Towards App-based Formative Feedback to Support Summarizing Skills

Peter van Rosmalen, Liesbeth Kester, and Els Boshuizen

Centre for Learning Sciences and Technologies (CELSTEC), Open Universiteit Nederland, Heerlen, The Netherlands
{peter.vanrosmalen, liesbeth.kester, els.boshuizen}@ou.nl

Abstract. Secondary education students have difficulty to comprehend text, let alone, hypertext. Summarizing is an effective strategy to improve text comprehension. It enables students to link the text content to existing prior knowledge, promotes self-testing which helps them to identify their comprehension gaps and fix them and directs students' attention to important content parts. However, summarizing takes skill that secondary education students often lack. This paper discusses the design of an app which aims to enhance summarizing skill acquisition and, hence, text comprehension of secondary education students by providing just-in-time, formative feedback as part of summarization activities. The app discussed will offer a formative assessment of a student's summary through visualization of salient aspects of it as compared to a peer's or teacher's work with additional guidance. Visualisation and guidance will be highly automated thus easing access and use in real practice. It builds on prior, recent research, showing that automatically created visualisations can be used to support writing.

Keywords: formative feedback, summary writing, language technology, technology enhanced learning, text comprehension, visualisation

1 Introduction

Secondary education students, especially those on the preparatory-secondary-vocational-education level, have difficulty comprehending expository text (Land, 2009). This problem is even more prominent for comprehending multiple-source learning-material, such as, hypertext which is frequently used nowadays (Rouet, 2006). Summarizing is a highly effective strategy to improve text comprehension (e.g., Friend, 2001). It is a cognitive process of extracting the most important information from a text and paraphrase it in a concise form (Beesley & Apthorpe, 2010). Firstly, it enables students to build relations between concepts in a text and connect these concepts and their relations to existing prior knowledge (i.e., elaboration). Secondly, it promotes self-testing which helps students to become better aware of comprehension gaps which might stimulate them to close these gaps and thirdly, it directs students' attention to important parts of the text (Thiede & Anderson, 2003). However, it takes skill to write a good summary. Often, students lack this skill (Graham &

Perin, 2007). This prevents them from taking advantage of this learning strategy for text comprehension. Training could overcome this problem. Teaching students a summarizing strategy helps them to develop good summarizing skills (Graham & Perin, 2007).

Several reasons exist to look into modern, mobile technology to foster summarising skills:

- It closely aligns with the increased use of hypertext material. The use of technology in education naturally embraces hypertext as opposed to single source linear text book material. Nevertheless, so far only few studies investigated the effects of summarization on hypertext comprehension (e.g., Gil, Bråten, Vidal-Abarca & Strømsø, 2010a). In order to comprehend hypertext, students need to "... to locate, evaluate, and use diverse sources of information for the purpose of constructing and communicating an integrated, meaningful representation of a particular issue, subject or situation." (pp. 157-158; Gil et al., 2010a). Gil and colleagues (Gil, Bråten, Vidal-Abarca & Strømsø, 2010b) found that a summarization instruction supported this process and led to a better hypertext comprehension.
- It makes it possible to deliver such a training just-in-time which might improve learning even more (Kester, Kirschner, van Merriënboer, & Bäumer, 2001).
- It may open up to (partly) automating the training and guidance required. Guiding hands-on practice in summary writing and offering supportive feedback tends to be a time consuming tasks. Assessment of student work has been rated to be a student support activity which easily leads to staff work overload (van Rosmalen et al., 2008).

