A Placement Web-Service for Lifelong Learners

PREPRINT – please cite as: Kalz, M., Drachsler, H., Van der Vegt, W., Van Bruggen, J., Glahn, C., & Koper, R. (2009). A Placement Web-Service for Lifelong Learners. In Tochtermann, K. & Maurer, H. (Eds.): Proceedings of the 9th International Conference on Knowledge Management and Knowledge Technologies (pp. 289-298), 2-4 September 2009, Verlag der Technischen Universität Graz:Graz, Austria.

A Placement Web-Service for Lifelong Learners

Marco Kalz
(Open University of the Netherlands, Heerlen, The Netherlands
marco.kalz@ou.nl)

Hendrik Drachsler
(Open University of the Netherlands, Heerlen, The Netherlands
hendrik.drachsler@ou.nl)

Wim van der Vegt
(Open University of the Netherlands, Heerlen, The Netherlands
wim.vandervegt@ou.nl)

Jan van Bruggen
(Open University of the Netherlands, Heerlen, The Netherlands
jan.vanbruggen@ou.nl)

Christian Glahn
(Open University of the Netherlands, Heerlen, The Netherlands
christian.glahn@ou.nl)

Rob Koper
(Open University of the Netherlands, Heerlen, The Netherlands
rob.koper@ou.nl)

Abstract: This contribution introduces a placement web-service which has been developed in the context of the TENCompetence Integrated Project. The web-service uses Latent Semantic Analysis (LSA) to calculate similarity between documents in learner portfolios and content in the current learning network of the learner. In the first part of the paper the placement problem in learning networks is introduced. Next we discuss the architecture of the current web-service prototype and describe our evaluation approach. Several limitations and future developments are discussed at the end of the paper.

Keywords: Latent Semantic Analysis, Placement, Web-Service, TENCompetence
Categories: L.1.0, L.2.2, L.3.0, L.6.0, L.3.6

1 Introduction

Aside from traditional educational offerings by institutional providers of technology-enhanced learning more and more focus is given to learning scenarios outside these institutions. Especially in professional education the importance of learning networks [Koper, Sloep and Rusman 05] for non-formal education is rising [Kamtsiou et al. 08]. This is, besides other factors, also provoked through the use of social software for self-directed learning and competence development [Klamma et al.

07]. In these learning networks people get together who share one or more competence development goals. Several providers can contribute learning activities to this network. Based on these contributed resources, learning activities and the behaviour of the participants emerging effects can be used as the basis for several service offers.

In the TENCompetence project one focus is the development of models and tools which help to bridge the worlds of formal education and non-formal approaches all subsumed under the concept of competence development programmes [Koper & Specht 08]. In the TENCompetence Integrated Project two independent but intertwined Web services have been developed to provide learners with orientation and feedback in learning networks. The placement Web service provides learners with a starting position in learning networks while the navigation web-service [Kalz et al. 08a] leads the learners through the learning network based on the position provided by the placement service. These services have been integrated in one recommender system for learning resources which can use and reason with information from several sources and backgrounds. While data from formal providers of competence development programs can most of the time offer well structured information about the programmes and e.g. dependencies between learning activities in non-formal learning scenarios this can not be assumed.

For this purpose the hybrid personalizer has been developed [Herder & Kärger 08]. The hybrid personalizer tries to overcome the problems of both learning contexts (formal & non-formal) through a collection of “top-down” and “bottom-up’ Web Services which contribute to a recommendation of learning activities. The recommendation is computed based on information available about the learner, the learning activities, and the behaviour of other (successful) learners. The recommendation provided by the system is two-fold: on the one hand, it ranks learning activities based on how close they are to the learner’s current knowledge level; that is, learning activities that are still too advanced are scheduled at a later point in the initial recommended visualization of the curriculum (or rather, competence development program).

On the other hand, as a second dimension, the system ranks learning activities based on to what extent they match the learners’ preferences—as explicitly indicated in their profiles and as estimated from the behaviour of similar users. The placement service was used here mainly to compute the similarity between the content of the learner’s portfolio and the content attached to the learning activities from non-formal sources without an extensive set of metadata. To test the reliability and performance of the placement web-service we have conducted a technical evaluation for further development.

