Journal of Universal Computer Science, vol. 22, no. 8 (2016), 1175-1200
submitted: 1/7/15, accepted: 29/7/16, appeared: 1/8/16 © J.UCS

RecSim: A Model for Learning Objects Recommendation using Similarity of Sessions

Tiago Wiedmann
(University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil
tiagowrs@gmail.com)

Jorge Luis Victória Barbosa
(University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil
jbarbosa@unisinos.br)

Sandro José Rigo
(University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil
rigo@unisinos.br)

Débora Nice Ferrari Barbosa
(FEEVALE University, Novo Hamburgo, Brazil
deboranice@feevale.br)

Abstract: A learning object (LO) is any entity or resource that can be used in computer-aided learning. This can take the form of text, multimedia content, presentations, programs or any other type of digital content, generally made available through web portals or distance learning systems. The LOs consulted by a student while accessing such portals are related to the interests of the student for the duration of the session. This article proposes a model for LOs recommendation using similarity of sessions, called RecSim. The model receives the sequence of LOs consulted during the current user session along with sessions whose sequences are similar to the LOs consulted in the current session. LOs found in similar sessions are then recommended to the user. A prototype was developed and applied into two controlled experiments. The results were encouraging and show potential for implementing RecSim in real-life situations.

Keywords: Recommendation Systems, Similarity Analysis, Learning Objects.
Categories: L.1.2, L.3.0, L.3.2

1 Introduction

A learning object (LO) is any resource that can be used in computer-aided learning. Learning objects can take the form of text, multimedia content, presentations, programs or any other type of digital content [Polsani, 2003]. These objects are made available on the web in order to be reused in different educational settings, thereby reducing the costs of production.

The increasing number of available LOs and the growth of the web itself have created the necessity of organized databases for storing and cataloguing. These online databases, known as *repositories*, provide LOs to students and educators and are available individually or organized into thematic groups. However, even with organization of LOs in dedicated repositories, finding objects that meet the needs of users is still a complex task, given the volume of information available. To solve this problem, *recommendation systems* [Bobadilla et al., 2013; Weng et al., 2009] were developed in order to assist searching for LOs that correspond to the user profile [Cazella et al., 2010]. The profile is a structure containing information on user preferences, behavior and context [Ghosh and Dekhil, 2009; Wagner et al., 2014].

In general terms, a recommendation system is able to recommend useful resources to users or groups by means of an assessment that determines the usefulness of a particular recommendation to the user [Bobadilla et al., 2013; Resnick and Varian, 1997; Weng et al., 2009]. In education, the utility of a recommendation or resource can be estimated from the content of the object together with the profile of the user who is performing the search. For example, Al-Khalifa [2008] proposes a recommendation algorithm that analyzes each user profile together with the history of all LOs that have been accessed and the ranking given to each object by the user.

However, analysis based on the profile considers the history of LOs consulted by the user, or even general characteristics of the profile, such as the area of expertise. This approach may not meet user expectations since it does not take current learning interests into consideration. This has led researchers to study context-aware approaches to recommendation [Barbosa et al., 2011; Moore et al., 2010].

This article proposes a model for LOs recommendation called RecSim (*Recommendation Based on Similarity of Sessions*), which considers the similarity between the sequence of objects consulted during the current user session with the sequences stored in the repository's sessions history. RecSim deals with LOs recommendation as a context-aware application [Dey, 2001; Bellavista et al, 2012; Knappmeyer et al., 2013].

According to the definition given by Dey, Abowd and Salber [2001], a context is considered a set of information describing the status of an entity. In the specific case of LOs, each object accessed by a student during a classroom activity or individual search constitutes the context of that student at a specific time.

To evaluate the similarity between two objects, several authors have proposed different techniques. In general, these techniques evaluate attributes of the objects under analysis in order to calculate a numeric value that determines how near or far the two objects are from one another. Furthermore, in the case of time series [Fu, 2011] we can apply techniques that individually evaluate each point of the series as well as techniques such as *Dynamic Time Warping* (DTW), which calculates an average time series between two series [Berndt and Clifford, 1994].

This article discusses related works on the topics *context sequence similarity* (section 5.1) and *learning objects recommendation* (section 5.2). Based on this literature review, we conclude (section 5.3) that RecSim contribution is the use of DTW to conduct real-time LOs recommendation during an access session in a content repository.

The remainder of this article is organized as follows. First, the background concepts are introduced in section 2. Section 3 delineates the RecSim model. Section 4 approaches evaluation aspects, describing experiments and results. Section 5 discusses related works focusing on contribution of the proposed model. Finally, section 6 presents the concluding remarks and directions of future works.

2 Theoretical Foundation

This section approaches two strategic themes to understand the proposed model. Section 2.1 discusses concepts and research on context awareness and context history. In turn, section 2.2 focuses on similarity analysis, mainly addressing the *Dynamic Time Warping* technique (DTW) [Berndt and Clifford, 1994].

2.1 Context awareness and Context history

Brown et al. [1997] state that context-aware applications are those whose operation is primarily directed by the current context of the user. According to Schilit and Theimer [1994], the context consists of data regarding location, the identity of people and nearby objects, and changes in these objects. Bellavista et al. [2012] and Knappmeyer et al. [2013] present surveys focusing on aspects of context awareness. RecSim adopts the definition of Dey, Abowd and Salber [2001], where context is any information that can be used to characterize the status of entities that are considered relevant for the interaction between a user and an application, including the user and the application themselves. Contexts are typically: location, identity and status of people, groups and physical and computational objects.