Rule-based summarization training has been successfully applied to develop the student's summarizing skills. Such a training teaches the following summarizing rules (e.g., Bean & Steenwijk, 1984): 1) deleting unnecessary or trivial material, 2) deleting material that is important but redundant, 3) substituting a superordinate term for a list of items, 4) substituting a superordinate term for components of an action, 5) selecting a topic sentence and, 6) inventing a topic sentence if there is none. Graham & Perin (2007) identified three conditions that have to be met to enable students to independently use a writing strategy that is instructed to them: 1) a modelling example or worked example of how to use the strategy should be shown to the students, 2) the instruction should be given over a longer time period (i.e., at least three days), and 3) the instruction should be delivered according to a scaffolding principle, that is, the instructional support should gradually fade to help students independently use it. Moreover, Beesley and Apthorpe (2010) put forward that summarization training might work best in combination with other instructional interventions.

Additional interventions strengthening summary writing and text comprehension could, for instance, focus on:

- *Prior knowledge activation*. Even simple instruction to activate prior knowledge can help students to comprehend and learn from text (see Machiels-Bongaerts, Schmidt, & Boshuizen, 1995; Wetzels, Kester & van Merriënboer, 2011).

- *Self-testing*. Self-testing consists of generating and answering questions during reading a text. It aims to enhance summarization by supporting the identification of comprehension gaps. It seems that answering questions helps students to more accurately judge their own learning, or in other words, identify their comprehension gaps (Dirkx, Kester & Kirschner, 2012).
- *Visualisation* (i.e., the presentation of visual representations of students' summaries). Berlanga, Van Rosmalen, Boshuizen, and Sloep (2012) explored and compared the use of automatically generated concept maps and word clouds to give formative feedback on verbal assignments. Word clouds of students' writings were used as visual tools to discuss writing development and lexical acquisition in foreign language writing (Brydon-Miller, Greenwood, & Maguire, 2011).

In this paper, we will explore the latter, i.e. the use of visualisations in combination with regular summarisation training. It builds on the premises that summarizing training improves students' summarizing skills and thus, their comprehension of both text and hypertext and, next, that supporting the meta-cognitive processes involved in summarizing through the app-based guidance with the help of visualisation, will improve summarizing skill acquisition during summarization training. Moreover, that by offering the guidance app-based it should be possible to deliver its' support just in time, make it easily accessible and to economise its use by making use of various technologies which can automate the creation and use of the visualisation to a high extend.

In the following sections we will first review the background of the intended app. Next, we will describe the initial prototype and how it builds on prior research experience. We will close with a discussion and our plans for future work.

2 App-based guidance with the help of visualisations

Summarizing is a verbal reporting method. Whereas the instructions to summary writing can be delivered following a scaffolding approach with stepwise fading support. It still may require additional interventions and one or more detailed formative assessments of the summary created. Hitherto, it has been very laborious and complex to analyse verbal data and subsequently to give feedback. However, some initiatives are taken to change this. For example, Shute et al. (2009) report on HIMATT, a family of tools that produce visualisations to provide students and teachers information on how well the students conceptualise a content area. One of the tools, MITOCAR (Pirnay-Dummer, 2006), parses natural language to extract the most frequent concepts and analyses these to derive graphical models. Furthermore, in recent work automatically created concept maps have been used to support the writing process. Villalon, & Calvo (2011) provided a concept map as a form of scaffolding so students can see their composition and evaluate if their concepts and relationships are what they expected. Reategui, Klemann, & Finco (2012) give a map of a text to show the main idea as starting point before writing a summary. Berlanga et al (2012) discuss a number of these approaches including other less technically demanding options such as word clouds. The latter being of particular interest since the creation of word clouds

and alike do not depend of a large corpus of sample texts and extensive training or specific expertise to be used and can be done with commonly available Natural Language Processing software.

WritLe, the app discussed below, aims to enhance summarizing training by providing graphic knowledge visualisations to help students identify important content parts. Visualisations or graphic knowledge representations are graphical overviews of someone's knowledge that are directly (e.g., concept maps) or indirectly (e.g., pathfinder nets) derived from a knowledge assessment. Jonassen, Beissner, & Yacci (1993) describe a large set of methods to assess knowledge including verbal reporting methods (e.g., think aloud, answering essay or other questions, summarizing). Hitherto, it has been very laborious and complex to analyse verbal data. However, the availability of language technologies is changing this. These techniques are now capable to condense learning content or a person's knowledge state into a visualisation of the most salient aspects of this knowledge. Visualisations of learning content or an expert's knowledge can be used as standard to compare a student's visualization to. In this way, the important content parts a student missed can be identified and the visualisations become a useful tool to support meta-cognitive activities.