In this paper we focus on the placement web-service for lifelong learning. Based on the discussion of learners needs in non-formal professional learning we introduce the problem of placement in learning networks. The architecture of the placement web-service is discussed and a first technical evaluation and empirical results are presented. Several limitations are discussed and a research and development outlook is provided.

2 The Placement Problem in Learning Networks

2.1 Learner needs in non-formal professional technology enhanced learning

Lifelong learners which make use of Technology-Enhanced-Learning have different needs and different expectations of technology-support for learning then learners in formal educational settings. Both learners share the wish to get personalized educational offers which fit to their prior knowledge and their learning goal. Most approaches from the traditional Adaptive Hypermedia (AH) literature assume for this problem the existence of a top-down-design for learning and a clear structure with pre-requisites defined and adaptation based on these predefined structures. In research on adaptive sequencing of learning resources for example a domain expert is assumed that models the learning goals and the domain concept ontology [Karampiperis & Sampson 05].

In non-formal education these assumptions often do not hold, because the sources and routes through the learning networks are not clearly defined and structured. This is also recognized in recent literature from adaptive hypermedia and referred to as the “open corpus problem” [Brusilovsky and Henze 07].

While these traditional models calculate the prior knowledge of learners only internal in their systems through a logic which is based on which pages a learner has visited inside the system and which pages contain similar content, for learning networks for non-formal professional development other techniques to take into account prior knowledge of learners are needed due to the open and dynamic structure of these networks. One approach to solve this problem is to ask contributors of learning networks to define the related learning goals to every learning activity added to the network and retrieve goal-related learning activities [Lindstaedt et al. 2008]. While this approach may work well in more hierarchically organized learning and working environments its success depends to a very large extend on the willingness and ability of contributors to add these information.

In learning networks other techniques have to be applied which can handle the dynamic nature of the learner profiles and the learning content in the network as well. Because of this dynamic nature we have chose an approach which takes the similarity between the content in the learning network and the content in the learner portfolio as a proxy for prior knowledge analysis [Kalz et al. 08b]. This is what we refer to as the “placement problem” in learning networks which we discuss in the next section in detail.

2.2 The placement problem in learning networks

In a nutshell the placement problems boils down to the question which learning activities a learner should take in the learning network and which can be omitted taking into account his prior knowledge. In traditional formal education this problem is mirrored in a process called Accreditation of Prior Learning (APL). In this process learners apply with their portfolio for a study programme. Domain experts study these portfolios and decide based on the submitted material which parts of the study programme can be exempted and which personalized curriculum the learner gets. This process is time-consuming and expensive so a need is to support this process with technology.

In the UK this process has been recently covered in a JISC funded project which focused on the development of services for students and assessor in this procedure. An “estimator” service should serve in the pre-entry phase of the APL process to estimate the “importance” of a specific part of the learner portfolio for a potential claim [Haldane et al. 07]. In the context of the TENCompetence project we have evaluated and developed a solution for the same purpose which is based on the extensive use of Language Technology. In the next part we describe the architectural design of this application and present the context in which this service is embedded.

3 Placement Web-Service Prototype

3.1 Latent Semantic Analysis

The core technology of the Placement Web Service Prototype (PWSP) is Latent Semantic Analysis (LSA) [Landauer 2008]. LSA is a method for extracting and representing the contextual-usage meaning of words by statistical computations. The whole process of this analysis consists of several steps like the pre-processing of the text involving weighting and normalizing mechanisms, the construction of a term-document matrix, a mathematical function called singular-value decomposition (SVD) similar to factor-analysis, the rank reduction of the Term Document matrix and finally the projection of a query vector into the latent semantic space. In this latent semantic space the main concepts of the input are represented as vectors. Concepts or documents containing these concepts in this space are similar if their vector representations are close together in the space which provides a measurement for the similarity of text and concepts. The power of LSA in comparison to classical keyword techniques is the ability to match similarity even if documents do not contain any joint terms. LSA has been applied in several fields like medicine, psychology or computer science and it has shown good performance in essay scoring [Foltz, Laham, & Landauer 99] and tutoring systems [Graesser, Chipman, Haynes, & Olney 05].

