Kalz, M., Van Bruggen, J., Giesbers, B., Waterink, W., Eshuis, J., & Koper, R. (2014). A Study about Placement Support Using Semantic Similarity. *Educational Technology & Society*, *17* (3), 54–64.

A Study about Placement Support Using Semantic Similarity

Marco Kalz[1*], Jan van Bruggen[1], Bas Giesbers[2], Wim Waterink[3], Jannes Eshuis[3] and Rob Koper[1]

[1]Welten Institute – Research Centre for Learning, Teaching and Technology, Open Universiteit Nederland, 6401 DL Heerlen, The Netherlands // [2]Department of Technology and Operations Management, Rotterdam School of Management, Rotterdam, The Netherlands // [3]School of Psychology, Open Universiteit Nederland, 6401 DL Heerlen, The Netherlands // marco.kalz@ou.nl // jan.vanbruggen@ou.nl // bgiesbers@rsm.nl // wim.waterink@ou.nl // jannes.eshuis@ou.nl // rob.koper@ou.nl

[*]Corresponding author

(Submitted May 5, 2013; Revised September 5, 2013; Accepted November 23, 2013)

ABSTRACT

This paper discusses Latent Semantic Analysis (LSA) as a method for the assessment of prior learning. The Accreditation of Prior Learning (APL) is a procedure to offer learners an individualized curriculum based on their prior experiences and knowledge. The placement decisions in this process are based on the analysis of student material by domain experts, making it a time-consuming and expensive process. In order to reduce the workload of these domain experts we are seeking ways in which the preprocessing and selection of student submitted material can be achieved with technological support. This approach can at the same time stimulate research about assessment in open and networked learning environments. The study was conducted in the context of a Psychology Course of the Open University of the Netherlands. The results of the study confirm our earlier findings regarding the identification of the ideal number of dimensions and the use of stopwords for small-scale corpora. Furthermore the study indicates that the application of the vector space model and dimensionality reduction produces a well performing classification model for deciding about relevant documents for APL procedures. Together we discuss methodological issues and limitations of our study whilst also providing an outlook on future research in this area.

Keywords

Methods, Assessment, Placement, Latent semantic analysis, Accreditation of prior learning, Networked learning

Introduction

In open and networked learning environments (Koper, Rusman, & Sloep, 2005) it is an important problem to support learners to find appropriate learning opportunities that fit to their competence level and prior learning. While traditional technology-enhanced learning environments are currently facing an innovation phase due to the widespread use of social and mobile media most of the time the assessment systems lag far behind the innovation processes. We see the development of assessment systems for open and networked learning environments as one of the grand challenges for technology-enhanced learning that makes it necessary to develop and evaluate alternative methods to approximate the prior knowledge of learners and to construct individual learning paths through their learning network (Janssen et al., 2010). Traditionally this problem has been addresses by intelligent tutoring systems with methods from adaptive hypermedia research like scalar models, overlay models, perturbation models or genetic models (Brusilovsky & Millan, 2007). However, it is rather a closed world in which these models operate; systems can only take account of what they represent/know about the learner in the current learning environment. All "system experiences" are lost after changing the learning environment and cannot be reused in another system. This problem has been recognized as the “open corpus” problem (Brusilovsky & Henze, 2007). While these models have their value in traditional e-learning processes in learning networks they are not applicable at all. Alternative bottom-up approaches are needed to offer personalized learning paths that do not need extensive sets of metadata to reason about prior knowledge of learners. Traditionally, higher education institutions in several European countries support the accreditation of prior learning (APL) (Merrifield et al., 2000). A typical APL procedure consists of four main phases (Van Bruggen et al., 2004):

- In a *profiling* phase the institution collects information about the learner’s needs and personal background.
- In the second phase learners collect and *present evidence* about their qualifications and experience. This evidence should support a claim for credit for the new qualification they are seeking.

- In the *assessment phase* the evidence submitted by the learner is analysed and reviewed conforming to the local assessment standards. The result of this phase of the procedure is an answer to the question of whether the student should be granted recognition of their prior learning.
- In the *accreditation phase* the results are verified by the department responsible for awarding the credit or recognizing the outcome of the assessment.

The procedures of APL are costly and time-consuming because they involve domain experts to assess the contents of the portfolios submitted by the students. There are two different approaches to accreditation in higher education institutions. On the one hand there is a generalized accreditation procedure based for example on certificates from vocational education which are expected to be equivalent to local certificates. On the other hand there is an individual accreditation procedure that also takes into account prior learning from non-formal and informal contexts. This second type of accreditation is seeking technological support models (Joosten-ten Brinke et al., 2008). These support models can range from a form of pre-advising the experts about which documents are relevant for the target course or study programme or it could help students to only fill their portfolios with material that is relevant to possible exemption decisions. At the same time these models have the potential to contribute to a future research agenda for technology-enhanced assessment in open and networked learning environments.

