettably, meaningful and timely feedback is difficult and expensive to provide, so researchers have been working on automated and semi-automated approaches to feedback. So far, most of the work on automated feedback has been on assessment or written suggestions on features that students should improve on (Beals, 1998; Graesser & McNamara, in press; Thiesmeyer & Thiesmeyer, 1990; Wade-Stein & Kintsch, 2004; Wiemer-Hastings & Graesser, 2000). Alternative visual representations of their own writing provide a new form of feedback (Calvo & Ellis, 2010).

Concept Maps

Concept Maps (CM) were introduced by Joseph Novak as a way to assess children's understanding of science with graphical tools to organize and represent knowledge (Novak & Gowin, 1984). In a CM, concepts are represented in boxes that are linked by labeled relationships; two related concepts (including their link) form a proposition or semantic unit. Concepts are also arranged hierarchically such that more general concepts are located higher on the map and specific concepts such as examples are located lower. Novak defines a concept as "a perceived regularity in events or objects", or "records of events or objects" designated by a label. A concept by itself does not provide meaning, but when two concepts are connected using linking words or phrases, they form a meaningful proposition.

Since inception, CMs have been widely used and tested as pedagogical tools (c.f. Novak, 2007). A study by Hauser, Nuckles, and Renkl (2006) compared a control group to several groups using CMs in four ways: constructing CMs from scratch, constructing CMs from a list of concepts, constructing CMs from spatially arranged concepts, and finally studying previously built CMs. The results showed that constructing maps from scratch and studying previous constructed ones led to significantly better learning outcomes than the other conditions. A similar study by McClure presented evidence on the validity and reliability of concept maps as an assessment tool (McClure, Sonak, & Suen, 1999). Another study validated the use of concept maps to improve text comprehension and summarization (Chang, Sung, & Chen, 2002). This study compared a control group to three concept mapping approaches: showing an expert generated CM; scaffolded concept mapping (completing a partially-completed expert CM), and constructing CMs from scratch. The results indicated that all approaches to concept mapping improve text comprehension and summarization skills, with the scaffolded concept mapping approach leading to the best outcomes. CM can also be used to scaffold students' reflection. Studies by McAleese show that CMs can be used as mirrors or assistants to the learner. These studies also describe the process of "off-loading" concepts from the mind to the map, and identify several cognitive steps that people take when constructing CMs (McAleese, 1998).

CMs are not the only cognitive visualization technique for documents and text. Recently proposed visual representations including Word Trees (Wattenberg, & Viegas 2008) and Radial Document Visualization (Collins, Carpendale, & Penn 2009) are incorporating relevant semantic information that are improvements over simple statistical visualization like Tag Clouds. Even though such representations are a step forward to a semantic view of a document, they present two limitations when compared to CMs. Word Trees present keywords in the context they appear (typically sentences), however different keywords are not connected to each other as concepts are connected in CMs. Such connections are key to represent meaningful propositions, that give CMs their knowledge richness. Radial Document Visualizations are built in a more semantic way, using connections between concepts, however these connections are limited to "is-a" relationships (hyponymia), our own research has shown that such connections in student essays count for only a small part of the knowledge (Villalon, Calvo, & Montenegro 2010).

Previous studies on automatic concept mapping

A study by Valerio and Leake (2006) analyzed the requirements for generating CMs from text for educational purposes. These requirements can be summarized into the three principles of Educational utility, Simplicity, and Subjectivity (Villalon & Calvo, 2008), elaborated below:

- *Educational utility*: There is scientific evidence for the educational utility of concept mapping following Novak's method. CMs must include concepts, which should be connected by linking words to form propositions. They must also have a topology where more general concepts are placed higher in the map and specific concepts lower. Concepts with the same level of generalization should be placed at the same level (Novak, 2007).
- *Simplicity*: CMs are mainly used for human analysis, giving teachers and students an alternative, structured, and cheaper representation of students' understanding. Novak's definition also indicates that a CM should require no more than 25 concepts to answer its focus question. Since there can be many more 'concepts' in a medium length document, CMs (either human made or automatic) can be conceived as a visual summary of the complete document.
- *Subjectivity*: As mentioned earlier, CMs represent both the author's knowledge and her writing skills. In an educational context, the terminology used by the student is also important for assessing the outcome, so CMs should be represented using the terms that the author used in her text. If a student uses a certain word to refer to a particular concept, this choice inevitably reflects her vocabulary level, and hence a concept map should retain

this information. This means that poorly written essays should be able to produce a CM with meaningful propositions that should also reflect the author's writing skills limitations (like spelling and grammar errors).

