Question-answering – Connecting and Supporting the Learner

Peter Van Rosmalen, Peter Sloep, Francis Brouns, Liesbeth Kester, Malik Koné, and Rob Koper,
Open University of the Netherlands. Contact: peter.vanrosmalen@ou.nl.

Abstract— **Tutors have only limited time to support the learning process. In this paper, we introduce a model that helps answering the questions of students. The model invokes the knowledge and skills of fellow students by bringing them together based on the combination of question posed and their study progress and supports them with text fragments selected from the material studied. We will explain how we used LSA to select and support these peers; examine the calibration of the LSA-parameters and conclude with a small practical simulation to show that the results of our model are fit for use in experiments with students.**

I. INTRODUCTION – THE MODEL

The prototype (see also Table 1) of the model (for a detailed description see [1]) consists of five modules: a Moodle learning environment; a wiki; GTP an LSA implementation [2]; GUP to ease the calibration of LSA (GTP Usability Prototype [3]); and ATL (A Tutor Locator [4]) for the selection of the peers who will assist, based on the topic involved and the users' background and performance.

Pre	A course with a set of topics and users with progress profiles.

Main	1. *Anne* poses a question.
	 2. The *system* determines:
	    - the most relevant text fragments (LSA);
	    - the appropriate topics (LSA);
	    - the most suitable users (LSA + user profiles).
	 3. The *system* sets up a wiki with the question, the text fragments and guidelines.
	 4. The selected *users* receive an invitation to assist.
	 5. *Anne and the users* discuss and phrase an answer in the wiki.
	 6. If answered (or after a given period of time) *Anne* closes the discussion and rates the answer.

Post	The answer is stored

Table 1: The main steps of the model.

II. CALIBRATION

The domain of the course we used is 'Internet Basics', a collection of texts, links and tasks that aim to instigate a basic understanding of the Internet [5]. It contains 11 topics, each of which introduces a different aspect of the Internet.. The topics consist of an introduction, exercises, references to external web pages for further study and an assessment. The corpus was extracted manually. It contained the Moodle pages and external web pages; the assessment questions were left out, however. These questions were used to calibrate the model. The documents were used as raw input; this means that no further corrections were applied such as removing irrelevant documents, diacritical signs or misspellings. The final corpus was relatively small. It consisted of 327 documents ranging in size from 50 to 23534 bytes (41 documents smaller than 250 bytes, 50 documents above 3000 bytes). The corpus contained a total of 82986 words divided over 10601 terms, 4440 of which occur in at least two documents.

In addition to the calibration, we investigated if it was possible to define the parameters with a predefined, limited number of steps that can be repeated and automated at a later stage. An overview of applications with LSA [6] revealed that there is no straightforward procedure to determine the LSA parameters. The parameters are influenced by the corpus and the way LSA is applied. We selected the five steps [2], [7] that should be the most important: the definition of a correlation measure and method, corpus preprocessing, normalisation, weighting and dimensionality. We did not carry out, however, an exhaustive test with all possible combinations of parameters. Instead, we started with an initial combination of parameters based on results reported [7]-[8], and in each step, we tested one parameter and continued to the next step using only used the best result(s) (Figure 1).

Correlation measure and method. For our correlation measure, we used cosine similarity. Our method directly follows from our model. First, we used LSA to identify to which topic(s) the question posed fits best. This information is used to identify peers that are competent in the pertinent topic. Second, we wanted to select the three documents that were most suited to assist the peers in answering the question. We combined the two by selecting the three best correlating documents. We used the result of the mapping on the topics to select the parameter combination with which to continue. The questions, 16 in total, were chosen from the original topic assessment questions. Therefore each question should map to one known topic. *Preprocessing* (run 1-3). Because we did not have access to a stemming application for Dutch, we only considered stopping. Moreover, given the size of our corpus, we created our own stop lists based on the term frequency in the corpus [8]. The stop list consisted of the terms that covered 33% (22 terms) and 50% (91 terms) respectively of the overall term frequencies with the exception of corpus specific terms. By way of comparison, we also used a 'general' Dutch stop list (Oracle Text Reference: Release 9.2). For our corpus, this resulted in a reduction of 188 terms. Finally, in each run (until the actual dimensionality step), we

chose to limit the number of singular values to 40% of the sum of the singular values (Wild et al, 2005). *Normalisation* (run 4-5). Next, with a limited number of documents per topic and quite a spread in document lengths we tested the use of normalisation. This has the effect that documents with the same semantic content are ranked equal in the question query. *Weighting* (run 6-8). Subsequently, we applied the three available types of Global Weighting. *Dimensionality* (run 9-10). In the last step, we determined the best value for the dimensionality by comparing the initial value of 40% of the sum of the singular values to 30 and 50%. Finally, in this step (run 11) we did one additional test i.e. we used the 50% stop list in order to check if this would improve our results. The other parameters followed the settings of Run 9. The result was good (15 out of 16) but not an improvement.
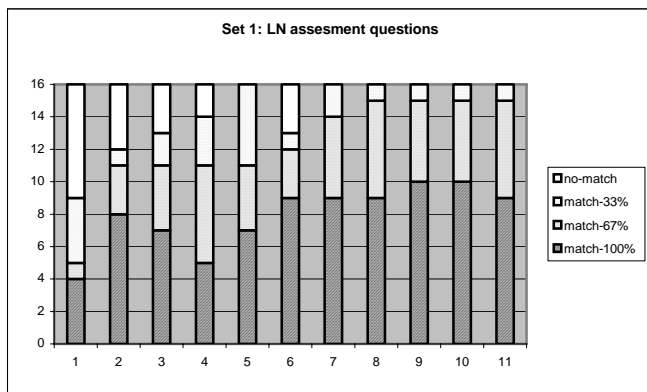