The basic assumption of our research is that prior knowledge of learners can be approximated by the content of the learner portfolio and therefore overlap between the documents in the portfolio and the courses of the plan/curriculum can act as a proxy to give exemptions and provide a personalized curriculum. The learner portfolio can consist of a variety of documents for other educational or work contexts like study assignments, thesis work or a project report. In an earlier publication (Kalz et al., 2007) we have analyzed existing technological solutions to assess prior knowledge with technological support. We have focused there on three categories of technologies, namely content similarity, metadata and ontologies. In this study we have focused on the content similarity aspect. The similarity calculation is executed in our study with the help of a dimensionality reduction technique similar to Latent Semantic Analysis (LSA).

LSA for prior knowledge approximation

Latent Semantic Analysis (LSA) is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations (Landauer, 2007). The latent semantic analysis process consists of several steps: In the *indexing phase* all words and documents construct a so called Term-Document-Matrix (TDM), in which all terms are listed in the columns and all documents in the rows. After the counting of the occurrence of each word in all documents several weighting and normalization options are possible. In the *dimensionality reduction* phase a mathematical function called singular-value decomposition (SVD) is applied which is similar to factor-analysis. The end result of this process is a latent semantic space, in which the input documents are represented as vectors. Documents in this space are similar if they contain words which appear in the same context and so their vectors are close together in the space providing a measurement for the similarity of text. In the *retrieval or query phase* the query text is projected into the space and the distance to the document vectors is calculated via the cosine or the Euclidian mean. Since we cannot elaborate on LSA in more detail in this paper we recommend the papers mentioned above as an introduction to LSA.

Latent Semantic Analysis has been applied to several problems in the domain of Technology-Enhanced Learning such as peer tutoring (Van Rosmalen, 2008), provision of feedback (Graesser et al., 2005), automated essay scoring (Foltz, Laham & Landauer, 1999) and the selection of educational material (Graesser et al., 2000). Zampa and Lemaire (2002) applied LSA to the user modeling problem. In their model learners learn a domain by acquiring the most important “lexemes” of a domain. The most important crucial concepts are identified beforehand and the end result is a recommendation to look at the next best item based on the theory of proximal development (Vygotsky, 1978). Wolfe et al., (1998) focused on the use of LSA for the selection of educational material.

Our application of LSA is similar to those presented here but differs in a number of important aspects. In contrast to the approach of Zampa and Lemaire (2002) we do not aim to model the student background and the learning resources beforehand. Since we used text students have written in their prior education as a proxy to their prior knowledge we have a more dynamical model which could build on learning progress documented e.g., in an electronic portfolio. Our application of LSA also differs from a “simple” essay scoring scenario where a student text is compared to one ore more clearly defined gold standard texts which provide the basis for judging about the quality of an essay. In our case the topical range of the student documents can be potentially very broad while the target documents or course units can be very similar to each other. This makes it very important to find an ideal dimensionality that results in the best discrimination between target documents.

Study

In this paper we present a study that has been conducted at a Psychology Course of the Open University of the Netherlands. The foundations for this study are described in (Van Bruggen et al., 2004), and research agenda for it has been presented in (Kalz et al., 2008). The study employed a vector space model and dimensionality reduction technique inspired by Latent Semantic Analysis (LSA) to calculate a similarity between documents in the learner portfolios and the content of the course units. To test the validity of the results an expert validation with domain experts is performed and the performance of using LSA as a classifier for relevant or irrelevant documents was evaluated with a Receiver Operating Characteristic (ROC curve). ROC curves are a method from signal detection theory and they are often applied to binary classification experiments and model performance evaluation.

The hypothesis tested with this study is as follows:
H1: LSA can classify documents as relevant/irrelevant in a manner comparable to human experts.

Our target for this study is to minimize the false-negative and the false-positive cases under a threshold level of 10%. Too many false-negative cases would hinder students from exemptions while too many false-positive cases would result in unnecessary work for the domain experts. In addition we checked the results we reported in a previous study about the identification of an ideal number of dimension for small corpora (Van Bruggen et al, 2006). However it is important to note that this study was not intended to evaluate the use of automated exemption rules for study programs but is intended to evaluate the applicability of the vector space model and dimensionality reduction techniques for learner placement in general. The exemption rules and problems associated with them such as trust issues, validity of thresholds etc. are not a specific problem of the method we are evaluating but of the general exemption procedure. This means that exemptions standards can differ between institutions and that they have to set their own thresholds for exemptions.

Method

Study context and participants

The study was carried out in the context of an introductory psychology course of the Open University of the Netherlands. Study participants were drawn from the 244 students that enrolled in the course around 7% submitted material to apply for the accreditation of prior learning. Overall we had a total number of 18 participants providing 28 documents to be compared to the units of the target course. Thus the total number of similarity ratings was 504 (28 documents x 18 course units).