Comparison of the most commonly used visualisation, concept maps, is also for teachers (i.e., experts), a difficult task which has to take into account differences in layout and nomenclature to define the concepts and relations (De Souza, Boeres, Cury, De Menezes, Carlesso, 2008). Word clouds, as compared to concept maps are relatively simple visualisations, they mainly focus on content and therewith are less complex to compare. Berlanga, Van Rosmalen, Boshuizen, and Sloep (2012) explored and compared the use of automatically generated concept maps and word clouds to give formative feedback on verbal assignments. Their study indicated that relatively simple visualisations such as word clouds, which can be generated with widely accessible tools, adequately cover the original text. From this study, no firm conclusions about the use of word clouds as meta-cognitive learning tools can be drawn. Research on this purpose of word clouds is still limited. Partly related examples are, for example, the use of word clouds as navigation tool to support a web search (Gottron, 2009); foreign language writing, that is, word clouds of students' writings were used as visual tools to discuss writing development and lexical acquisition (Brydon-Miller, Greenwood, & Maguire, 2011) and exploratory data analysis, that is, word clouds to compare documents of two studies on a single issue (Cidell, 2010). The app discussed will be used to study and to find out more about how to use word clouds as meta-cognitive tools for learning, or in other words, how relatively simple visualisations of verbal reports can be used to identify important content parts and regulate further learning.

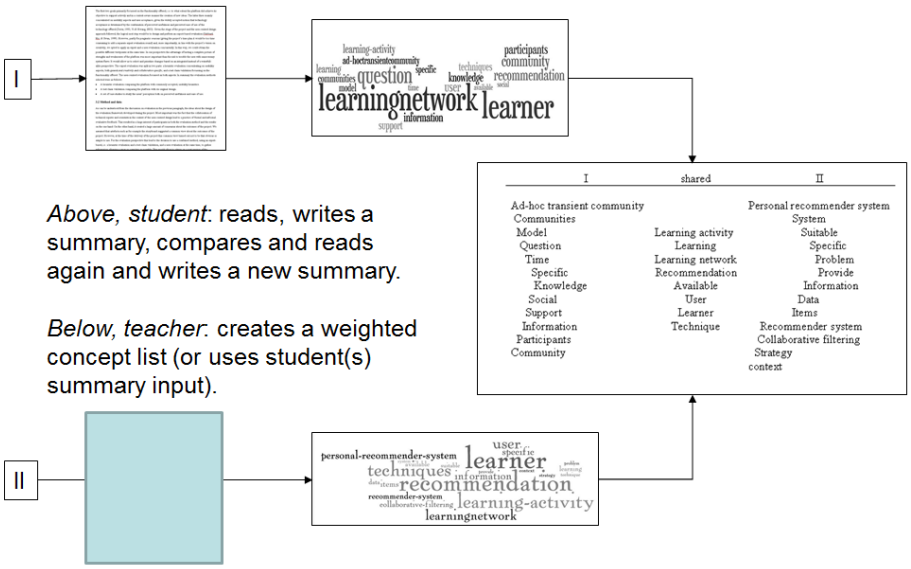

Fig. 1. Outline of the workflow. Step 1: the summary of the student (or an equivalent), step 2: concept (single word or multiple word concepts) extraction and frequency count, step 3 visual comparison.