In this article, the goal of ascertaining context is to determine the user's learning interests in an LO repository access session. Since it is difficult to determine this information directly, contextual data can help the application to estimate subjects of interest to the user and then provide recommendations based on these estimates. We consider not only the set of LOs accessed, but also the order in which they were accessed so as to take into account the student's learning strategy. The sequence of LOs accessed by users and recorded in the repository history therefore corresponds to the sequence of activities performed by the user within their program of study, where each activity corresponds to the context of the student at any given time.

The idea of a sequence of contexts visited by an entity over time is known as *context history* [Hong et al., 2009] or *trail* [Driver and Clarke, 2008; Silva et al., 2010; Cambruzzi et al., 2015]. According to Clarke and Driver [2004], a trail is a collection of locations accompanied by the associated information and a recommended visiting order. Spence, Driver and Clarke [2005] extend this definition by considering that the history of recommended trails can also be analyzed. The authors use *trail histories* to refer to information on how the user made use of trails suggested in the past. Subsequently, in Driver and Clarke [2008], a trail is described as a collection of activities to be performed by the user, similar to a *to-do list* but whose sequence of execution is ordered according to the context and the user's preferences.

Another way to view the sequences of contexts is to consider them as *time series*. According to Fu [2011], time series are sets of observations in chronological order. Esling and Agon [2012] argue that in almost all areas of science it is necessary to measure variables over a period of time, leading to the creation of these sets of ordered data. An LO accessed at a particular time is not a quantifiable variable. In this article, however, we are interested in computing the similarity between two LOs at an equivalent point in time in two distinct visiting sequences. We can therefore use a similarity function that evaluates two objects that are, in relative terms, at the same point in two sequences and then return a numeric value that will comprise the time series for similarities between the two sequences of LOs accessed.

2.2 Similarity Analysis

According to Cha [2007], from a mathematical and scientific point of view, distance or dissimilarity is a value that expresses how far two objects are from each other. Quantitative measurements are generally referred to as metrics, while other non-metric measurements are sometimes known as divergences. The similarity or proximity between objects is known as the similarity coefficient.

RecSim uses the *Dynamic Time Warping* (DTW) [Berndt and Clifford, 1994] similarity function. This function is used in the analysis of similarity between time series, such as the analysis of share price history and speech recognition, for example.

Considering two time series, $S = \{s_1, s_2, ..., s_i, ..., s_n\}$ and $T = \{t_1, t_2, ..., t_j, ..., t_m\}$, *DTW* aligns both series in the form of a *n* x *m grid*, where *n* and *m* are the sizes of the respective series and each point $\{i,j\}$ of the *grid* corresponds to an alignment between the elements of the series.

The value of each point is obtained by measuring the distance δ between the points, as the magnitude of the difference ($\delta(i,j) = |i - j|$) or the square of the difference ($\delta(i,j) = (i - j)^2$). The distance is calculated using dynamic programming techniques to find the sequence of points in the grid $W = \{w_1, w_2, ..., w_k\}$ that correspond to the smallest sum of distances.

According to Berndt and Clifford [1994], this can be formally defined as a minimization problem over all possible *Warping Paths*, formalized as shown in equation 1, where δ is the distance function at point *wk* in the grid and *p* is the last point of comparison between the series.

$$DTW(S, T) = min_w \left[\sum_{k=1}^{p} \delta(w_k) \right] \qquad (1)$$

To solve this type of problem, we need to use a stage variable, state variables and decision variables. In the case of DTW, the stage variable imposes a monotone order on the events represented by the series and is given by the time dimension of the series. The state variables are the individual points of the grid. And the decision variables, which define the transitions of possible states, correspond to the restrictions that define the paths that can be traversed, thereby reducing the size of the space of solutions, known as the *Warping Window*.

The solution to the minimization problem is obtained using the recursive relationship $\gamma(i,j) = \delta(i,j) + \min[\gamma(i - 1,j), \gamma(i - 1, j - 1), \gamma(i, j - 1)]$ that calculates the cumulative distance for each point of the grid that is within the warping window. The cumulative distance is given by the sum of the distance between the current point and the shortest cumulative distance from its neighboring points. This is a symmetric algorithm since both the preceding diagonal points, $(i - 1, j)$ and $(i, j - 1)$, are used.

3 The RecSim Model

This section describes the proposed model. Section 3.1 discusses the principles and architecture of RecSim, focusing primarily on its organization in layers. Section 3.2 approaches the recommendation process, describing mainly how the DTW technique was employed.

3.1 Principles and Architecture of RecSim

RecSim is based on the following principles:

1. the LOs sequence for the current session implicitly indicates the current interest of the user, thereby providing an appropriate basis for the recommendation with respect to the current context;

2. the current session may also implicitly provide the current context of the user, such as, for example, what course or subject is being studied at the time, who the student's classmates are, as well as other inferable information;

3. the number of LOs consulted in a session is usually small and this improves the performance of similarity analysis;

4. a simplified technique is applied, which can easily be adapted to more complex models.

RecSim uses the technique known as DTW to analyze the similarity of the current session with histories of sessions stored in the repository. Each session consists of a sequence of LOs consulted by the user, which the model treats as an artificial time series [Matuschek et al., 2008]. This interpretation determined the decision to use DTW as the similarity analysis technique, since it was developed specifically for the analysis of time series.

The architecture consists of five layers illustrated in Figure 1. The RecSim can be better understood through a bottom-up approach. The bottom layer is the physical storage repository and the layer above is used to access this repository, abstracting the data in an internal LO and session representation. The third layer performs an analysis of the similarity between the current session and the sessions stored in the repository. In turn, the next layer provides the recommendation for a session. The client application uses the *Recommender component*, specifying which component of the data access layer will be used to load the sessions data (*sessionReader*), the maximum number of sessions to be analyzed (*maxSessions*) and the location of the object repository (*sessionURL*). A call is then made to the *recommend method*, transferring the LOs sequence from the current session and the number of LOs to be recommended.