Study procedures

The introductory psychology course consisted of 18 units each of them dedicated to a subtopic of psychology. The course was offered in an online environment. Before students could enter the course they had to read an introduction about the content of the units of the course. Thereafter, the students filled out a questionnaire on any prior knowledge for the course or parts of the course. Students were invited to submit materials they had produced in their prior education or working environments. Documents submitted were work reports, (bachelor / master) theses, technical reports, essays, reference lists and presentations.

We have chosen the ideal dimensionality for the study based on two performance criteria. On the one hand we employed the connection between singular values and the variance they account for in the Term-Document matrices analyzed by LSA. Our target was to reduce the variance accounted for under a threshold of 70 – 80 % of variance represented in the data. This approach is in line with other studies that have a comparable purpose. Then we used the corpus documents as queries to control the discrimination between the chapters. These queries were tested for selfcorrelations and discrimination to the other chapters under different conditions. We varied the dimensions used between 5 and 1000 and we used different stopword settings (no stopword list, 30% stopword, 50% stopword list). Stopwords were removed because they either appeared very infrequently or very often in the corpus. In addition we also evaluated the use of local and global weighting options. In this process we followed a method by (Rosmalen et al., 2006) to calibrate and test several LSA parameters.

In order to evaluate the results of the study two domain experts independently rated the similarity between the student documents and the domain documents on a 5-point Likert-scale. For each student document the experts evaluated the semantic similarity and marked if they would give an exemption. In addition they wrote down how much time it took to review the material. We also interviewed one of the domain experts. We calculated a raw overall percentage of agreement between the two raters. We calculated the interrater agreement according to the consensus and the constistency of the ratings by the two judges (Stemler, 2004). The consensus of the ratings by the two domain experts was calculated using Cohen's Kappa. The Spearman rank coefficient was used to calculate the consistency among the ratings. For the performance assessment we recoded the Likert scale into a binary scale. Here we used the most optimistic rule that a document is seen as relevant when at least one of the raters has rated the document with a value higher than 1.

The model performance assessment of our method as a classifier for relevant and irrelevant documents for APL procedures was analyzed via a confusion matrix and ROC-curve (Fawcett, 2003). ROC-Curves (Receiver Operating Characteristics) are a method from signal detection theory that has been applied to evaluate model performance assessment of classification models. In ROC curves the true positive rate (tpr) of a classifier is plotted against the false positive rate (fpr) while varying the thresholds used for the classification. One of the main advantages of the application of ROC-curves is that they are not sensitive to class skew (Hamel, 2009). With this approach we have also compared the effect of applying different weighting functions reported in the literature to contribute positively to performance in small-scale corpora (Nakov et al., 2001; Wild et al., 2005). We have compared the use of a logarithmic weighting function for a local weighting and the use of entropy and inverse document frequency for the global weighting and the combination of logarithmic and entropy weighting.

In addition we have calculated raw success scores on the document level (How many documents would have been recommended right?) and the person level (How many learners would have been exempted right if LSA would have decided about exemptions?). The corpus for the experiment consisted of the content of the 18 psychology learning units. This corpus had 28165 terms with 490431 occurrences. The corpus size was 3,1 MB. 13283 terms only appeared once in the corpus. After keeping only terms that appeared more than once the corpus was reduced to 14882 terms with 477147 occurrences.

All corpus and learner documents have been manually cut into paragraphs. The paragraph length was between 250 and 500 words. The corpus consisted in the end of 2246 paragraphs. In this study all analysis was done using the Text to Matrix Generator (TMG) - a Matlab implementation of Latent Semantic Analysis and other vector space techniques (Zeimpekis & Gallopoulos, 2006). For the experiment a script was written that calculates the mean of cosines for all paragraph to paragraph comparisons and writes down the mean correlation to all 18 chapters in a spreadsheet file.

Results

Dimensionality reduction and sensitivity

For the estimation of the ideal dimensionality we were able to reproduce results obtained in a prior explorative study (Van Bruggen et al., 2006). Figure 1 shows our research corpus with different stopping strategies applied and with different numbers of singular values.

In this figure we can see that the variance zone of 80% variance accounted for or higher starts with different numbers of singular values depending on the stopping strategy used. For no stopping this zone starts at a reduction to 194 singular values. For the 30% stopping strategy this zone starts at 530 singular values while it starts for the 50% stopping strategy at 713 singular values. The second performance criterion was tested with the discrimination between the target learning units. Figures 2 – 4 illustrate the self-correlation and the correlation to the other learning units for the learning unit 1. These figures show the relation between singular values extracted for the analysis in relation to cosine similarity between the units. The ultimate goal for the ideal number of singular values and stopwords would be that there is a clear distinction between the units to be able to discriminate between the relevant and obsolete target learning activities. Note that we did not analyze the full range of singular values but only a selection since the figures are only used for illustrative purposes.

