Evaluation of personalized search for learning resources

TOMAŽ KLOBUČAR
Laboratory for Open Systems and Networks
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana
SLOVENIA

Abstract: - One of the problems in technology-enhanced learning is how to efficiently find in the network of educational nodes those learning resources that contribute most to learner's personal goals. This paper presents a personalized search component that enables a learner to find such learning resources in the network, and a methodology and results of its evaluation. The component has been developed in the EU IST ELENA project as part of the HCD (Human Capital Development) Suite. We have tested its usability, i.e. effectiveness, efficiency and user satisfaction. The testing results show some interesting aspects that have effect on personalized search performance. Based on these results improvements are proposed and some guidelines for such components are given.

Key-Words: - Personalization, learning resources, search, evaluation, learning repositories, ranking

1 Introduction

Recently, a number of learning resource (LR) repositories have been set up in the world, for example EducaNext (www.educanext.org), Ariadne, Merlot, eduSource, NIME, EdNA, etc. Many digital libraries are also starting to make their resource widely accessible. In such an environment, rich with learning resources, a learner is faced with a new problem: how to efficiently find those learning resources that contribute most to his personal goals or interests.

The EU IST ELENA project (www.elena-project.org) developed several systems and standards to solve the problem [7]. The main goal of the project was to create a Smart Space for Learning, which aimed at managing the distribution and consumption of learning services via a personal learning assistant (PLA). One of the instances of a PLA developed within the project for corporate settings is called HCD (Human Capital Development) Suite (http://www.hcd-online.com). It has been designed to support goal-driven human capital development process, and supports organizations and their employees in searching for, selecting and contracting learning services, as well as in managing the whole learning life-cycle.

In order to connect different learning resource repositories and enable federated search Simple Query Interface (SQI) has been developed [8]. SQI is a common interface for querying learning repositories. It was accepted by CEN/ISSS workshop on Learning Technologies last year and has recently become an official CEN/ISSS Workshop Agreement. A set of heterogeneous educational nodes and networks have already been connected by means of SQI. The network includes nodes such as the learning brokerage system EducaNext, a gateway to Amazon.com, a commercial learning management system CLIX, a German educational network ULI Campus, and course providers and learning resources catalogues in Austria and Germany, e.g. Knowledgebay, LASON and Seminarshop [9]. Recently, iCamp [5] (www.icamp-project.org) has integrated SQI into several open source learning management systems, in particular Moodle, .LRN, IVA, Course Online, as well as provided an SQI gateway for OAIster (oaister.umdl.umich.edu/o/oaister/).

One of the main parts of the HCD Suite is a component for searching for LRs in a network of learning repositories via SQI. The component enables a learner to find in the network of educational repositories those learning resources that contribute most to the learner's personal development plan. Results presentation is improved by implemented ranking algorithms and learner's personal information (interests, goals, learning history).

In this paper we present the evaluation results of this personalized search component. The main purposes of the trials were to test its usability, i.e. effectiveness, efficiency, and satisfaction. The paper is organized as follows. In Section 2, the search component is presented. A test methodology is described in Section 3 and test results in Section 4. At the end, conclusions are given.