The top layer corresponds to the application that makes use of the RecSim, for example, a module on the repository web portal. The application must be able to represent the LOs consulted during the current session using the *Session* and *LearningObject* classes from the *Data Access Layer*. Optionally, data of a specific session can be loaded from the sessions history by using one of the implementations of *Reader class*.

The *URL parameter* in the *read method* is a string containing the location of the repository. The content of this string depends on the implementation of the *Reader* *class* used and may be the full path to a file or the *jdbc* string for connecting to a database. The application must then use the *Recommender class* from the *Recommendation Layer* in order to obtain the list of objects to be recommended to the user. The class must be initialized by creating an instance of the *Reader class*, which will access the repository, the maximum number of sessions to be read from the history and the URL for accessing the repository. Lastly, a call is made to the *recommend method*, transferring, as a parameter, the objects of the current session and the maximum number of objects to be recommended.

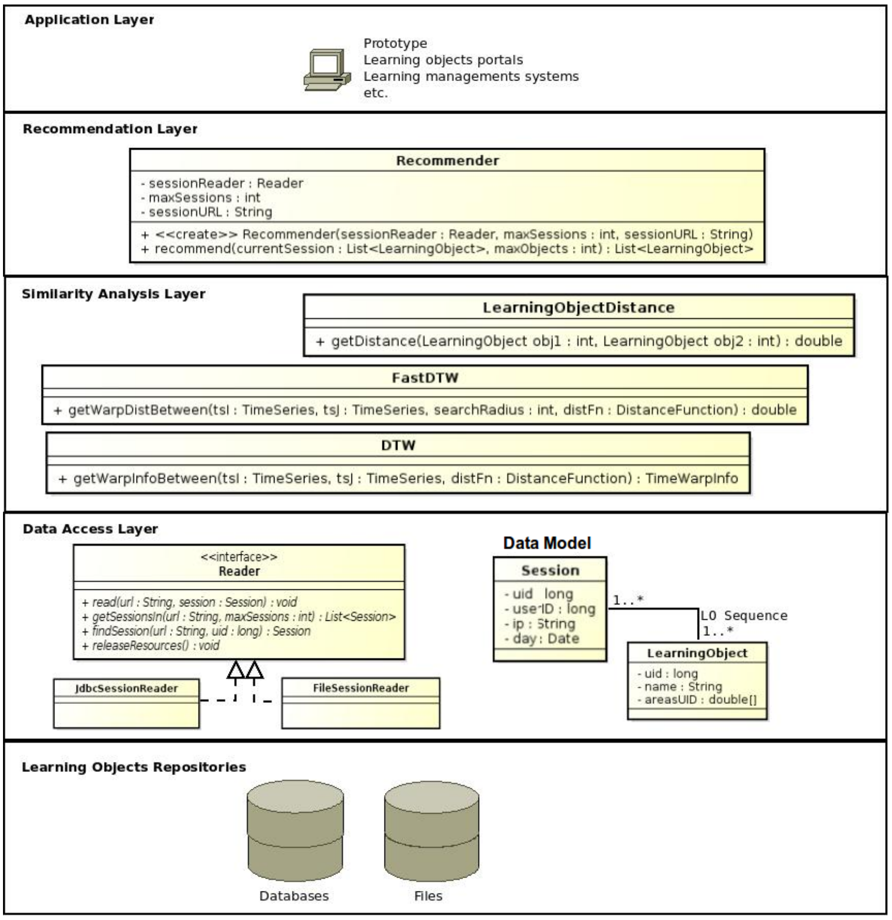

Figure 1: The RecSim Architecture

3.2 Recommendation Process

The recommendation process is performed in two steps. In the first step, sessions that are most similiar to the current session are identified using the DTW technique. Initially, the sessions are obtained from the repository, limited by the maximum number of sessions to be analyzed, using the *Reader class* and the URL of the specified repository.

The sequence of LOs for each session is then read and represented as a time series, as well as the sequence of current session. The *DTW class* is then used to calculate the value for the distance between the current session series and each series obtained from the history.

Lastly, after calculating the distances for the series in the history, in relation to the current series, the sequences are sorted in ascending order of distances and used in the second step for obtaining the list of objects to be recommended.

In the second step, the ordered list of similar sessions is examined and each LO that is not present in the sequence of LOs consulted in the current session is then selected to be recommended to the user. This process is repeated until the maximum number of recommendations is reached or until the end of the sessions list.

The *Similarity Analysis* layer calculates the distance of each session found in the repository, in relation to the current session, using the DTW technique. Initially, both sequences of LOs are treated as time series *S* and *T* (see section 2.2) and are arranged orthogonally, forming a grid *n x m*.

Each point of the grid, within the warping window, takes the value of the distance between the LOs. At the end of this process, the standard algorithm of the DTW technique is applied, which searches for the shortest path among the possible paths in the grid.

For implementing the function of the distance between the LOs, a technique was developed that compares the subject area of the objects, returning a value between (0..1), according to the determined distance. The technique is based on the work of Abraham and Lai [2012], where the equation 2 was applied to determine the distance between object trajectories through sequence alignment of travel locations. RecSim uses the same equation to calculate the distance of hierarchies of subject areas of LOs. The principle on both applications is the same, i.e., the distance of hierarchies is calculated by means of binary comparison between levels. In the case of the RecSim, (x, y) are the LOs to be compared and (X, Y) are the respective hierarchies of the subject area of each object.

$$\delta_{lo}(x,y) = 1 - \frac{X \cap Y}{max(size(X), size(Y))} \tag{2}$$

Table 1 illustrates the calculation of the distance between two distinct LOs obtained from the *International Bank of Educational Objects* [ILOD, 2015].