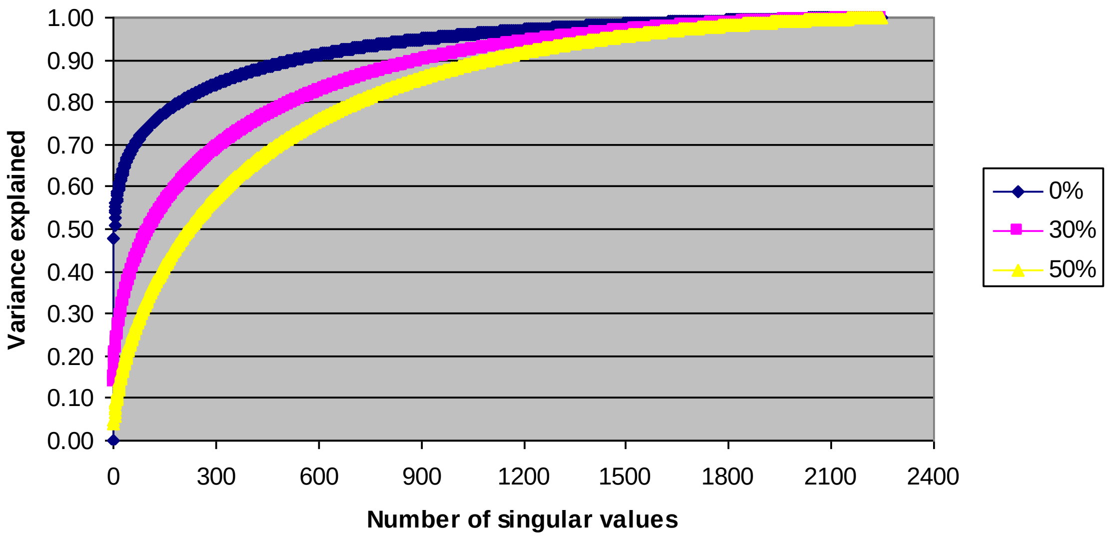

Figure 1. Correlations and variance explained for different numbers of singular values and 3 different stopping strategies

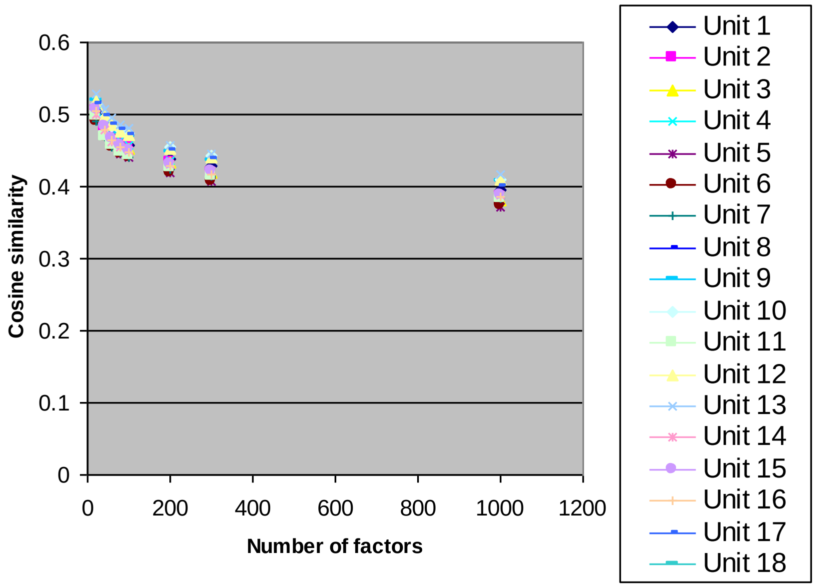

Figure 2. Correlations and self-correlation for learning unit 1 with no stopping

As we can see in figure 2 it is not possible to discriminate between unit 1 and the other learning units of the course. All units are too similar to each other to be able to discriminate between them. In figure 3 we have increased our stopword strategy to 30 %.