3 WritLe, a first prototype

3.1 Design considerations

Following the introduction above, the design of the proposed app is grounded in the idea that an intervention offering an formative assessment of a student's summary through visualization of salient aspects of their summary with additional guidance, on top of a summarisation training, is of great value. Even while following a training in many cases for students it is difficult to determine the scope and quality of their summary. Actively writing (or summarizing) on a course subject is an good approach to see what one understands. However, to be able to do so it is essential for students to get an assessment of one or more versions of their writing and actively learn how to improve it themselves. Unfortunately, a formative assessment is relatively rare given the scarce resources of teachers. The app proposed considers:

- The scope and quality of a student's summary is reflected by which concepts they use;
- Both use of app and visualization do align well with the experiences and preferences of secondary education students;
- Visualization, directing attention to the summary under construction, together with guidance can actively involve students in their learning process
- Students can be provided with diverse ways of comparing their level of performance.

Based on prior work (Berlanga et al., 2012), we will investigate the use of word clouds alike (Figure 1) to show students which content is most prominently present in their summary 'under construction'. Different versions of this app will be developed and tested, for example, one version may present students a visualisation of their summary in contrast to a visualisation of a summary of their peers (i.e., a peer reference model), the other presents a visualisation of their summary in contrast to a visualisation of the text or hypertext (i.e., an expert reference model). To compare and assess knowledge both peer an expert reference model are in use. Steinhart (2011), for example, uses a collection of peer summaries to establish a golden standard. Shute et al. (2009) use both peer reference models and expert reference models depending of the context and the tool used. Domain novices and naïve students might benefit most from 'student - peer' visualisation comparisons, as at this stage a peer visualisation and their vocabulary would correspond most to their Zone of Proximal Development (Vygotsky, 1978). As expertise develops, the peer visualisations may still be appropriate, depending on the development stage of the peers, but content visualisations representing the 'expert knowledge state' may be more suitable to more advanced students. However, alternatively it can be argued that content visualisations will correspond best with the original text and the presence and absence of relevant features of this text and therefore better suited to compare to. So far, the effectiveness of these two approaches has not been contrasted as is the case in this design.

The word clouds that will be used, will improve on regular word clouds such as Wordle (www.wordle.net), they will take into account, for example, bending of words and multiple word concepts (Kaptein, Hiemstra, & Kamps, 2010) and use advanced visualisations i.e. word clouds that integrate and contrast two independent word clouds. The source of the independent word clouds may vary between a summary of a student, a previous version of a the summary, a summary or a group of summaries of peers of the students, the original text studied, an expert summary or a frequency tagged list of key concepts prepared by the teacher.

The idea behind the guidance is that students use the visualisation to challenge them to think about strength and weakness of their text. The visualisation prompts them to their key concepts, the key concepts they share and the key concepts of the other text. Questions to be answered by the student are, for example, (1) identify and map synonyms; (2a) motivate why you *did not* mention concepts of the second text or (2b) why you *did*, (3) identify trivial, irrelevant concepts (4) identify substitutes i.e. a concept replacing a set of concepts (5) identify look-up concepts to be studied. The final app may be implemented as a game, a collaborative task, with or without scaffolding to guide the interactions and with one or more rounds depending on the overall summarisation training.

Finally, in our case, we aim to build the first full prototype in Dutch. For the Dutch language, software such as Termtreffer (http://www.inl.nl/tst-centrale/nl/over-de-tst-centrale/projecten/termtreffer) or Alpino parser (http://www.let.rug.nl/vannoord/alp/Alpino/) are available to support in the required linguistic parsing to automatically extract the terms of the text. The current first prototype –as will be discussed below- has been developed in English.

3.2 WritLe, the first prototype

Taking into account the design considerations we designed and successfully prepared a first functional prototype concentrating on two of the main aspects discussed above:

- The application should be able to automatically extract the concepts of a text, both single and multiple word concepts and sort them on frequency and compare them with another text.
- The application should be able to visualise the differences in a Wordle-alike format.