Based on these experiences we have evaluated the use of LSA for approximating prior knowledge of learners based on their writings and results of this evaluation are promising. For the TENCompetence project a web-service prototype has been developed whose architecture is explained next.

3.2 Architecture of the PWSP

The Web-Service is built on top of an LSA implementation in PHP. The PHP implementation offers compiled methods for all basic LSA steps from importing documents, through cleaning up, building the Term-Document-Matrix (TDM) matrix, decomposing and reducing it and performing queries. This significantly reduces the code complexity without removing the option to implement custom scripted steps in between. It is possible to switch between a built-in and (two) external decomposition applications: As built in method we have used the ALGLIB library [Bochkanov and Bystritsky 08], while users can use General Text Parser [Giles, Wo and Berry 03] or

SvdLibC [Rohde 05] as well if they request them from the developers. No matter which engine is used, the output is always formatted in Harwell-Boeing format, making it easy to compare the various engines and to switch between them. The PHP implementation also allows various locations for input documents like ftp, (local) disk and nntp (usenet). Other document sources can be implemented easily. It also has a wide range of textual cleaning functions built-in. The implementation can work directly with documents in txt-format or Word, PowerPoint and PDF. It also allows these type of documents as query. Because it saves all intermediate output in an easy readable non compressed format if the matrices are small enough, the implementation is suitable for research purposes. Figurer 1 shows a high level overview about the web-service prototype.

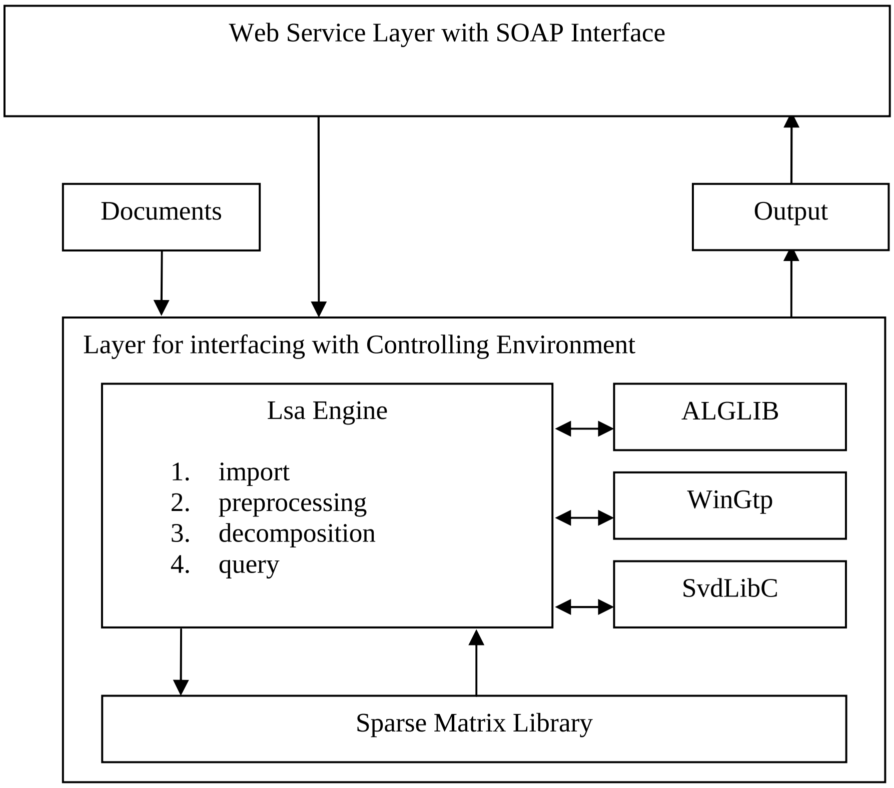

Figure 1: Architecture of the PWSP

On top of the Sparse Matrix Library the Lsa engine is built that performs the various actions necessary for Latent Semantic Analysis. These steps are:

1. importing documents
2. extraction of plain text from text files, Microsoft Word, Microsoft PowerPoint and Adobe Acrobat documents
3. pre-processing the imported text

4. constructing the Term x Document matrix (TDM)
5. performing the singular values decomposition and transforming the output of the decomposition
6. reducing the rank of the Term x Document matrix (TDM)
7. building a query vector
8. projecting a query vector onto the reduced Term x Document matrix
9. performing the query by calculating distances, dot products and cosines between the query vector and the reduced Term x Document space
10. allow retrieval of results

The Web-Service layer on top of the implementation offers a SOAP interface which can be called from external applications. As a result the web-service delivers a list of annotated documents with their similarity scores attached. Table 1 shows an overview of the API of the placement web-service prototype.

Placement Web-Service Interface (API)				
Name	**Method**	**Description**	**Input**	**Output**
getPositionValues	Get	Return a list of UoL annotated with cosine values	Iduser=xx	2-dimensional array of floats. Each UoL with its calculated cosine values.
Frequency		DATA Fields		Format
On request.	EventType	Iduser = Integer Learninggoal = Array (Strings)		

Table 1: Placement Web-Service Interface (API)

In the next part of the paper we discuss the evaluation of the web-service prototype.

4 Evaluation

4.1 Performance Evaluation

In two prior studies we have evaluated the performance of using LSA for prior knowledge approximation. Since LSA was used in the past on the basis of large corpora from information retrieval we first had to evaluate its applicability on much smaller corpora which we expect to be available in learning networks. In this first study we could show the importance of applying stopwords and we have proposed a method to estimate the number of factors to be retained in LSA for small corpora [Kalz et al. 09a].

In a second empirical study we have collected real data from students and compared LSA results with results from two domain experts who analyzed the documents according to semantic similarity and prior knowledge of students [Kalz et al 09b]. In this study we could show that we can optimize the LSA procedure by applying the stopword and dimensionality method from the former study and we could reach a false negative and false positive rate under 10%. In this regard the correctness of the web-service results was satisfactory.

4.2 Technical Evaluation

There are several approaches how to evaluate semantic web-services and frameworks from a technical perspective. While we cannot discuss here extensively the several approaches for evaluation of semantic web-services we would like to present only the categories in which these web-services and their frameworks are evaluated. [Küster et al. 08] present the following aspects for such an evaluation: Performance & Scalability, Usability & Effort, Correctness, Decoupling and Scope & Automation. In Table 2 we show the evaluation results from a first evaluation of the PSWP.

Category	Result
Performance & Scalability	30 seconds for a dataset of around 800 documents (total size 7 MB).
Usability & Effort	Usability is high. Currently low effort, but might increase in future versions.
Correctness	Has been evaluated in an empirical study, but we expect that a “training phase” is needed for every implementation.
Decoupling	Not tested/Does not apply
Scope & Automation	Not tested/Does not apply

Table 2: First Evaluation Results of the PWSP

We have tested the performance and stability of the web-service with the same dataset which we have used for the empirical evaluation of LSA for placement purposes. For the dataset which had a size of 7MB in total the service needed around 30 seconds to respond and deliver the annotated list of learning activities (Core 2 Duo T7200 (2 GHz) PC with 1 GB of RAM). All documents used in the evaluation were text files so that no conversion into text format was needed. Scalability has been tested with datasets of different size and the time needed to get the result list increased on the one hand with the number of documents that have to be converted into textformat and on the other hand with the number and size of documents.

The usability of the current prototype is high, since we have coupled it with a WAMP (Windows, Apache, MySQL, PhP) package and so the setup time for the web-service is low. We could not test the aspects of decoupling and scope/automation since these aspect fit better to large frameworks than to a single web-service. In the next part of the paper we discuss the results of the evaluation and provide a perspective for further development of the web-service.