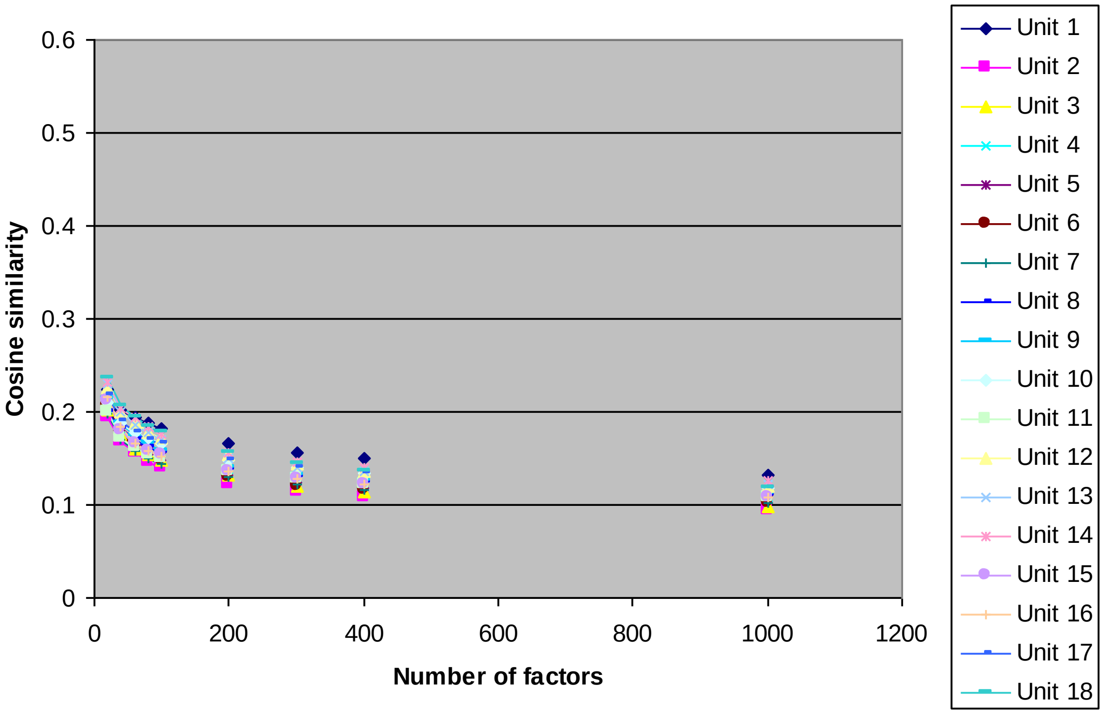

Figure 3. Correlations and self-correlation for learning unit 1 with 30% stopping

As we can see in figure 3 the cosine values drop and discrimination between the chapters improves with a 30% stopping strategy. But still it is not easy to clearly discriminate between the chapters. In figure 4 we increased our stopping strategy to 50% stopwords.

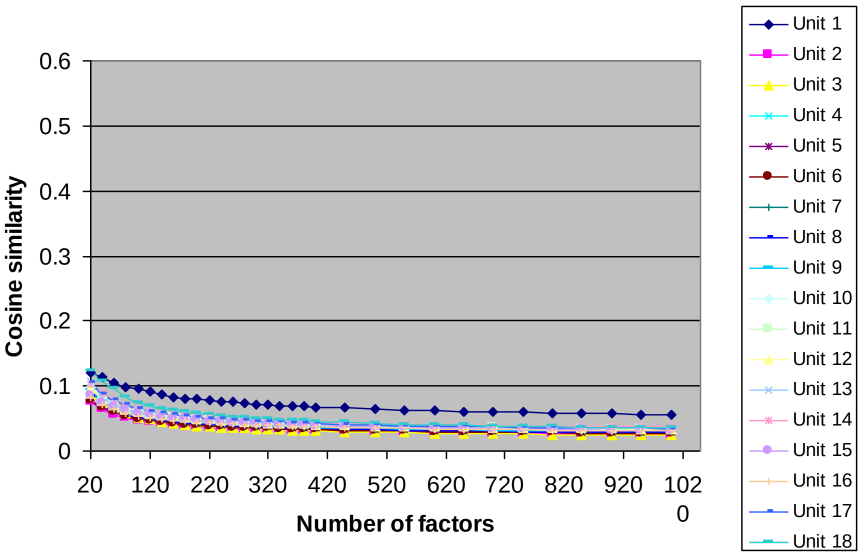

Figure 4. Correlations and self-correlation for learning unit 1 with 50% stopping

Overall we can see that the cosine values drop the more stopwords we apply, but at the same time the discrimination between the target documents in the corpus improves sufficiently to discriminate between the chapters. This effect could be replicated for all 18 units. After having defined the ideal number of factors and stopwords for the analysis we have used these parameters (800 factors with a 50% stopword list) for querying the student documents and comparing the cosine values to the 18 learning units.

Model performance assessment

The raw agreement between the two domain experts was 95%. This high agreement between raters was mainly based on the high number of submissions not rated as relevant for the APL procedure. The interrater reliability for the raters was found to be Kappa = 0.77 ($p < 0.001$). According to Spearmans rho (1904) there was a high consistency between the ratings (rho = 0.75, $p < 0.01$). After the recoding into a binary classification the interrater agreement was Kappa = 0.74 ($p < 0.001$).

Only 8 % of all cases (40 cases) were evaluated as relevant for APL. Because of this the data were negatively skewed. The expert data confirmed our basic assumption that content similarity is related to exemptions in APL procedures. If we take into account only cases with content similarity higher 2 on the Likert scale then 83 % of the cases have been proposed for an exemption in the mean of both raters. Seven learners would have been given exemptions based on the decisions of experts which equals 38 % of all participants who submitted material for the study.

In our evaluation of the impact of different weighting functions we have compared five parameters as shown in Figure 5. This figure is a plot of the true positive rate on the y-axis and the false positive rate on the x-axis.