WritLe, the resulting application, goes through four main step to produce a visualisation (Figure 2). It has been build in Python with the help of public available libraries (including pytagcloud, pygame and pyglet):

Step 1 Input. The two input files to be compared are read. As discussed earlier, the inputs can vary e.g. an essay of student 1 and student 2; or of student 1 and the teacher; or of student 1 and a grouped text of a number of students.

Step 2 Parsing. In step 2 the input is parsed. This includes the removal of 'irrelevant' words (so called stop words), determination of the nouns (the concepts in the text), mapping plurals to their singular form (so book and books are mapped onto book), identify clusters (n-grams) of words which point to one concept (so e.g. secondary school 'secondary school' or learning network 'learning network') and finally counting the concepts and sort them on frequency. As an intermediate result WritLe returns a sorted list with for all concepts the pair (concept, frequency), e.g.: (learning networks, 16); (essay, 12); (school, 4); (secondary school, 2). The parsing can be tuned by for instance adjusting the maximum number of concepts or the cluster (n-gram) length.

Step 3 Comparing. In step 3 the two input files are compared and sorted with regard to most frequently shared concepts and the most frequent unique concepts of both text 1 and text 2.

Step 4 Visualisation. Finally, the results are visualised where a function of the relative frequency is used for the x-position and the frequency for the size of the visualisation of each concept.

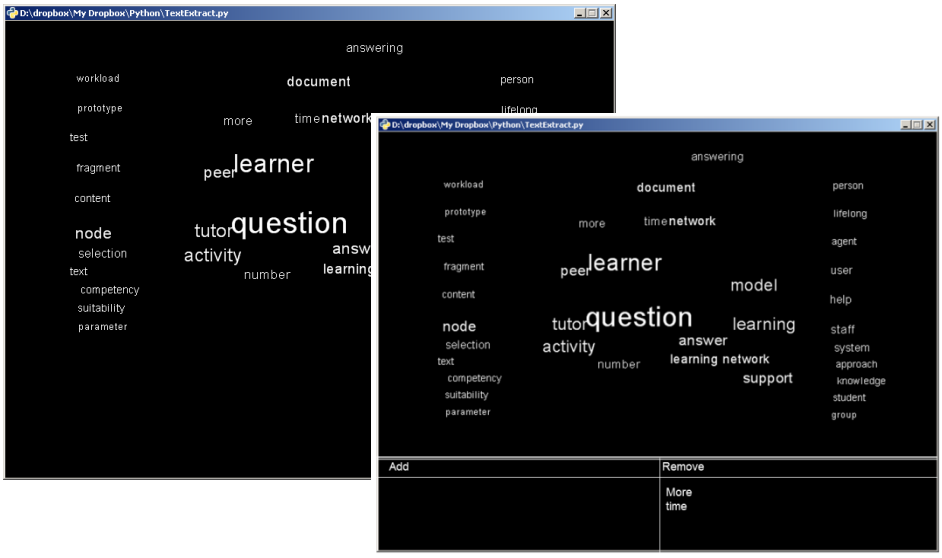

Fig. 2. To *the left*, the output of Writle visualising the differences between two summaries. The x-position indicates in which summary a concept occurs most or exclusively (completely left or right). The size of a concept indicates the relative frequency. *To the right*, a mock-up drawing showing how the visualisation can be enhanced with activities. In this case dragging and pointing out important additional concepts or superficial concepts.

4 Discussion and future work

It is well-established that summarizing text aids text comprehension and that rule-based summarization training helps develop summarizing skills. Summarizing is part of the curriculum and end-terms of each level of secondary education. For students, it is highly relevant to acquire this skill. The app proposed aims to bring this research a few steps further by establishing the worth of summarization training for summarizing hypertext, and thus hypertext comprehension which has not yet been researched; by enhancing summarization training by providing additional guidance through the use of mobile technology and last but not least by offering a highly automated service providing a formative assessment and additional guidance without large efforts of teachers.