Several studies have focused on the automatic generation of CMs from text. These studies expose challenges at two levels: An inconsistent definition of CMs, and a lack of an evaluation framework. The first problem occurs when authors report on the automatic creation of CMs, but are actually creating a different type of knowledge representation with its unique characteristics, which do not follow Novak's definition. The second problem is that each study uses a different method to evaluate its contribution, making it very hard to compare them. For a detailed description of previous attempts see (Villalon & Calvo, 2008).

Three studies described the construction of CMs from documents to facilitate the visualization of domain knowledge (Chen et al., 2008; Clariana & Koul, 2004). These studies create maps with concepts that are connected by unlabeled relationships and generate propositions which do not conform to Novak's (2007) method. Three other studies addressed these limitations, creating CMs which include some type of labeled relationships. Oliveira, Pereira, and Cardoso (2001) presented a system called TextStorm which generates 'raw CMs' from texts for students to elaborate further. These authors used a rule-based approach that can analyze only affirmative and declarative sentences. The CM quality was evaluated by manually assessing propositions generated from a set of articles, essays and book chapters. The second study, by Valerio and Leake (2006) also proposed the idea of using automatically generated CMs as an educational tool to 'jump-start' the CM construction. In this study CMs were evaluated in an application based fashion. The generated concepts were used to link new documents to improve precision and recall in Information Retrieval. The approach does not evaluate relationships or the CM's fidelity to the text source. A study by Zouaq and Nkambou (2008) presented a system named TEXCOMON that creates CMs as part of an Ontology Learning from Text (OLT) process. Using linguistic analysis they generate a CM from a corpus and use it to infer an Ontology that represents the corpus knowledge. They evaluated the quality of the generated ontology by comparing it with an ontology generated using state-of-the-art OLT tools.

The Concept Map Miner (CMM)

Within our research program for providing automated feedback for writing activities we have shown progress in writing support tools allow that feedback to be provided at any stage of the writing process (Calvo et. al., 2011). This is a significant improvement over previous efforts that focused on providing feedback on the final product that students submit. We have also showed new approaches to help students reflect on their writing (Calvo & Ellis, 2010) and how students understand the use of these new tools. Concept map visualization fall within this design strategy where the tools help students reflect about their own writing, helping them develop metacognitive skills.

Mathematically, a CM is defined as a triplet $CM = \{C, R, T\}$ where C is a set of concepts, R a set of relationships between concepts, and T is the map's topology or spatial distribution of the concepts. The CM mining process can be expressed as the identification of a concept map CM from a document D. This process has three steps: Concept Identification (CI); Relationship Identification (RI); and Summarization, which reduces the size of the CM by keeping only the most relevant concepts. As both concepts and linking words come from a single document, D itself defines all potential words (or phrases) that could become part of the CM. This can be formalized by defining a document as a duplet $D = \{C_d, R_d\}$ where C_d corresponds to all the concepts, and R_d corresponds to all the relationships that explicitly appear in the text. The whole process is summarized in *Figure 1*.

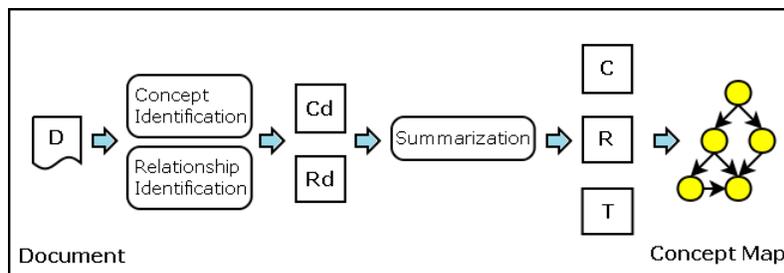


Figure 1. CMM process

Both Concept and Relationship Identification are implemented as operations in TML, using the extracted terminological maps with all terminological map rules applied to obtain a reduced map. The new Concept Identification operation extracts all vertices in the map corresponding to nouns or compound nouns, instead of extracting them from a grammar tree. The Relationship Identification (RI) operation uses the terminological map and a set of concepts as input (the results of a Concept Identification operation). RI has been implemented by finding the shortest path between each pair of concepts, using Dijkstra's algorithm in the terminological map. Redundant connections were avoided by discarding paths containing concepts, and paths that were equivalent but in an opposite direction. Remaining paths are used to identify relationships, and the relationship labels (linking words) are obtained as the concatenation of the words in the path. Some dependencies such as conjunctions and prepositions also include words, which were cleaned and then included. The final summarization operation was implemented using Latent Semantic Analysis (LSA), as suggested in our previous work. LSA is a statistical technique used to find patterns on big corpora (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), that was found to be useful to identify the relative importance of terms within a document (Villalon & Calvo, 2009). A complete CM for a sample essay is shown in Figure 3.