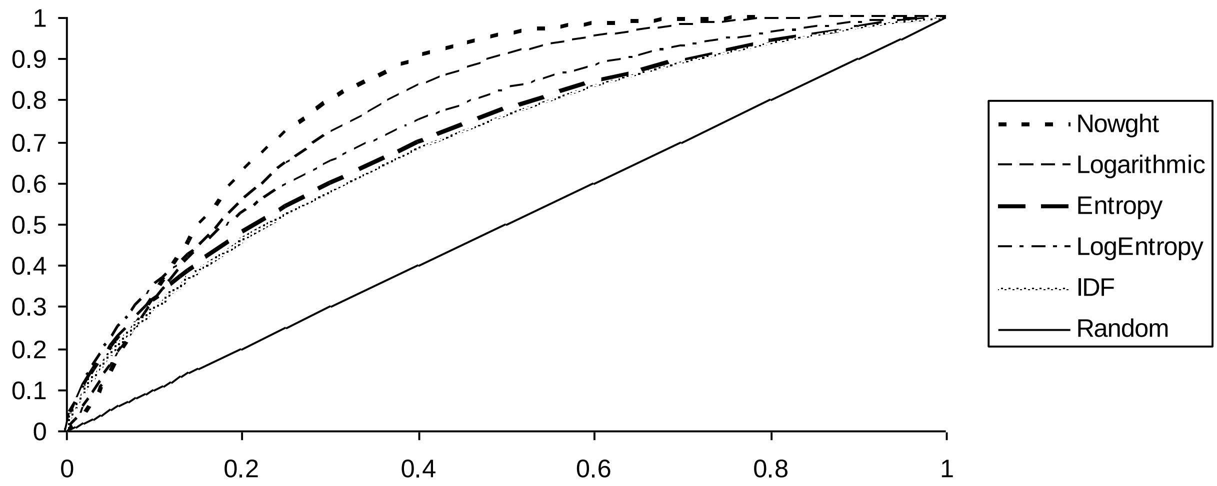

Figure 5. Receiver operating characteristics curve (ROC curve) for LSA with different weighting parameters

There are several important aspects to be reported. In general, we can report that without taking into account the different weighting functions our method can be clearly distinguished from a random classifier that would rate documents by chance into each of the categories. Overall the use of the weighting functions reported in the literature has slightly decreased the performance of our model and the best performance could be reached without any local and global weighting. To compare this performance in more detail we provide in table 1 an overview of the size of the Area und the ROC curve (AUC) which can be used to evaluate the performance of a classification model.

*Table 1***.** Influence of weighting to performance measured with the size of the Area under the ROC-Curve and the standard error

	AUC	Std. Error.
No wght.	0.811	0.0262
Logarithmic	0.777	0.0313
Entropy	0.708	0.0413
Log./Entropy	0.741	0.0386
IDF	0.697	0.0417

We can see that the inverse document frequency function (idf) performs worst for our model and that the logarithmic function was the second best option. If we also take into account the confidence intervals of the ROC curves we have to summarize that the weighting functions decrease the performance but not significantly because the confidence intervals of all settings overlap as we can see in figure 6.

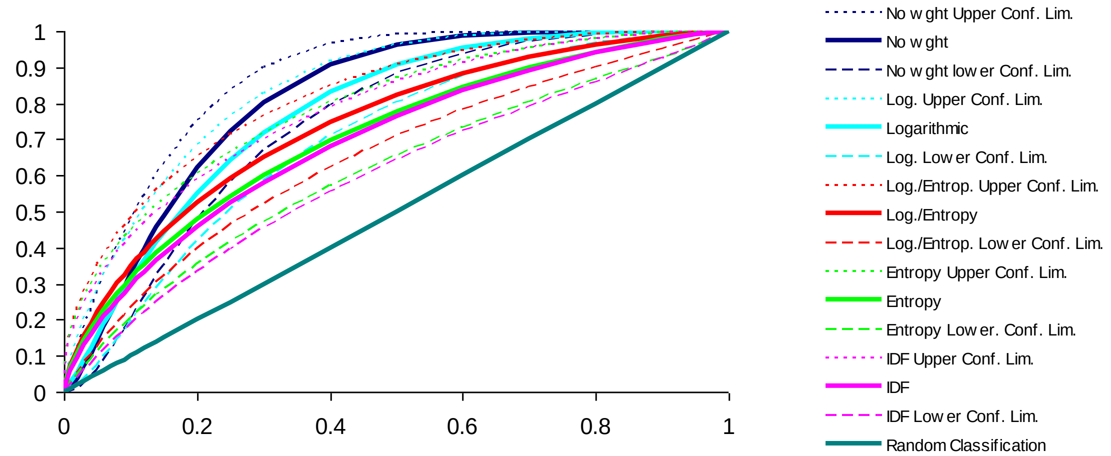

Figure 6. Impact of local and gobal weighting to the model performance assessment

Thus we have decided to use the 800 dimensions, a 50% stopword list and no weighting for our model. For this settings human raters and LSA results show only a moderate agreement of Kappa = 0.33 ($p < .0.001$). The mean of the ratings was 0.11 (Std. error = 0.21) while the mean variance was 0.23 and the mean standard deviation 0.47. To discuss the performance of our model more into detail we provide an overview of the results with a confusion matrix (table 1). We can see that LSA could successfully classify 88 % of all cases right. The false positive cases are 8 % of all cases while 2 % of the cases fall into the false negative category.