5 Conclusions and Outlook

In this contribution we have introduced a placement web-service which has been developed in the context of the TENCompetence Integrated project. While we think that the work presented here is an important contribution for formal and non-formal contexts in technology-enhanced learning, there are several limitations of our approach. Since the services focuses on word usage in texts it can only work in domains where textual expression is important. Another limitation of using LSA for placement support is the inability to recognize more structured information like metadata. For this purpose other techniques have to be applied to allow for a matching like e.g. case-based reasoning approaches. The result of the first technical evaluation was that the web service performance was not suited for a productive system. This result was mainly due to scalability and effort issues. Response times higher than 30 seconds could be observed. Although this long reaction time depends to a large extent on the hardware of the Web Service machine we also analyzed improvements of this response time in the architecture. An important architectural aspect which leads to these response times was the combination of updating the Term-Document-Matrix (TDM) and singular value decomposition every time before a query is executed. Since an update of the matrix is only needed when sufficient new content has bee added to the learning network or the portfolio of the learner this process does not have to be executed before every query. For further development of the web-service we will work on the performance issues and we will split the process of updating the TDM and execution of the query. This is a well know problem in LSA research [Zha & Simon 99] and several solutions for this problem have been proposed which will be explored in the future.

Another important aspect of future work is related to the effort to set up the web-service to work correctly. Although we could evaluate a method to improve the web-service results sufficiently on a data set from psychology studies a new “training phase” will be needed for implementations in other context. For the further development of the web-service a rating system should be introduced that aligns the behaviour to the ratings of web-service users. For this purpose an adjustment layer should be implemented that can “learn” from the ratings of the results by users.

While we have focused in this project only on the application of Latent Semantic Analysis for prior knowledge approximation we combined text and data mining approaches in another recent project. The Semantic Weblog Monitoring Framework (SWeMoF) will enable users to conduct similarity analysis, classification and clustering experiments on corpora generated via RSS-feeds from weblogs [Kalz et al 09c]. With this framework we will be able to improve the construction of test corpora on the one hand and we will extend our approach on other methods and approaches that go beyond the application of Latent Semantic Analysis alone.

Acknowledgements

The work on this contribution has been sponsored by the TENCompetence Integrated Project that is funded by the European Commission's 6th Framework Programme, priority IST/Technology Enhanced Learning. Contract 027087 (www.tencompetence.org)

References

[Bochkanov and Bystritsky 08]. Bochkanov, S. and Bystritsky, V. ALGLIB libraries, 2008. http://www.alglib.net/

[Brusilovsky and Henze 07] Brusilovsky, P. & Henze, N. Open Corpus Adaptive Educational Hypermedia. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.) The Adaptive Web. Springer, 2007, ISBN 978-3-540-72078-2.

[Foltz, Laham, & Landauer 99] Foltz, P., Laham, D., & Landauer, T. (1999). Automated essay scoring: Applications to educational technology. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (pp. 939-944). Chesapeake, VA: AACE.

[Giles, Wo, Berry 03] Giles, J.T., Wo, L. and Berry, M.W. (2003). GTP (General Text Parser) Software for Text Mining. Statistical Data Mining and Knowledge Discovery, H. Bozdogan (Ed.), CRC Press, Boca Raton, FL, 2003.

[Graesser, Chipman, Haynes, & Olney 05] 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.

[Haldane et al. 07] Haldane, A., Meijer, R., Newman, C., Wallace, J. To what extent can the APEL process be facilitated through the use of technology? Work-based Learning Futures. Proceedings from the Work-based Learning Futures Confierence, Buxton, UK 2007.

[Herder & Kärger 08] Herder, E., & Kärger, P. (2008). Hybrid Personalization for Recommendations. In J. Baumeister & M. Atzmüller, Proceedings of the 16th Workshop on Adaptivity and User Modeling in Interactive System, (pp. 20-25). October, 6-10, 2008, Würzburg, Germany: Universität Würzburg.