The distance function implemented by the technique consists of a binary comparison between the hierarchies of the subject areas, dividing the number of equal areas by the size of the largest hierarchy. A characteristic of this technique is that, like the work of Abraham and Lai [2012], the algorithm continues to calculate the binary comparison only when correspondence is found between the levels of the two hierarchies.

The purpose of this control is to reduce the number of false positives during comparison. For example, if an object *X* is linked to the field of "*science::Physics::Atomic Theory*" and an object *Y* is linked to the field of "*science::History of* *science::Atomic Theory*", then the similarity between them occurs only at the top the hierarchy (*science*).

LO	Subject Area
x *y*	Basic Education::Secondary Education::Chemistry::Constitution models Basic Education::Secondary Education::Chemistry::Properties of matter
	Binary comparison: 1 :: 1 :: 1 :: 0 $\delta_{LO}(x,y) = 1 - 3/4 \rightarrow 0.25$

Table 1: Sample Distance Calculation

Lastly, if the compared objects have the same unique identifier (UID) then they are the same object and the distance value can therefore be considered zero, without needing to calculate similarity between the subject areas.

Once the distance function is determined, the DTW technique is used to calculate the similarity of the sequence accessed during the current session in relation to the sessions stored in the repository histories. We can limit the number of sessions compared, considering for example, only the most recent sessions, or sessions of users with profiles similar to the current user. The technique returns the sessions most similar to the current session, in relation to the sequence of consulted LOs.

The *Recommendation* layer then organizes the sessions obtained in order of increasing distance, disregarding sessions with a distance value equal to zero. The reason for this is that zero indicates an LO sequence equal to the current session, and these sequences will not, therefore, contain any new LO to be recommended. Finally, the *Recommender* component examines the sessions list, selecting the LOs that are not contained in the list of LOs for the current session, until the limit of LOs to be recommended is reached, and returns them to the *Application* layer.

4 Evaluation Aspects

In this section we describe the evaluation methodology and two experiments based on the prototype of RecSim.

4.1 Prototype and Methodology

A prototype was developed using the Java programming language and implemented in order to evaluate the proposed model. The data set used in the experiments is composed by a set of real data and a set of data from a synthetic repository. Both data sets were stored in a PostgreSQL database. To implement the DTW technique, we used the open source FastDTW framework, available at *http://code.google.com/p/fastdtw*, which was adapted to implement the distance function used in RecSim.

The experiments were conducted on a DELL VostroTM Notebook, equipped with an Intel CoreTM i5 2.40GHz processor and 6GB of RAM memory. The Ubuntu Linux 12.04 64-bit operating system running kernel version 3.2.0-51 was used as operating system.

The first experiment, described in section 4.2, was developed in order to evaluate the overall process of RecSim, to assess the correctness of results and also to expose some performance and scalability issues. This experiment was conducted with a real data set, obtained from a Moodle Learning Environment.

The second experiment used a controlled scenario to obtain standard metrics values (e.g., precision and recall) in order to evaluate the recommendation performance. The controlled context allowed the adequate conditions to measure the results of our recommendation method. A set of data was generated to simulate the expected situations in a classroom. In this manner, it was feasible to consider different situations expected in real life. As discussed in section 6, a future experiment will apply RecSim in classrooms for a wide period with the participation of students to evaluate the effectiveness of the recommendation by subjective and objective measurements. The second experiment allowed only objective measurements.

4.2 First Experiment

The data used to perform the first experiment was obtained from the access logs of the Moodle system (*http://moodle.org*) used for an undergraduate Information Systems course, during the period from August 2012 to August 2013. The *mdl_log* table records were used to determine students' access to the course materials made available by teachers.

These materials can be considered examples of LOs [Moore, Jackson and Wan, 2010] and are described in the *mdl_resource* table, from which the description of each object was obtained. The subject area of each LO was given by the "*course*" and "*course module*" identification codes, stored respectively in the *course* and *cmid* fields of the *mdl_log table*, thereby creating a hierarchy pattern for identifying the subject area of each learning object. The general form of this hierarchy is exemplified as follow: "*Course::Course_Module*". The "*Course*" element represents the identification of an undergraduate course and the "*Course_Module*" contains the diverse sections or modules of this course.

Figure 2 illustrates the structure of the data used for evaluation. The *mdl_log* table did not originally have a user session identifier. To perform the first experiment, a *session* table and a *session identifier* were artificially created. To construct the *session* table, the data set was grouped considering the following data fields: *User* (*userid*), *IP address* and *access date*. The *datetime* and *day* fields were created to assist in this process.

Based on this procedure, a number of 11.039 user sessions were obtained in a set of 33.778 LO access records. Figure 3 shows the distribution of the number of records (in the y-axis) per user session (represented in the x-axis). We can see that the greater session has 154 records, while the majority of sessions have less than 25 records. This context is also relevant to justify the necessity of an algorithm that is capable of generating results based on small data sets. The real data set analyzed is a valid source that can be used to indicate a tendency regarding the size of the user sessions.

The RecSim algorithm operation starts with one of the sessions available in the database and, based on the sequence of LOs in this session, identifies similar sessions and recommends new objects. In this first experiment, the prototype was configured to recommend three LOs for the selected session. The result generated by the prototype is a processing report, indicating the most similar sessions, the recommended LOs and the processing time.