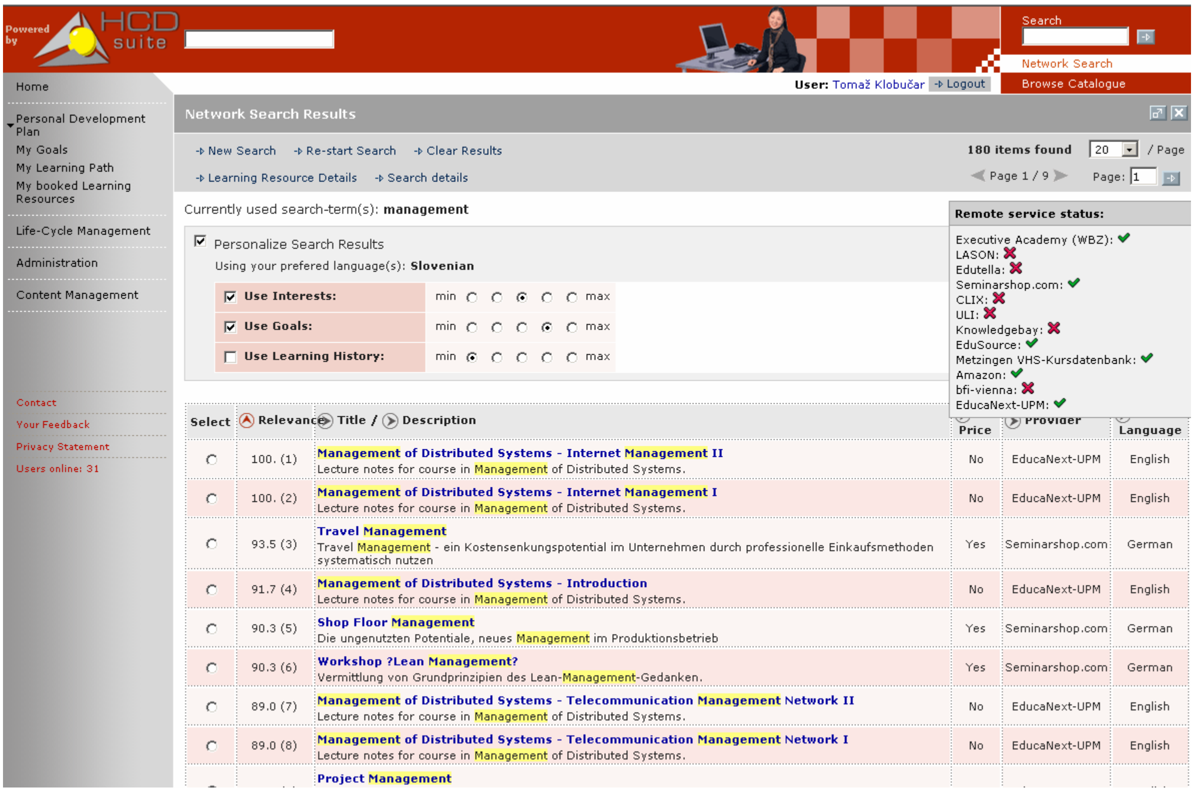

Fig. 1: User interface of a personalized search engine

2 Personalized search component

ELENA personalized search component can be seen as a federated search tool that improves the search results by taking into account user profiles and by implementing mechanisms such as ranking algorithms. Federated search is defined as "support for finding items that are scattered among a distributed collection of information sources or services, typically involving sending queries to a number of servers and then merging the results to present in an integrated, consistent, coordinated format" [6]. It allows learners to search across a number of learning repositories simultaneously.

The main goal of personalized adaptive learning systems is to increase effectiveness and efficiency of learning by taking into account learner profiles. In the ELENA search component, query definitions and ranking of the results are adapted to learner goals and interests. Querying a particular LR, the same list of results is presented to all users, but the order in which the items are to be presented is specifically tailored for each individual user and depends on criteria such as the user's goal, interests or foreknowledge.

The personalized search component is used in the following way:

- Prerequisites – a learner first specifies his personal goals and interests in his profile.
- Query definition and modification – the learner creates a query by specifying keywords as search terms. The query can be further restricted by the category (material, activity), price, copyright and other restrictions, and LR language.
- Querying educational repositories - the query is sent to several educational repositories that are integrated in the network, such as Amazon, CLIX, EducaNext-UPM, Knowledgebay, LASON, Metzingen VHS-Kursdatenbank, Weiterbildungszentrum WU Wien, Seminarshop or BFI Wien.
- Resource merging, ranking and selection - retrieved results from a set of repositories are merged and ranked on the basis of LR metadata and learner profile.

A user interface of the search engine with displayed search results is given on Fig. 1.

During the trials, the users had an opportunity to search through a variety of learning activities and learning material from several providers. Table 1 shows the number of learning resources in the educational nodes that were part of the test network and their categories (LM – learning material; LA – learning activity).

Table 1: Network nodes

Node	No. of LRs	Category	Language
Amazon	millions	LM	Various
BFI	350	LA	German
EducaNext	730	LM, LA	Various
Knowledgebay	150	LA	German
LASON	350	LA	German
Metzingen	500	LA	German
Seminarshop	10000	LA	German
WBZ	400	LA	German

3 Trials

The International Standards Organisation (1998) defines usability in the following way: "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." [4] The main objective of our evaluation was thus to test effectiveness, efficiency and satisfaction of use of the personalized network search component. The following 4 questions were investigated during the trials:

- *Question 1* – Does ELENA search component enable a learner to find relevant learning resources in the ELENA network?
- *Question 2* – Are implemented ranking algorithms effective, i.e. do most relevant learning resources appear on the top of the result list?
- *Question 3* – How does knowledge about the learner, i.e. his goals, interests, preferences, history or other information from the learner profile, change ranking of the results?
- *Question 4* – Is ELENA search component interface user friendly?