[Kalz, Hornung-Prähauser and Koper, 08] Kalz, M., Hornung-Prähauser, V. & Koper, R. Proceedings of the 1st Workshop on Technology Support for Self-Organized Learners. June, 2-3, 2008, Salzburg, Austria: CEUR Workshop Proceedings, ISSN 1613-0073.

[Kalz et al. 08a] Kalz, M., Drachsler, H., Van Bruggen, J., Hummel, H. G. K., & Koper, R.. Wayfinding Services for Open Educational Practices. International Journal of Emerging Technologies in Learning, 3(2), 2008, 24-28.

[Kalz et al. 08b] Kalz, M., Van Bruggen, J., Giesbers, B., Eshuis, J., Waterink, W., & Koper, R. (2008). A Model for New Linkages for Prior Learning Assessment. Campus Wide Information Systems, 25(4), 233-243.

[Kalz et al. 09a] Kalz, M., Van Bruggen, J., Rusman, E., Giesbers, B., & Koper, R. (submitted). Latent Semantic Analysis of Small-Scale Corpora for Placement in Learning Networks.

[Kalz et al. 09b]. Kalz, M., Van Bruggen, J., Waterink, W., Eshuis, J., & Koper, R. Where am I? – A Study on Learner Placement based on Semantic Similarity. Journal of Computer Assisted Learning.

[Kalz et al 09c] Kalz, M., Beekman, N., Karsten, A., Oudshoorn, D., van Rosmalen, P., van Bruggen, J., & Koper, R. (accepted). SWeMoF: A semantic framework to discover patterns in learning networks. Proceedings of the Fourth European Conference on Technology-Enhanced Learning. Nice, France.

[Klamma et al. 07] Klamma, R., Chatti, M. A., Duval, E., Hummel, H., Hvannberg, E. H., Kravcik, M., Law, E., Naeve, A., & Scott, P. (2007). Social Software for Life-long Learning. Educational Technology & Society, 10 (3), 72-83.

[Kamtsiou et al. 08] Kamtsiou, V., Naeve, A., Kravcik, M., Burgos, D., Zimmermann, V., Klamma, R., Chatti, M. A., Lefrere, P., Dang, J., & Koskinen, T. (2008). A Roadmap for Technology Enhanced Professional Learning (TEPL). Prolearn Consortium. Deliverable D12.15.

[Karampiperis & Sampson 05] Karampiperis, P., & Sampson, D. (2005). Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. Educational Technology & Society, 8 (4), 128-147.

[Koper, Sloep and Rusman 05] Koper, R., Sloep, P., Rusman, E. Effective Learning Networks. Lifelong Learning in Europe, 1, 2005, 18-27.

[Koper & Specht 08] Koper, R., & Specht, M. (2008). Ten-Competence: Life-Long Competence Development and Learning. In M. Sicilia, *Competencies in Organizational e-learning: concepts and tools* (pp. 234-252). Hershey: IGI Global.

[Küster et al. 08] Küster, U., König-Ries, B., Petrie, C., & Klusch, M. On the Evaluation of Semantic Web Service Frameworks. International Journal On Semantic Web and Information Systems, 2008, 4 (4).

[Landauer 2008] Landauer, T.K. 2007. LSA as a theory of meaning. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of Latent Semantic Analysis (pp. 379-400). Mahwah, NJ: Erlbaum.

[Lindstaedt, Scheir, Lokaiczyk, Kump, Beham & Pammer 2008] Lindstaedt, S. N., Scheir, P., Lokaiczyk, R., Kump, B., Beham, G., Pammer, V. Knowledge Services for Work-integrated Learning in: Proceedings of the European Conference on Technology Enhanced Learning (ECTEL) 2008, Maastricht, The Netherlands, September 16-19, pp 234-244

[Rohde 05] Rohde, D. (2005). Doug Rohde's SVD C Library, Version 1.34./

[Zha & Simon 99]. Zha, H. and Simon, H. D. On Updating Problems in Latent Semantic Indexing. *SIAM J. Sci. Comput.* 21, 2 (Sep. 1999), 782-791.