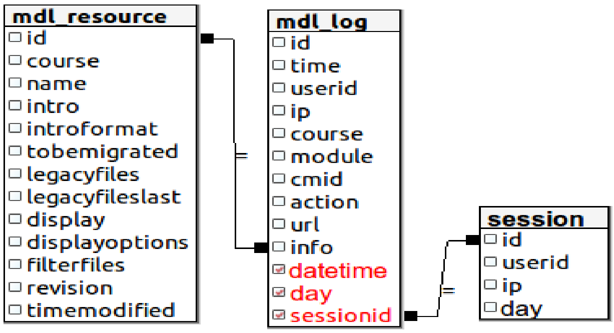

Figure 2: The data model used in the first experiment

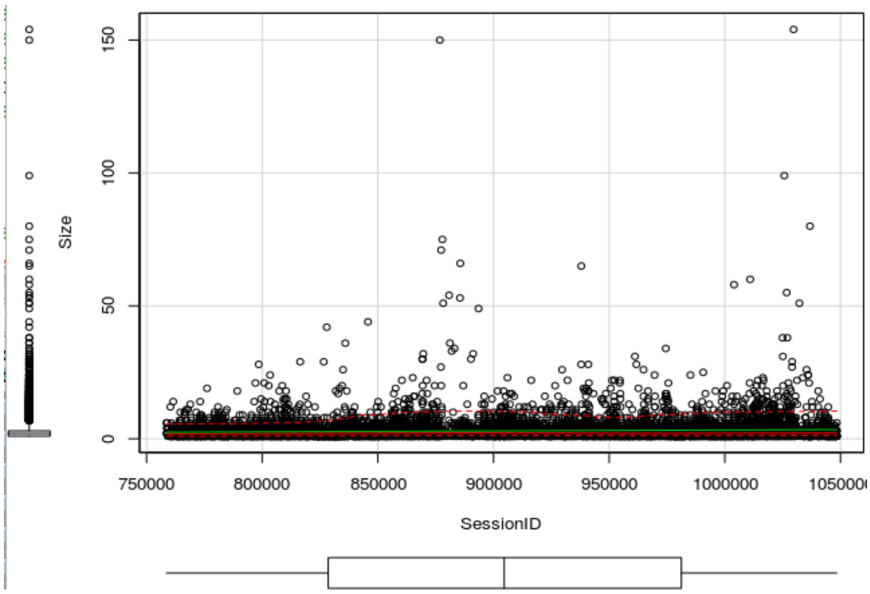

Figure 3: Distribution of the number of records per user session

The FastDTW implements the DTW technique for calculating the distance between two time series, which are encapsulated in a class named *TimeSeries*. This class uses vectors to store internally the numeric values that comprise each series. Some modifications were necessary in order to use this library in calculating the distance between sequences of LOs. After these modifications, the prototype was able to use the FastDTW for calculating the distance between two sequences of LOs, treating them as time series. To achieve this, the data obtained in the Moodle Learning Environment were stored into a PostgreSQL database and the *jdbcSessionReader* class (see *Data Access* layer in Figure 1) was configured to access this database. The prototype receives a session identifier number, as a parameter, which is loaded from the database and simulates the current user session. The recommendation process is then activated, searching the other sessions present in the repository and recommending the LOs from the most similar sessions, which are those with the lowest distance value calculated via FastDTW in relation to the current session.

Figure 4 shows the result of the recommendation for the session number [1044725], linked to the user number [399]. This session has four access records for the three LOs identified by codes [3183], [2761] and [290], respectively. The *Recommendation* component analyzed 11.039 sessions, at a rate 400 to 500 sessions per second, in approximately 24 seconds. Session number [1043715] was the most similar in relation to the current session, with a distance value of 2.0, according to the DTW technique. Lastly, the three LOs numbered [2529], [1043] and [292] were recommended.

```
run:
------------------------------------------------------------------------------------------------------------
Original Session:
   UID.....: 1044725
   User ID.: 399
   IP......: 200.132.2.240
   Day.....: 06/08/2013
Session Sequence:
   1. [3183] Tanenbaum: Chapter 1 |90::5439|
   2. [2761] Pedagogical Project (Internet Systems) |1::4637|
   3. [290] lesson 1 - Basic concepts of databases |13:435|
   4. [290] lesson 1 - Basic concepts of databases |13:435|
Start recommendation process:
   Retrieving Sessions: 11039 sessions found! 735ms
       * Finished Block: Block [1000 sessions] [2sec 652ms] Total [2sec 652ms] Avg [500.5 sessions/sec]
       ...
   1. Similar Session: id=[1043715] distance=[2.0]
   2. Similar Session: id=[1044622] distance=[2.0]
   ...
   getSimilarSessions(4): 24sec 239ms
Recommended L.O.:
   1. [2529] Prototype Sample Output |42::4156| source=[session:1044676]
   2. [1043] lesson 2 - basic concepts of databases |13:1914| source=[session:808227]
   3. [292] lesson 2 - introduction to conceptual modeling |13:437| source=[session:808227]
------------------------------------------------------------------------------------------------------------
BUILD SUCCESS!
```

Figure 4: Prototype test result

By analyzing the first recommended object, with number [2529] (description "*Prototype Sample Output*") and subject area identified by "42::4156", it can be seen that it is not possible to identify similarities with any of the objects present in the current user session. However, when analyzing the session that gave rise to this recommendation – session number [1044676], the records of which are shown in Figure 5 – it can be seen that the sequence of LOs is similar to the objects of session [1044725] and we can deduce that the learning interests of the user during session [1044676] were similar to the interests of the user in the current session. Based on this reasoning, the prototype recommends the LOs present in the session [1044676] that have not yet been consulted in the current session.