Relationship Extraction and CMM

The quality of automatic concept maps depends on the quality of the knowledge extraction techniques used. Measuring the performance of these techniques first requires that a group of human annotators build a 'gold standard' corpus with annotations that have high inter-coder agreement and are compared to those extracted automatically. Identifying knowledge in text is a subjective task which involves the judgment of humans in the face of ambiguity, such as when a word has many meanings, whether a sentence should be part of a summary, or if a phrase had a particular intention. Under the assumption that humans can reliably make such interpretations, subjectivity is tackled by using corpora annotated by two or more human coders who are required to identify the relevant pieces of knowledge in a corpus. Inter-coder reliability is the upper bound for accuracy for any automatic algorithm trying to identify the same set of knowledge (Artstein & Poesio, 2008).

Data

A set of essays (N=43) collected as a writing proficiency diagnostic activity for first year-university students were used to create the benchmarking corpus. In writing the essays, students first read three short papers on the topic of *English as a global language*, and then answered two questions: Is English becoming a global language? Is this a positive development? The essays had an average length of 468 words; the corpus had a total of 18,431 words.

A first version of the benchmarking corpus was annotated by two coders using a web based annotation tool described in (Villalon & Calvo, 2008), and used to measure the accuracy of our first version of the Concept Extraction algorithm proposed in (Villalon & Calvo, 2009), we refer to this corpus as the first annotation trial. Even though this corpus was useful for extracting concepts, it presented a very low inter-human agreement for relationships, the main problem found was that coders created relationships that were not explicitly present in the essay, but were an interpretation of several propositions. As each coder interpreted the text based on her own knowledge, the already subjective task of summarizing an essay using a CM became even more subjective, resulting in poor inter-rater agreement. To improve this, the protocol was modified, as described below. The tool was also modified to verify this at addition time. Two new human coders using the modified tool and the new protocol performed a second trial that improved inter-rater agreement. Quantitative data on this improvement is reported in the results section. The process averaged 23 minutes per CM, and produced a total of 440 concepts and 258 linking words per coder.

Annotation Method

To annotate the benchmarking corpus for the evaluation of the CMM, the human coders need to follow the same protocol (Artstein & Poesio, 2008). We followed Novak's (2007) method for constructing good and generic CMs that reflect the understanding contained in a document. The first step consists on stating a good focus question that will guide following steps. Then, a list of the most relevant concepts is created; concepts in this list must be ranked from the more general/inclusive to the more specific. Novak refers to this list as "the parking lot", from where

concepts are taken and put in the concept map one by one, starting from the most general ones, and linking each one to previous concepts when needed, forming good propositions. He explains that previously chosen concepts can be discarded if they do not provide new information to the map (Novak, 2007).

The annotation protocol can be summarized as follows:

- Identify the focus question that the essay is trying to answer.
- Make a list of concepts that are the most relevant in the document using words from the document, including spelling and grammar errors.
- Rank the concepts from the most general to the most particular according to what the document explains.
- Begin the concept map with 1 to 4 of the most general concepts.
- Choose explicit linking words to relate the concepts; the words must have appeared in the document and must relate the two concepts. To ensure that the concepts are related, both must appear in the same paragraph.
- Add concepts and relationships until the knowledge expressed in the document is accurately represented in the concept map.
- Reposition the concepts if necessary, to reflect the different levels of generalization of the concepts expressed in the document, in the concept map.

An analysis of the resultant Gold Standard was performed to validate that the CMs were compliant with their educational requirements. For Educational utility, all CMs had concepts and labeled relationships and only very few of them had isolated concepts, human annotators reliably reported that such cases represented isolated ideas in the essay. For simplicity, all CMs were of a considerable smaller size than their corresponding essay, even though the number of concepts used correlated positively with the length of the essay, the number of relationships (inter-connected concepts) did not correlate with the document extension. Finally, for subjectivity, the essay grades showed no correlation with the CMs size, measured in concepts and relationships. This result showed that poorly written essays can be represented by a CM even if they reflect poor writing skills. A more detailed discussion on how humans summarize text using CMs can be found in (Villalon, Calvo, & Montenegro 2010).