During the trials we used qualitative and quantitative methods. Methods for investigating the questions were the following:

- *Question 1*: Observing results of different types of queries, e.g. disjunctive queries, multi-terms, and search attributes, as well as measuring number of returned results and precision (fraction of retrieved results that are relevant).
- *Question 2*: Measuring relevancy of the top 10 ranked results, measuring the RHL indicator [2] and relevancy of the top 15 ranked results, and measuring distribution of the most relevant results through the list of all results
- *Question 3*: Comparing the ranking with and without personalization and comparing different adaptation options (goals, interests)
- *Question 4*: Questionnaires (standard questionnaire on usability: SUS – System Usability Scale; questions on search priorities and missing attributes)

A total of 13 people participated in the trials from Slovenia, Austria, Germany and Spain. The participants were mostly employees at research and academic institutions with rich experience in search engines and information and communication technology. Five of them were directly involved in the ELENA project work. Gender ratio: 10 males and 3 females.

The study was performed in 2005. As most of the users were researchers, they were first asked to imagine working in a company in a particular position, and to think of the learning goals and interest they might have in this situation. The test users then entered into their learner profiles at least 3 goals and 3 interests. The goals were free text descriptions, while interests were selected topics from 2-level taxonomy and accompanying descriptions. Two examples of the described goals are:

Information systems access control - *To investigate and learn about the most recent technologies and frameworks for access control to information systems, such as security policies, authentication mechanisms (e.g. biometrics), authorization mechanisms or privacy-enhancing technologies.*

Understanding macroeconomics - *I would like to understand how the economics works, how the world economics progresses, what factors play roles whether the economics of a particular country prospers.*

Below we also give two examples of interests (described by a type, classification and description):

Type: *Professional Interest*
Classification: *Network technology, network security*
Description: *Security services, mechanisms for data protection and privacy provision, digital signature, access control, public key infrastructure, privacy-enhancing technologies, digital certificate*
Type: *Private Interest*
Classification: *Communication training*
Description: *English*

After describing the goals and interests, the users performed several search tasks. A few search tasks were predefined, while in the other cases, the users were asked to find learning resources that contribute to their specified goals. For each search, the users entered one or more search keywords (as conjunction or disjunction) and restricted the search on the basis of the LR category, price (yes or no), copyrights and other restrictions (yes or no) and language.

3.1 Round 1 of the tests

For each of the tasks, the users counted the total number of results and the number of all results relevant for them, and selected the five most relevant results. Ranking was compared in four cases: when no personalization was used, when only interests were taken into account, when the ranking was adapted to user's goals and when both interests and goals were used as the basis for ranking. For each of the ranking, the users were asked to count how many of the top 10 ranked results were relevant for them, and to write down the ranking places of the previously selected 5 most relevant learning resources. The number 10 was chosen because most of the search engines usually show by default 10 results on a page. After the user had performed all tests, they filled in a post-test questionnaire.

In total, the users performed 47 complete search tasks. Apart from those tasks, additional partial tasks were performed to test effectiveness, i.e. whether the search component returns correct results (e.g. comparison with the results returned by native search engines), how it supports disjunctive queries, how it supports multiple words versus phrases, how the results are displayed, ranking adaptation only to certain information, etc.

3.2 Round 2 of the tests

In the second round two assessors performed another set of 23 goal-related tasks. The assessors looked at the 15 top ranked results for each query, and for each of those results wrote down, whether the result is very relevant, partly relevant or not relevant. Ranking was compared in two cases: when no personalization was used and when the knowledge about the goals was taken into account.

4 Results

4.1 Effectiveness and efficiency

4.1.1 Round 1

Average number of the returned results in *Round 1* was 28.9, while the average number of the relevant results for the users was 9.3. The number of results heavily depended on the number of LRs in a particular node of the ELENA network, result limits (Amazon.com returns only 10 results) and quality of a search term (general vs. specific one). Lower precision (fraction of retrieved results that are relevant) can perhaps be explained by multiple keywords, which have mostly been used by the test users. Some nodes, for example Seminarshop.at, support only disjunctions, i.e. adding further keywords to the search term only increases the number of results.