Lastly, in order to evaluate scalability of the model, a test was performed based on the selection of the largest stored session with 154 access records. The number of sessions processed per second did not change significantly as a function of the number of LOs in the current session, always maintaining an average of approximately 435 sessions per second, even when using the largest sessions available in the database.

	sessionid bigint	userid bigint	datetime timestamp without time zone	course bigint	cmid bigint	resource_uid bigint	resource_name character varying(255)
1	1044676	1444	2013-08-06 22:30:12	42	4156	2529	Prototype Sample Output
2	1044676	1444	2013-08-06 22:51:15	13	435	290	lesson 1 - Basic concepts of databases
3	1044676	1444	2013-08-06 22:51:16	13	435	290	lesson 1 - Basic concepts of databases

Figure 5: Sequence of LOs accessed in the session number 1044676

4.3 Second Experiment

This experiment aims to evaluate the RecSim with standard metrics, such as precision and recall. In order to do so, we adopted as strategy the use of a synthetic database in which possible situations of a real classroom were represented. The advantages of this approach are to facilitate the generation of the data set according to the needs of the evaluation process and, at the same time, to allow a precise result evaluation, due to the control of the general data set distribution.

The experiment created 100 test cases corresponding to user sessions. A program written in Java language generated these user sessions automatically. The examples of materials in the sessions were obtained from a dictionary of learning objects taken from the *Merlot repository* (*http://www.merlot.org/merlot/categories.htm*). The methodology adopted in this experiment can be summarized in the following steps: (1) creation of a LOs dictionary; (2) automatic generation of the test cases; (3) experiments with the RecSim prototype and results evaluation. Details of these steps are described in the following paragraphs.

The **first step** involved the creation of a script to extract a list of LOs with their respective categories from the Merlot website. This script cycles through the HTML code of the webpages in the Merlot and extracts the name and category of LOs, saving a list of all learning objects found. The dictionary is saved in a text file named *merlot.dat*, which contains three fields separated by the character "|". These fields are the LO identification, its thematic area with the hierarchy levels and, finally, the LO title. An example of the obtained material is shown in Figure 6.

```
842024 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations|Bookshare: An Accessible Online Library
910494 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations|Autism e Disability App for Android
843326 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations|About KDE.org
843323 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations| accessibi-IT
727014 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations/ Accommodation Services | SpeechText Access LLC
1016046 |Academic Support Services/Accessibility/Accessibility Centers and
Organizations/ Accommodation Services | Test
```

Figure 6: Extract of LOs dictionary

The **second step** comprised the automatic test cases generation. These test cases are described in text files, with the same structure of the dictionary text file, but separated into five blocks (see an example in Figure 7).

The first block contains the expected result for the test, i.e., which LOs should be recommended. The second block contains a user session generated from the dictionary of LOs. This block represents the current session, i.e., the objects found by the user in a typical session and that match the current interest. There are three more blocks, with a different configuration regarding the similarity of materials compared to the user session in the second block. One of these represents a very similar session, the other a near similar session and the last a very different session.

The purpose of the last four blocks is to simulate a user access to a repository in a session and also to generate related sessions, with different degrees of similarity. In Figure 7, the most similar is called "Expected Similar Session". The block "Almost Similar Session" is the session that has some objects from the same area of the current session. Finally, the last block is called "Not Similar Session", representing the session that has no object similar to the objects of the current session.

```
expectedResult
555766 |Business/Economics/History of Ec. Thought|14.71 Economic History of Financial Crises
661609 |Business/Economics/History of Ec. Thought|American Economic History
90454 |Business/Economics/Law and Economics|Market Failures and Government Regulation
session 1|Current User: Business -> Economic Crises History
524254 |Business/Finance/Money and Banking|Time Value of Money
883922 |Business/Economics/Industrial Organization|15.015 Macro and International Economics
(MIT)
84634 |Business/Economics/International|Trade and Globalization
833342 |Business/Finance/Financial Institutions|Banking: An Introduction
76857 |Business/Finance/Financial Markets|The Stock Market Game Worldwide
86150 |Business/Finance/International|A Collection of Financial Keywords and Phrases
625534 |Business/Economics/Law and Economics|Optimal Regulation: The Economic Theory of
Natural Monopoly
session 2|Expected Similar Session User 1
76857 |Business/Finance/Financial Markets|The Stock Market Game Worldwide
86150 |Business/Finance/International|A Collection of Financial Keywords and Phrases
84634 |Business/Economics/International|Trade and Globalization
84634 |Business/Economics/International|Trade and Globalization
86150 |Business/Finance/International|A Collection of Financial Keywords and Phrases
833342 |Business/Finance/Financial Institutions|Banking: An Introduction
833342 |Business/Finance/Financial Institutions|Banking: An Introduction
883922 |Business/Economics/Industrial Organization|15.015 Macro and International Economics
(MIT)
625534 |Business/Economics/Law and Economics|Optimal Regulation: The Economic Theory of
Natural Monopoly
555766 |Business/Economics/History of Ec. Thought|14.71 Economic History of Financial Crises
session 3|Almost Similar Session User 2
667025 |Workforce Development/Construction|Carpentry video
79436 |Science and Technology/Information Technology/Web|Core Rules of Netiquette
822007 |Social Sciences/Sociology/Global|Planet Money Makes a T-Shirt
716035 |Humanities/History/Resources/Document Archives|Writing, and Researching for History:
A Guide for College Students by Patrick Rael
83948 |Science and Technology/Nanotechnology/Nano-electronics|CMOS Nanotechnology 101
635700 |Academic Support Services/EPortfolios/Teaching-centered ePortfolios/Health and Human
Sciences|L4a Spinal Cord Clinical Notes - Part 2
697654 |Social Sciences/Geography/Physical|Sustainability: A Comprehensive Foundation
524254 |Business/Finance/Money and Banking|Time Value of Money
883922 |Business/Economics/Industrial Organization|15.015 Macro and International Economics
(MIT)
84634 |Business/Economics/International|Trade and Globalization
833342 |Business/Finance/Financial Institutions|Banking: An Introduction
session 4|Not Similar Session User 3
622177 |Humanities/History/Theory and Method/Methods|Teaching history with technology
85079 |Humanities/World Languages/Latin/culture|Stanford Digital Forma Urbis Romae Project
85237 |Mathematics and Statistics/Mathematics/Mathematical Modeling/Discrete|The
Mathematical Art of M. C. Escher
883963 |Humanities/Women and Gender Studies|21A.470J Gender and Representation of Asian Women
(MIT)
90956 |Science and Technology/Biology/Botany|Virtual Foliage
836272 |Education/Educational Leadership/Leadership Specialization/Student Development and
Learning|Utilizing Effective Instructional Technologies
77225 |Science and Technology/Computer Science/Computer Simulation|D Flip-Flop Simulator
```