Figure 1: The mapping of the questions on the topics in the calibration runs.

III. A SIMULATION OF THE MODEL

For a final check, we formulated a new set of 16 questions, each connected to one topic. The questions were once again mapped on the topics, and the results were compared with their known topics. The parameter combination of the calibration run 9 and 10 were applied. The model identified the topic correct for 12 out of 16 questions. Case one (the settings of Run 9) did slightly better in the 100% recognition category. For this case we asked two of the designers of the course to rate, on a 5-point scale, the suitability of the text fragments selected through the application of LSA. The suitability of the text fragments is far less accurate; approximately 40% of the questions received one or more fragments rated 3 (5-point scale) or above. The designers of the course, however, indicated that approximately 35% of the questions posed were beyond the scope of the contents of the topic studied; as a consequence the topic did not contain any useful fragments at all. Together the results are promising. The corpus used is rather small, so the chances to find an answer are limited. Also the results may be stepwise improved by making use of successfully resolved questions and their answers. Finally, it is not merely the answer that matters. It is an important aid, but the first concern is to identify the appropriate topic so the right peers can be selected. With 75% recognition we think we are in a good position to achieve this.

IV. CONCLUSION

We introduced a model that intends to help the learner and alleviate the support task of tutors. We described how we calibrated LSA for an existing course. Subsequently, and for the same course, we checked with a simulation whether the model is fit for experimentation with students. In our opinion, the results shown are promising. Moreover, we were able to arrive at our results in a systematic way. The same steps can be followed for a new corpus or if the changes to an exiting corpus are relatively small, the known settings can be reapplied in just one additional run. Obviously, one should be open to retrace one's steps, in particular, if the results are very close (as in the normalisation step) and improvements develop insufficiently.

Clearly, there are still a number of issues to be considered. First, the model has only been applied once and to questions that exactly match one topic. It is fair to expect that, in real practice, part of the questions will cover not just one but more topics. This may complicate the recognition and thus dilute the results. Next, as shown by some of the results, the approach is sensitive to the size (and content) of the available corpus.

ACKNOWLEDGMENT

REFERENCES

[1] Van Rosmalen, P., Sloep, P., Brouns, F., Kester, L., Kone, M. and Koper, R. (2006) Knowledge matchmaking in courses: Alleviating the tutor load by mutually connecting course users. British Journal of Educational Technology, Vol. 37 (6), 881-895.

[2] Giles, J. T., Wo, L., & Berry, M. W. (2001). *GTP (General Text Parser) Software for Text Mining*. Retrieved online January, 2005 at: http://www.cs.utk.edu/~berry/papers02/GTPchap.pdf.

[3] De Jong, A., Brouns, F., Van Rosmalen, P., Sloep, P., Kester, L., & Koper, R. (2006). GUP GTP Usability Prototype. URL: http://sourceforge.net/projects/gup.

[4] Brouwers, M., Brouns, F., Van Rosmalen, P., Sloep, P. B., Kester, L. & Koper, R. (2006). ASA Tutor Locator. URL: http://sourceforge.net/projects/asa-atl.

[5] Janssen, J., Tattersall, C., Waterink, W.,Van den Berg, B., Van Es, R., Bolman, C., & Koper, R. (in press). Self-organising navigational support in lifelong learning: how predecessors can lead the way. *Computers & Education*.

[6] Haley, D. T., Thomas, P., De Roeck, A. & Petre, M. (2005). A research taxonomy for latent semantic analysis-based educational applications. Technical report no. 2005/09. Retrieved April 10, 2006, from http://computing-reports.open.ac.uk/index.php/content/download/ 187/1136/file/TR2005_09.pdf.

[7] Wild, F., Stahl, C., Stermsek, G. & Neumann, G. (2005) Parameters driving effectiveness of automated essay scoring with LSA. In *Proceedings of the 9th International Computer Assisted Assessment (CAA) Conference*, Loughborough, UK, July 5, 2005 (pp. 485–494). Retrieved March 6, 2006, from http://nm.wu-wien.ac.at/research/pu blications/b497.pdf.

[8] Van Bruggen, J. M., Rusman, E., Giesbers, B. & Koper, R. (submitted). Latent semantic analysis of small-scale corpora for positioning in learning networks. Retrieved February 24, 2006, from http://hdl.handle.net/1820/561.