Comparative Measures for CMs

Two quantitative measures that can be applied to compare CMs have been proposed in the literature: The Kappa statistic that measures the extent to which two coders agree above chance, and Information Retrieval for Ontology Learning from Text (OLT) measures (Dellschaft & Staab, 2006). A measure first proposed in the information retrieval literature, corresponds to the lexical layer, which is the identification of 'concepts' in the text. These methods are evaluated using the Lexical term Precision (LP). The hierarchical layer corresponds to relationships between concepts. Several measures have been proposed and discussed for this, and the current state of the art corresponds to Taxonomic Overlap Precision (TP) based on the common semantic cotopy proposed by Dellschaft and Staab (2006). LP measures how well the learned lexical terms (concepts for CMM) cover the target domain and TP measures how well the concepts are arranged in the hierarchy defined by an ontology (Dellschaft & Staab, 2006). An important quality of such measures is that they are influenced only by one dimension, and therefore affected by one type of error only. As Dellschaft and Staab point out "if one uses measures for evaluating the lexical term layer of an ontology (e.g. the lexical precision and recall) and one also wants to evaluate the quality of the learned concept hierarchy (e.g. with the taxonomic overlap), then a dependency between those measures should be avoided." (Dellschaft & Staab, 2006) In order to achieve this, the common semantic cotopy must be used to evaluate TP only for those concepts that are common to both maps.

Results

The results of the inter-human agreement show that the modified annotation protocol improved its reliability. Inter-human agreement was measured for the two annotation trials. In the first trial, making a list of the most important concepts, showed an average of 62% agreement for LP and 27% for TP. This shows that humans tend to agree when choosing the most relevant concepts in a document; however, they tend to disagree about linking concepts if given too much freedom. The second annotation trial showed an average of 77% agreement for LP and 85% for TP, indicating a small increase in agreement on concepts and a large increase in agreement on relationships. This increase cannot be caused by chance, because the number of possible linking words for two concepts are only

restricted to all words (not just a category like nouns or verbs) appearing in the same paragraph. Therefore, the possibility of connecting two concepts in the same way by chance is approximately 0.

The automatically generated CMs were compared to the human generated gold standard, showing good results that are affected by the summarization rate. As summary visualizations, an automatic CM should contain only the 'most important' information in an essay. The main characteristic of summaries is their 'size', which is defined as a percentage of the total text units (units can be any subpart of the text, such as sentences, words, terms, paragraphs). For example, a CM reduced to 10% will contain only 10% of the total concepts identified and all relationships between those concepts. In the Concept Map Miner tool, the strategy for summarization was to apply LSA to each essay, forming single document semantic spaces that sort topics within the essay by their relative importance. *Figure 4* shows the effect of summarization on agreement measures.

A high lower bound for LP of 50% validates LSA as a good strategy for summarization. LP grows rapidly reaching 80% with CMs of less than half their original size. Summarization causes a reduction of up to 20% on TP, this results are good because improvements on relationship extraction should improve results even at strong summarization ratios. Finally, LSA does not use the information about the CM, making connected concepts to be removed if they do not provide information, and causing all its relationships to disappear. An approach that would consider both the information provided and the original CM and its connections while summarizing, may provide better results.