Figure 7: Example of test cases file

A specific algorithm was developed to generate the test cases. Firstly, the system generates four files called *currentSession*, where each one contains a coherent session description, that is, the objects accessed have no relation to each other and reflect the user's current research topic. These files refer to LOs described in the same dictionary structure and store the LOs list selected by the algorithm.

With these four sessions as a starting point, the system generates the test cases with the following steps. First, three LOs are selected to be recommended. Then, the remaining objects are shuffled to form the current session in the tests. After this, the algorithm generates 10 sessions with random objects. In these sessions, three are selected to have the similar sessions characteristics as exemplified in Figure 7. The first one is the similar session generated from the current session with objects from the list of selected objects. The second one is the near similar session, which is generated with some objects from the current session and with the random objects files. The number of these near similar sessions will be determined randomly and varies from one to seven sessions. The last one is the non-similar session generated from random objects from the file, which contains no objects in the current session. The algorithm follows until it has 10 sessions in total file, comprising a mixture of session with the three similar profiles. At the beginning of the file the expected learning objects are stored. A total of 100 files were generated with 25 for each one of the original four sessions.

In the **third step**, RecSIM was applied to test cases. Each set of 25 files generated was separated into folders, one for each *currentSession* generated. The RecSim then was configured to read the files in each folder, to apply the recommendation method for each file, and then to assess whether the recommended objects were those expected. In this experiment, the maximum number of LOs to be recommended in each test case was configured using parameters of 3, 6 and 9 objects. This was done, in order to allow a future evaluation regarding the objects generated from each *currentSession* obtained. Thus, to perform the precision and recall evaluation of the results, three batteries of tests were done, by switching the value of this parameter into 3, 6 and 9 objects to be recommended.

The prototype was adjusted to generate a worksheet (see example in Figure 8), showing the results of each test performed. The items expressed in the columns are the following. The column "*Test Case*" has the name of the test case file used. The "*Original User*" column expresses the name of the user and the thematic area in this test. The user name is generated randomly just for the test. The "*Number of LOs*" column exhibits the number of unique learning objects in the file. The column "*Relevant LO*" contains the number of LOs considered relevant and used in the recall measure. Details of precision and recall measures are described in the following paragraphs. The "*Recommended LO*" column indicates the number of recommended LOs. The column "*Hits*" contains the number of recommended LOs that fits in the criteria of expected recommended LOs. This criteria refers to the LOs present in the first part of the test case files (as exemplified in Figure 7), which are the correct choices for recommendation according to the automatic test cases generation procedure.

In order to evaluate the precision and recall measure, the RecSIM is applied on 4 sets of test cases files, setting the maximum number of objects to be recommended to 3, 6 and 9, which is the number of expected learning objects configured on each test case file. Then, the sessions described in each test case file will be assessed to confirm if some of the previously selected LOs (the LOs in the first block of the text case file) were recommended. Each of these is considered as a *True Positive* result (TP). If a leaning object that is not present in the selected object list was recommended, then this is considered a *False Positive* (FP). The precision measure is given by the equation "*Precision = TP/(TP+FP)*".

For the recall measure are considered LOs attributes, such as the knowledge field registered for each LO in the dictionary. The recall measure is calculated considering as relevant all those LOs that share the same knowledge field with at least one LO present in the current section. The value of recall is obtained considering how many relevant LOs were recommended, over the total number of relevant LOs available in test case set. In other words, the recall consists in the relation of the True Positive (TP) objects and the sum of the TP and False Negative (FN) objects, i.e., the recall measure is given by the equation "*Recall = TP/(TP+FN)*".