Table 2. Classification results Human ratings vs. LSA ratings (n = 504)

		Human rating	
		Relevant	Irrelevant
LSA rating	Relevant	28	40
	Irrelevant	12	417

The *sensitivity/true positive rate* of our model is 0.7 while the *specificity* is 0.91. The *false positive rate* is 0.08. Overall the *negative predictive value* of our model is 0.97 while the *false discovery rate* is 0.58. The Area Under the ROC-Curve (AUC) for our model was 0.81 (95% CI, Std Err. = 0.0262). Overall this is a sign of a good predictive model.

On the document level we have analyzed how many documents of the 36 would have been passed to the human experts if our LSA model would have been implemented to evaluate relevant documents. Here we have compared two different methods: The mean of all cosine values to each chapter for every learner document and the maximum cosine value of the comparison between learner document and all 18 chapters. The raw percentage of right classified documents of the mean method was 85 % while it was only 43 % with the maximum method. On the level of the learner our model would have recognized 7 of 12 given course unit exemptions of the human raters but it would have added 40 false positive exemptions on top for these 7 learners. Overall human experts needed in the mean 255 minutes to analyze the material provided by the students.

Discussion

Based on a method to estimate the ideal number of dimensions retained in an LSA space we could reach sufficient discrimination between the target documents and identify the ideal number of dimensions for our study. With these results we confirm our findings from an earlier study on LSA and small-scale corpora. This lets us conclude, that the application of a stopword strategy between 30% and 50% is needed to reach sufficient discrimination between documents for placement support on small-scale corpora. After testing different weighting options we could show that weighting does not improve the performance of our model. Overall we could reach a satisfactory classification model because of two reasons. First we could reach our self-set target of less than 10% false positive and false negative cases. Second, an AUC value of 0.81 is seen as a good performance indicator for a classification model.

In summary we have to reject the null-hypotheses showing that LSA can successfully discriminate sufficiently between relevant and irrelevant documents according to the targets we have set ourselves.

But there are several limitations to the here presented study. First of all the low number of participants and especially relevant documents is problematic in case of generalizability of the results. The data we collected for an APL procedure resulted in negatively skewed data. After discussing these findings with an APL expert we discovered that students are often not aware of what to submit for exemptions. This means that collecting more student data would likely lead to a similar skewed dataset. In fact, it is a part of the problem we are trying to solve with this technology.

Some of the false negative cases reveal that human experts decide about exemptions on more factors then just the semantic similarity between learner documents and target documents. One case was especially interesting in this regard: One of the learners submitted a very detailed description about an experiment they had conducted. This description did not contain sufficient semantic concepts which illustrated a relationship to the target documents. But human raters deducted from the document that this learner must have specific prior knowledge in psychology to be able to write such a document. For this purpose other techniques and approaches to approximate prior knowledge are needed which go beyond semantic similarity. In related experiments we have focused on semantic networks and other ways to visualize content of sets of documents for visual inspection (Kalz et al., 2009; Berlanga et al., 2011). This could address the problem of the false negative since there would be a combination of methods which would address most false negative cases from the study.

The qualitative interview showed us that the analysis model of the human experts is based on semantic similarity of documents but the cognitive process is more complicated. One domain expert described the analysis process with different steps involving keyword analysis, semantic analysis and quality ranking of the student documents. While our study confirms that the application of dimensionality reduction techniques like LSA for the support of APL procedures and the approximation of prior knowledge is a promising research and development direction the approach needs to be validated in several different contexts. In this regard we expect that a training phase is needed in all implementations to align the approach to local thresholds and local decision boundaries.

We believe that the work presented here can stimulate research about alternative bottom-up assessment methods that will play an important role in future open and networked learning environments. From a European perspective especially the development and evaluation of similar approaches in a multilingual context is a challenging research direction.

References

Berlanga, A., Kalz, M., Stoyanov, S., van Rosmalen, P., Smithies, A., & Braidman, I. (2011). Language technologies to support formative feedback. *Educational Technology & Society, 14*(4), 11–20.

Joosten-ten Brinke, D. J., Brand-Gruwel, S., Sluijsmans, D., & Jochems, W. (2008). The quality of procedures to assess and credit prior learning: Implications for design. *Educational Research Review, 3*(1), 51-65.

Brusilovsky, P., & Henze, N. (2007). Open corpus adaptive educational hypermedia. In P. Brusilovsky, A. Kobsa, & W. Nejdl, *The Adaptive Web* (pp. 671-696). Berlin, Germany: Springer.