Ranking without personalization placed on average 5.7 relevant results among top 10 ranked. The most important result for a user was top ranked 9 times (19 % of the time), on the second place 6 times, and on the third place 7 times (i.e. 47 % of the time among top 3).

For the tasks that were related to users' interests and goals, we calculated the ranking differences with and without personalization applied: for the number of relevant results among 10 top ranked results and for the rankings of the 5 most relevant learning resources. Quantitative analysis shows that for the selected 25 interest-related tasks and 24 goal-related tasks, *on average* there is practically no difference when the personalization is switched on or off. In the case of the goals, for example, the number of relevant results among top 10 ranked increased is almost exactly the same, while the average ranking of the 5 most relevant LRs is slightly worse after personalization (by 0.22 place). Out of 112 monitored rankings, 46 were the same, 26 better and 40 worse.

4.1.2 Round 2

In *Round 2*, we measured the number of relevant LRs the search component retrieves and places among the top 15 results, and investigated the Ranked Half-Life (RHL) indicator [2], a positional measure of ranked retrieval results. RHL shows the ability of the system to place relevant results high in the ranked list of retrieved results. The RHL value is the median case, i.e. the point which divides the continuous area exactly into two parts. Each LR in the algorithmically ranked list represents a class of grouped data where the frequency equals the assigned relevance value. The lower the RHL value, the higher the relevant LRs are placed in the ranked output, i.e. the better the retrieval engine [1]. In order to make the RHL value more comparable, it can be normalized into the RHL index, i.e. divided by the fraction of the relevant results among the top 15 ranked LR. For the exact formula how the RHL value is computed, we refer the reader to the two papers mentioned in this paragraph. As a summary of the quantitative results of Round 2, we can say that the average number of relevant results among the top 15 ranked LRs was 8.8. The average RHL index value was slightly lower (better ranking results) when no personalization was used (RHL index = 11.71) than when the goals were taken into account (RHL index = 12.32).

4.1.3 General observations
More interesting than quantitative analysis is perhaps the analysis of a few single examples that can explain the quantitative results and suggests future improvements. The following two patterns have been identified:

- The results rankings improve if a user gives a general search keyword and the goal description significantly narrows down the result. For example, if user's goal is to learn about risk management and he searches for LRs about management, after personalization LRs about risk management are ranked higher at the top.
- The results can become worse when a user has several heterogeneous goals, and information about other goals affects rankings for particular goals. For example, one of the users had two goals; one of them was to investigate IT access control and the other was related to risk and project management. When a user entered access control as a search term and switched on personalization feature, the highest ranked LR became IT project management, because it was related to the other user's goal.

The second pattern is mainly a user interface issue of the current implementation, i.e. the user does not have an option to specify to which goals he would like to apply personalization. If the learners had possibility to restrict personalization to specific goals, the total average results would improve and prove successfulness of the applied ranking algorithm.

4.2 User satisfaction
User satisfaction was tested by a standard SUS (System Usability Scale) questionnaire. SUS yields a single number representing a composite measure of the overall usability of the system being studied. SUS scores have a range of 0 to 100, where 100 means that all respondents strongly agreed with all questionnaire items, (the system is very effective, efficient and satisfactory) and 0 that they strongly disagreed on all the items. Value of the SUS questionnaire for a personalized search component was 66.9, which tells that the users perceived the search component as quite usable (on average they agreed on questionnaire items).

Important information for further development of the personalized search component and LR metadata scheme are users' wishes for new search attributes. Network search interface currently offers a learner to restrict category (learning material or learning activity), price (yes or no), copyright and other restrictions (yes or no) and language. Among the missing attributes, the users, on average, assigned the highest priority to:

- LR format, e.g. Power Point presentation, movie, book, seminar, and
- Favorite results, i.e. they would like to find resources that other users judged as good.

As priorities depend on the type of learning resource, it might be a good idea in the future to separate an interface according to the LR category. For example, copyright and other restrictions are not that important for learning activities (e.g. seminars), comparing to learning material. More important meta-data for learning activities are the schedule (when do they take place) or location (where do they take place).