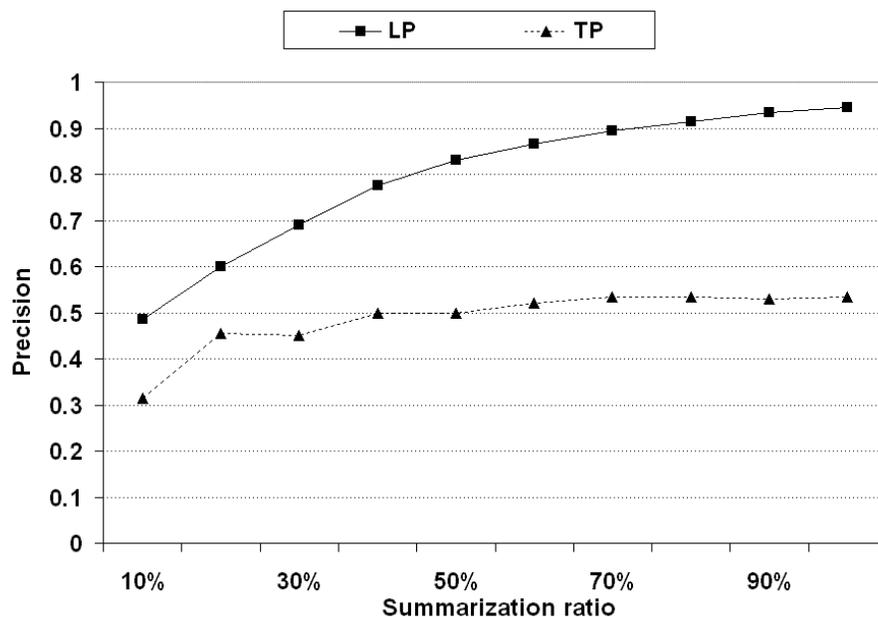


Figure 4. Lexical and Taxonomical Precision versus Summary size

Integration of CMM as Writing Support Tool

Cognitive Visualizations provide quality feedback because they make the author's thinking visible, making explicit the mental model learners are using. CVs can also function as metaconceptual scaffolds, allowing learners to develop an awareness of their own mental representations and inferential processes (Jacobson, 2004). Within the writing process, providing automatic feedback in the form of CVs allows learners to reflect on their own work and their own mental models that guided its construction, facilitating the development of metacognitive skills. One way to better support the writing process is by improving feedback's availability through automatic feedback, in this way the learner can obtain such feedback as many times as she requires, therefore allowing metacognitive skills to develop through the writing process. This is a common design principle in the set of tools (a.k.a. Glosser) that our research group has been developing by using TM techniques to provide automatic feedback to students (Villalon, Kearney, Calvo, & Reimman, 2008). By integrating CMM into our Glosser tool (Calvo and Ellis, 2010) and using cloud computing technologies (Calvo et.al., 2011) students can access feedback on the current state of their assignment at

any point in time. A common learning activity using CMM to support essay writing would be: A teacher uploads readings to the University's Learning Management System and creates an assignment, students then write their essays in the Google Docs tools, with the documents produced by the assignment manager in iWrite. Before (or after) the deadline students can use CMM or any of the other Glosser feedback tools. This is a significant improvement over having to wait a few weeks to receive feedback from the instructors.

CMM is designed to be used in real writing activities aimed at helping a student learn about a topic, rather than focused on communication as a standalone skill. In many writing activities students are asked to write an essay about a topic. This activity can take from a few hours to a full semester. Using CMM, at any point in time, the student can see a CM of the composition and evaluate if those concepts and relationships were what he expected (and what is expected from him). CMM can also be used by the instructors, as other studies have shown that visualizations of the essays can improve the time and reliability of human assessors (O'Rourke, Calvo, McNamara, 2011).

Home Flow Topics **Concept Map** Language Participation
Different forms of feedback

Concept Map

- Do you think the most relevant concepts you covered in the essay are present in the map?>
- Could you improve the Concept Map by adding or removing concepts and relationships?
- Do you think someone could understand your argument only by analyzing this map?

To help you reflect on these questions, Glosser has attempted to extract a **concept map** that represents a summary of your essay, this was done by extracting the most relevant topics and connecting them through the sentences in which you wrote about them.

Revision: 0-15 |

Concept map ('Gloss') as a form of cognitive visualization

Customizable trigger questions

Guidelines/helps

Revision being analyzed

Figure 5. Concept Map Miner showing an automatic CM within Glosser

We have integrated CMM into the Glosser environment for automatic feedback. Glosser's designed is aimed at provoking reflection in students during the writing process and details on its pedagogical design are discussed in (Villalon, Kearney, Calvo, & Reimman, 2008). The feedback is in the form of descriptive information about different aspects of the document. For example, by analyzing the words contained in each paragraph, Glosser can measure how close two adjoining paragraphs are. If the paragraphs are too far this can be a sign of a lack of flow or what is called lexical in cohesiveness and Glosser flags a small warning sign. Glosser provides feedback on several aspects of the writing: structure, coherence, topics, keywords. Each of these areas is identified by a tab on the homepage of the web application as shown in Figure 5. The Questions section contains triggers to prompt student reflection on a particular set of features of the composition, and can be customized for each course or activity. The instructions section describes how to interpret the feedback, which is provided as 'gloss' on the lower half of the page. The feedback is an alternate type of visual or textual representation of features that have been automatically sourced from

the text. The history of versions provides a record of the writing process, which includes information on the time of each revision and the author responsible for it.