Brusilovsky, P., & Millan, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web* (pp. 3 - 53). Berlin, Germany: Springer.

Fawcett, T. (2003). *ROC graphs: Notes and practical considerations for researchers.* (HP Labs Tech Report No. HPL-2003-4). Paolo Alto, CA: HP Laboratories.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, *1*(2). Retrieved May 15, 2013, from http://imej.wfu.edu/articles/1999/2/04/index.asp

Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4), 612-618.

Graesser, A., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., & Person, N. (2000). Using latent semantic analysis to evaluate the contributions of students in autotutor. *Interactive Learning Environments, 8*(2), 129-147.

Hamel, L. (2009). *Model Assessment with ROC Curves.* In J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining* (2nd ed.) (pp. 1316-1323). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-010-3.ch204

Janssen, J., Berlanga, A. J., Heyenrath, S., Martens, H., Vogten, H., Finders, A., … Koper, R. (2010). Assessing the learning path specification: A pragmatic quality approach. *Journal of Universal Computer Science, 16*(21), 3191-3209.

Kalz, M., Van Bruggen, J., Rusman, E., Giesbers, B., & Koper, R. (2007). Positioning of learners in learning networks with content, metadata and ontologies. *Interactive Learning Environments, 15*(2), 191-200.

Kalz, M., Van Bruggen, J., Giesbers, B., Waterink, W., Eshuis, J., & Koper, R. (2008). A model for new linkages for prior learning assessment. *Campus-Wide Information Systems, 25*(4), 233–243.

Kalz, M., Berlanga, A., Van Rosmalen, P., Stoyanov, S., Van Bruggen, J., & Koper, R. (2009). Semantic networks as means for goal-directed formative feedback. In V. Hornung-Prähauser (Ed.), *Kreativität und Innovationskompetenz im digitalen Netz - Creativity and Innovation Competencies in the Web, Sammlung von ausgewählten Fach- und Praxisbeiträgen der 5. EduMedia Fachtagung 2009* (pp. 88–95). Salzburg, Austria: Salzburg Research.

Koper, R., Rusman, E., & Sloep, P. (2005). Effective learning networks. *Lifelong Learning in Europe, 1*(1), 18-27.

Landauer, T. (2007). *LSA as a theory of meaning.* In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 3-34). Mahwah, NJ: Lawrence Erlbaum Associates.

Merrifield, J., McIntyre, D., & Osaigbov, R. (2000). Mapping APEL final report: Accreditation of prior experiential learning in English higher education. London, UK: Learning from Experience Trust.

Nakov, P., Popova, A., & Mateev, P. (2001). Weight functions impact on LSA performance. In R. Mitkov (Ed.), *Proceedings of the Recent Advances in Natural language* (pp. 187-193). Tzigov Chark, Bulgaria: Bulgarian Academy of Science.

Rosmalen, P., Sloep, P., Brouns, F., Kester, L., Kone, M., & Koper, R. (2006). Knowledge matchmaking in Learning Networks: Alleviating the tutor load by mutually connecting learning network users. *British Journal of Educational Technology, 37*(6), 881-895.

Stemler, S. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4).

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*,72–101.

Van Bruggen, J., Sloep, P., Rosmalen, P. V., Brouns, F., Vogten, H., & Koper, R., & Tattersall, C. (2004). Latent semantic analysis as a tool for learner positioning in learning networks for lifelong learning. *British Journal of Educational Technology, 35* (6), 729-738.

Van Bruggen, J., Rusman, E., Giesbers, B., & Koper, R. (2006). Content-based positioning in learning networks. In Kinshuk, R. Koper, P. Kommers, P. Kirschner, D. Sampson, & W. Didderen (Eds.), *Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies* (pp. 366-368). Kerkrade, The Netherlands: IEEE.

Van Rosmalen, P. (2008). Supporting the tutor in the design and support of adaptive e-learning. Heerlen, The Netherlands: Open University of the Netherlands.

Vygotsky, L. S. (1978). Mind in society: The development of higher mental process.

Wild, F., Stahl, C., Stermsek, G., & Neumann, G. (2005). Parameters driving effectiveness of automated essay scoring with LSA. *Proceedings of the 9th Computer Assisted Assessment Conference*. Retrieved from the CAA website: http://caaconference.co.uk/pastConferences/2005/proceedings/WildF_StahlC_StermsekG_NeumannG.pdf

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by Latent Semantic Analysis. *Discourse Processes, 25*, 309-336.

Zampa, V., & Lemaire, B. (2002). Latent Semantic Analysis for user modeling. *Journal of Intelligent Information Systems, 18*(1), 15-30.

Zeimpekis, D., & Gallopoulos, E. (2006). TMG: A MATLAB toolbox for generating term-document matrices from text collections. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping multidimensional data: Recent advances in Clustering* (pp. 187-210). Berlin, Germany: Springer.