We attempted to explain the background of our research both in the area of summarisation and how natural language processing has developed through the last decade now enabling various ways of assessing writing text. Moreover, we argued that word clouds alike, though simple, are of interest for what we want to achieve and we showed how WritLe, our first prototype, used word clouds alike to fulfil our main requirements. Nevertheless, it is obvious that WritLe is still in its infancy. Extensive research will be required to establish how learners can benefit most of WritLe taking

into account questions such as which text to initially compare, which kind of guidance and activities to offer and how to scaffold them best. Taking into account the potential benefits of an automated assessment of students´ their own work as compared to only superficial general rules, we do believe that it is worthwhile to continue on our path in exploring WritLe and its further extensions in real practice.

Acknowledgements

We would like to thank Rik van Rosmalen who did spend his free time to develop the current version of the WritLe prototype.

References

1. Bean, T. W., & Steenwijk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior, XVI*, 297-306.
2. Beesley, A. D., & Apthorp, H. S. (2010). *Classroom instruction that works, second edition: Research report.* Denver, CO: Mid-continent Research for Education and Learning.
3. Berlanga, A.J., Van Rosmalen, P., Boshuizen, H.P.A. and Sloep, P.B. (2012). Exploring formative feedback on textual assignments with the help of automatically created visual representations. *Journal of Computer Assisted Learning, 28*, 146–160.
4. Brydon-Miller, M., Greenwood, D., & Maguire, P. (2011). Using Wordles To Teach Foreign Language Writing. *Language Learning & Technology, 15*(2), 12–22.
5. Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area, 42*, 514–523.
6. De Souza, F. S. L., Boeres, M. C. S., Cury, D., De Menezes, C. S., & Carlesso, G. (2008). An Approach To Comparison Of Concept Maps Represented By Graphs. In: *Proceedings of the Third Int. Conference on Concept Mapping*, Tallinn, Estonia & Helsinki, Finland 2008.
7. Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2012, 22 June). Zelfbeoordelingen bij het lezen van teksten [Judgement of learning when reading texts]. In J. van Merriënboer (Chair), *Het verbeteren van zelfbeoordelingen bij het leren* [Improving judgments of learning when learning]. Symposium conducted at the meeting of the Onderwijs Research Dagen [Educational Research Days], Wageningen, The Netherlands.
8. Friend, R. (2001). Teaching summarization as a content area reading strategy. *Journal of Adolescent & Adult Literacy, 44(*4), 320-329.
9. Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010a). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology*, *35*, 157–173.
10. Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010b). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*, 30–68.
11. Gottron, T. (2009). *Document Word Clouds: Visualising Web Documents as Tag Clouds to Aid Users in Relevance Decisions.* Lecture Notes in Computer Science, 2009, Volume 5714, Research and Advanced Technology for Digital Libraries, Pages 94-105.
12. Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*, 445-496.
13. Jonassen D.H., Beissner K. & Yacci M. (1993). *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge.* Lawrence Erlbaum Associates, Hillsdale, NJ.

14. Kaptein, R., Hiemstra, D., & Kamps, J. (2010). *How Different Are Language Models and Word Clouds? Advances in Information Retrieval.* Last accessed June 2012 at http://www.springerlink.com/index/45547N3256202U81.pdf
15. Kester, L., Kirschner, P. A., Van Merriënboer, J. J. G., & Bäumer, A. (2001). Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior, 17*, 373-39.
16. Land, J. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen.* [Weak readers, strong texts? Effects of text and reader characteristics on text comprehension and text appreciation of preparatory secondary vocational education students.]. Delft: Eburon Uitgeverij.
17. Machiels-Bongaerts, M.,' Schmidt, H. G., & Boshuizen, H. P. A. (1995). The effect of prior knowledge activation on text recall: an investigation of two conflicting hypotheses. *British Journal of Educational Psychology*, *65*(4), 409-424.