Another way to improve the quality of feedback in the writing process is by improving the quality of the CVs used as feedback, and CMMs can provide such quality. As Jacobson argues: “One technique for doing this (making thinking visible) is the use of concept maps to represent knowledge” (Jacobson, 2004). However, until now it was impossible to automatically extract such a CV from students’ text. Using the CM to visualize their own mental models, students can compare them to a model map provided by the teacher (as in cognitive apprenticeship) or revise their writing so the CMM tool can clearly identify the parts of the model it could not automatically extract. Reflection for revising is scaffolded in Glosser by providing questions for the student to reflect when analyzing the CM.

The CMM integration consisted on a new tab labeled “Concept Map” that shows the generated CM, and presents the questions that can be customized by each instructor:

- Do you think the most relevant concepts you covered in the essay are present in the map?
- Could you improve the Concept Map by adding or removing concepts and relationships?
- Do you think someone could understand your argument only by analyzing this map?

Conclusion

Motivated by the pedagogical value of Concept Maps as cognitive visualizations, this paper presents a new Concept Map Mining (CMM) tool, and evaluates it using a collection of human annotated essays written by undergraduate students. The results show that the automatic generation of CMs from documents is feasible, despite the complexities of noisy data such as student-produced text. Humans tend to agree when summarizing a document using a CM (inter-coder agreement of 77% for Lexical Precision and 85% for Taxonomical Precision). Available OLT tools and the CMM reliably identified concepts, averaging 94% for LP with human coders.

The tool has been integrated into an e-learning environment, and future work involves designing pedagogical activities in which the impact on student learning will be assessed. As the motivation of the CMM is to support writing activities, it was also integrated into a tool for enhanced feedback on writing activities called Glosser (Villalon, Kearney, Calvo, & Reimman, 2008; Calvo and Ellis 2010). The tool supports writing activities by scaffolding authors' reflection during the process of writing, to encourage them to revise their work. At any time during their writing authors can *gloss* their documents by clicking a link that will take them to the Glosser website. Glosser can then present several issues related to the quality of the essay with provoking questions on how to assess them. Automatically generated features from the text are also presented for the student to guide and facilitate the answering of the questions.

Acknowledgment

The authors would like to thank Stephen O'Rourke for his contribution to this project and Anindito Aditomo for feedback on the manuscript. This project has been funded by an Australian Research Discovery Grant DP0665064, a Norman I Prize and a Chilean Bicentennial Scholarship.

References

- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34, 555-596.
- Beals, T. J. (1998). Between teachers and computers: Does text-checking software really improve student writing? *English Journal, National Council of Teachers of English*, 87(1), 67-72.
- Calvo, R. A., & Ellis, R. A. (2010). Students' conceptions of tutor and automated feedback in professional writing. *Journal of Engineering Education*, 99(4), 427-438.
- Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2011). Collaborative Writing Support Tools on the Cloud. *IEEE Transactions on Learning Technologies*, 4 (1), 88-97.