5 Conclusion
In general, the users liked the idea of combining a dedicated network of educational nodes, searching with restrictions that narrow down what they were looking for, and personalization of the results based on their interests and goals. The component can become a nice tool for the future for finding best learning offers. Additional nodes need to be integrated into the network to increase the number on non-German learning offers. This is currently being done in the EU PROLEARN (www.prolearn-project.org) [11] and iCamp projects. Success of semantic search heavily depends on the metadata quality of the resources. Since at the moment all nodes in the network support a small subset of the attributes that enable a user to restrict search for, it

might be necessary to further investigate user search priorities for different categories of learning resources.

Search component currently does not use personalization in the preparation of a query, but just for ranking of the results. The result set thus depends mostly on the quality of search terms and selected restrictions the user explicitly selects during the searching, and not on the learner's personal profile. This might be a problem when a general search produces too many results (performance), or when some of the educational nodes return only a limited number of results, e.g. Amazon in our case that returns only 10 results. An important aspect of query rewriting on the basis of personal information is of course privacy, as the search component should not leak sensitive personal data of the learner.

With regard to the personalized ranking it should be further investigated in which situations the personalization is useful, and what kind of information (content, form) is necessary in the learner profiles. These trials have shown that on average, taking into account usual queries, and goals and interests in free text, personalization has little effect on the ranking. The reasons seem to be that the learners already provide very specific query, or they have multiple heterogeneous goals in their profiles. The effect is clearly visible when search terms are general and information in the profile is used to significantly narrow down the results.

To allow assessing the effectiveness of the chosen algorithm in comparison to other existing algorithms (especially non-semantic retrieval algorithms), it might be useful to create a marked-up test corpus including pre-formulated queries and corresponding relevance judgments. The currently offered collections in TREC [10] and CLEF [3] unfortunately do not have the semantic mark-up needed to evaluate the strength of the ELENA common schema and the connected retrieval algorithms. In the future, further larger scale trials (larger in terms of the number of users and duration of the tests) should also be performed in the real work environments.

References:

[1] Blomgren, L., Vallo, H., Bystrm, K., Evaluation of an information system in an information seeking process. In: *Proceedings of ECDL 2004*, 2004, pp. 57–68.

[2] Borlund, P., Ingwersen, P., Measures of relative relevance and ranked half-life: Performance indicators for interactive IR. In: *Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval*, 1998, pp. 324–331.

[3] Cross Language Evaluation Forum (CLEF), http://clef.isti.cnr.it/.

[4] International Standards Organisation, ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability, 1998.

[5] Jerman-Blažič, B, Klobučar, T., Arh, T., Kieslinger, B., Wild, F., Lai-Chong Law, E., An approach in provision of interoperability of eLearning systems in enlarged EU - the case of iCamp project. In: *Advanced international conference on telecommunications and International conference on internet and web applications and services (AICT/ICIW 2006)*: 19-25 February, Guadeloupe, French Carribbean. Piscataway: IEEE, 2006, 7 pages.

[6] Ribeiro-Neto, B., Baeza-Yates, R.. *Modern Information Retrieval*. NY: ACM Press, 1999.

[7] Simon, B., Dolog, P., Miklós, Z., Olmedilla D., Sintek, M., Conceptualising Smart Spaces for Learning. *Journal of Interactive Media in Education, 2004 (9). Special Issue on the Educational Semantic Web*. ISSN:1365-893X, 2004.

[8] Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., and Miklos, Z., A Simple Query Interface for Interoperable Learning Repositories, in: *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems*, Chiba, Japan, May, 2005, pp. 11-18.

[9] Simon, B., Sobernig, S., Wild, F., Aguirre, S., Brantner, S., Dolog, P., Neumann, G., Huber, G., Klobucar, T., Markus, S., Miklós, Z., Nejdl, W., Olmedilla, D., Salvachua, J., Sintek, M., Zillinger, T., Building Blocks for a Smart Space for Learning. In: *Proceedings of ICALT'06*, Kerkrade, Los Alamitos: IEEE Computer Society, 2006, pp. 309-313.

[10] TREC (Text REtrieval Conference), http://trec.nist.gov/.

[11] Wolpers, M., Grohmann, G., PROLEARN: Technology Enhanced Learning and Knowledge Distribution for the Corporate World, *Int. J. Metadata, Semantics and Ontologies*, Vol. 1, Nos. 1/2, 2005, pp. 44-61.