- Chang, K. e., Sung, Y.-T., & Chen, I.-D. (2002). The Effect of Concept Mapping to Enhance Text Comprehension and Summarization. *The Journal of Experimental Education*, 71 (1), 5-23.
- Chen, N.-S., Kinshuk, P., Wei, C.-W., & Chen, H.-J. (2008). Mining e-Learning Domain Concept Map from Academic Articles. *Computers & Education*, 50, 694-698.
- Clariana, R. B., & Koul, R. (2004). A computer-based approach for translating text into concept map-like representations. In Cañas, A. J., Novak, J. D. & González, F. M. (Eds.) *Proceedings of the First International Conference on Concept Mapping*. Pamplona, Spain.
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive Apprenticeship: Making Thinking Visible. *American Educator*, 6,38-46.
- Collins, C., Carpendale, S., & Penn, G. (2009). DocuBurst: Visualizing Document Content using Language Structure. *Computer Graphics Forum*, 28 (3), 1039-1046.
- Deerwester, S., Dumais, S., Furnas, G., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41 (6), 391-407.
- Dellschaft, K., & Staab, S. (2006). On How to Perform a Gold Standard Based Evaluation of Ontology Learning. *Paper presented at the 5th International Semantic Web Conference*, November 5-9, Athens, GA, USA.
- Emig, J. (1977). Writing as a Mode of Learning. *College Composition and Communication*, 28, 122-128.
- Graesser, A. C., & McNamara, D. S. (in press). Use of computers to analyze and score essays and open-ended verbal responses. In H. Cooper, P. Camic, R. Gonzalez, D. Long & A. Panter (Eds.), *APA handbook of research methods in psychology*. Washington, DC: American Psychological Association.
- Hauser, S., Nuckles, M., & Renkl, A. (2006). Supporting concept mapping for learning from text. In Barab, S., Hay, K. & Hickey, D. (Eds.) *Proceedings of the 7th international conference on Learning sciences* (pp. 243-249). Bloomington, IN: ISLS.
- Jacobson, M. J. (2004). Cognitive visualisations and the design of learning technologies. *International Journal on Learning Technology*, 1, 40-60.
- Keh, C. L. (1990). Feedback in the writing process: a model and methods for implementation. *English Language Teaching Journal*, 44, 294-304.
- Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, 47 (3), 43-62.
- Marneffe, M.-C. d., MacCartney, B., & Manning, C. D. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J. & Tapias, D. (Eds.) *Proceedings of the 5th international conference on Language Resources and Evaluation* (pp. 449-454), Genoa, Italy.
- McAleese, R. (1998). The Knowledge Arena as an Extension to the Concept Map: Reflection in Action. *Interactive Learning Environments*, 6, 1-22.
- McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36, 475-492.
- Moore, T., & Morton, J. (1999). Authenticity in the IELTS academic module writing text, *IELTS Research Reports*, retrieved June 20, 2011, from, http://www.ielts.org/PDF/Vol2_Report4.pdf.
- Novak, J. D. (2007). *The theory underlying concept maps and how to construct them*. Retrieved June 10, 2011, from <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>.
- Novak, J. D., & Gowin, D. B. (1984). *Learning How To Learn*. Cambridge: Cambridge University Press.
- O'Rourke, S. Calvo R.A. & McNamara (2011). Visualizing Topic Flow in Students' Essays. *Educational Technology and Society*, 14 (3), 4-15.
- Oliveira, A., Pereira, F. C., & Cardoso, A. (2001). Automatic Reading and Learning from Text. *Paper presented at the International Symposium on Artificial Intelligence*, December 18-20, Panhala, India.
- Paltridge, B. (2004). Academic writing. *Language Teaching*, 37, 87-105.
- Thiesmeyer, E. C., & Thiesmeyer, J. E. (1990). *Editor*. New York: Modern Language Association.
- Valerio, A., & Leake, D. (2006). Jump-starting concept map construction with knowledge extracted from documents. In Cañas, A. J. & Novak, J. D. (Eds.) *Proceedings of the Second International Conference on Concept Mapping* (pp. 296-303). San Jose, Costa Rica.

- Villalon, J., & Calvo, R. A. (2008). Concept Map Mining: A definition and a framework for its evaluation. *Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, December 9-12, Sydney, Australia.
- Villalon, J., & Calvo, R. A. (2009). Single Document Semantic Spaces. *Paper presented at the Australian Joint Conference on Artificial Intelligence*, December 1-4, Melbourne, Australia.
- Villalon, J., Kearney, P., Calvo, R., & Reimman, P. (2008). Glosser: Enhanced feedback for student writing tasks. *Paper presented at the International Conference on Advanced Learning Technologies*, July 1-5, Santander, Spain.
- Villalon, J., Calvo, R. A., & Montenegro, R. (2010). Analysis of a gold standard for Concept Map Mining – How humans summarize text using concept maps. *Proceedings of the Fourth International Conference on Concept Mapping* (pp. 14-22). Viña del Mar, Chile.
- Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, 22(3), 333-362.
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language teaching research*, 10 (2), 157-180.
- Wattenberg, M., & Viegas, F. (2008). The Word Tree, an Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14 (6), 1221-1228.
- Wiemer-Hastings, P., & Graesser, A. C. (2000). Select-a-Kibitzer: A Computer Tool that Gives Meaningful Feedback on Student Compositions. *Interactive Learning Environments*, 8(2), 149-169.
- Zouaq, A., & Nkambou, R. (2008). Building Domain Ontologies from Text for Educational Purposes. *IEEE Transactions on learning technologies*, 1(1), 49-62.