Test Case	Original User	Number of L.O.	Relevants L.O.	Recommended L.O.	Hits	Precision	Recall	Most Similar Session
test01.txt	Current User: Engineering/Computer Engineering	70	10	3	2	0,66667	0,2	Expected Similar Session User 3 (dist=5.0)
test02.txt	Current User: Engineering/Computer Engineering	75	8	3	3	1	0,375	Expected Similar Session User 1 (dist=4.0)
test03.txt	Current User: Engineering/Computer Engineering	83	8	3	3	1	0,375	Expected Similar Session User 2 (dist=5.0)
test04.txt	Current User: Engineering/Computer Engineering	74	8	3	3	1	0,375	Expected Similar Session User 3 (dist=4.0)
test05.txt	Current User: Engineering/Computer Engineering	63	6	3	3	1	0,5	Expected Similar Session User 1 (dist=5.0)
test06.txt	Current User: Engineering/Computer Engineering	73	8	3	2	0,66667	0,25	Expected Similar Session User 1 (dist=6.0)
test07.txt	Current User: Engineering/Computer Engineering	83	7	3	2	0,66667	0,28571	Expected Similar Session User 2 (dist=4.0)
test08.txt	Current User: Engineering/Computer Engineering	88	7	3	2	0,66667	0,28571	Expected Similar Session User 3 (dist=4.0)
test09.txt	Current User: Engineering/Computer Engineering	99	9	3	3	1	0,33333	Expected Similar Session User 1 (dist=6.0)
test10.txt	Current User: Engineering/Computer Engineering	66	8	3	3	1	0,375	Expected Similar Session User 1 (dist=5.0)
test11.txt	Current User: Engineering/Computer Engineering	69	8	3	3	1	0,375	Expected Similar Session User 2 (dist=3.0)
test12.txt	Current User: Engineering/Computer Engineering	73	6	3	3	1	0,5	Expected Similar Session User 1 (dist=6.0)
test13.txt	Current User: Engineering/Computer Engineering	82	8	3	2	0,66667	0,25	Expected Similar Session User 2 (dist=4.0)
test14.txt	Current User: Engineering/Computer Engineering	71	9	3	3	1	0,33333	Expected Similar Session User 1 (dist=5.0)
test15.txt	Current User: Engineering/Computer Engineering	79	8	3	2	0,66667	0,25	Expected Similar Session User 1 (dist=6.0)
test16.txt	Current User: Engineering/Computer Engineering	85	8	3	1	0,33333	0,125	Expected Similar Session User 3 (dist=5.0)
test17.txt	Current User: Engineering/Computer Engineering	84	8	3	3	1	0,375	Expected Similar Session User 2 (dist=5.0)
test18.txt	Current User: Engineering/Computer Engineering	68	7	3	2	0,66667	0,28571	Expected Similar Session User 3 (dist=6.0)
test19.txt	Current User: Engineering/Computer Engineering	85	9	3	2	0,66667	0,22222	Expected Similar Session User 2 (dist=5.0)
test20.txt	Current User: Engineering/Computer Engineering	68	9	3	2	0,66667	0,22222	Expected Similar Session User 1 (dist=5.0)
test21.txt	Current User: Engineering/Computer Engineering	75	7	3	2	0,66667	0,28571	Expected Similar Session User 1 (dist=5.0)
test22.txt	Current User: Engineering/Computer Engineering	79	7	3	3	1	0,42857	Expected Similar Session User 1 (dist=6.0)
test23.txt	Current User: Engineering/Computer Engineering	74	8	3	1	0,33333	0,125	Expected Similar Session User 3 (dist=4.0)
test24.txt	Current User: Engineering/Computer Engineering	70	7	3	3	1	0,42857	Expected Similar Session User 3 (dist=5.0)
test25.txt	Current User: Engineering/Computer Engineering	72	7	3	3	1	0,42857	Expected Similar Session User 3 (dist=4.0)
	Total	1908	195	75	61	20,33336	7,98965	
	Average	76,32	7,8	3	2,44	81,33%	31,96%	

Figure 8: Preliminary results

The Figure 9 shows the results summarized to highlight the precision and the recall together with the number of cases in which the maximum precision was obtained. The average precision of 85% and the average recall of 28,58% reflect the results for the situation in which the number of LOs parameter of the simulation is 3. This is considered an adequate value for precision, although the recall value can be considered low. In other steps of the experiment, when using the number of 6 or 9, it was observed a decrease in the precision value and an increase in the recall value. When using 6 LOs, the values obtained were 45,34% for precision and 39,26% for recall. When using 9 objects the values were 30,89% for precision and 40,11% for recall.

Finally, as a complementary step in this second experiment, some test cases were manually created and submitted to the RecSim prototype evaluation. The objective of this last step was to observe the outcome in situations that can occur in real life but cannot always be verified with purely random session generation. The tests were based on the following steps. First were created five testing files, simulating a sequence of a fictional set of sessions, created with objects taken from the Merlot repository. Each file corresponds to a specific situation to be tested, covering some unusual situations. Then the RecSim was applied in each file and were generated the recommendations. The situations covered scenarios where the test cases have similar and non-similar sessions, differences in the number of LOs in the sessions, several sessions with similar LOs and also a session with only one object. The expected outcome was defined manually and was confirmed in all results. The outcome of this additional assessment is relevant considering that the situations represent the worst-case scenarios and were not obtained in random generated files.

	Test Cases	Precision AVG	Recall AVG	Cases with Max Precision (100%)
Test Sequence 01	25	85,33%	32,17%	16
Test Sequence 02	25	89,33%	28,44%	17
Test Sequence 03	25	84,00%	26,85%	14
Test Sequence 04	25	81,33%	26,87%	13
Total	*100*	*85,00%*	*28,58%*	*60*

Figure 9: Summary of results obtained

5 Related Works

In this section, related works are organized into two groups. Section 5.1 approaches proposals for similarity analysis of context sequences. In turn, section 5.2 addresses works involving learning objects recommendation. Finally, section 5.3 discusses the contribution of RecSim through a comparison with related works.

5.1 Context Sequence Similarity

According to Lv et al. [2013], similarity analysis of context sequences is a crucial factor for the new generation of location-based social networks (LBSN). However, several studies focus on analyzing sequences of geographic locations. The limitation of this approach is that physical trajectories generally correspond to short periods of time, thereby making a long-term similarity analysis impracticable. To resolve this issue, Lv et al. [2013] propose a two-stage approach. In the first stage, the notion of *Routine Activities* is proposed to record the regularity of a user's activities in the long term. In the second stage, similarity is calculated hierarchically, based on the extracted routine activities. The user location preference matrix is then calculated. The value in each cell corresponds to the number of minutes the user remained in the location during the time interval. The *Routine Activity* is a probability distribution matrix, where cell values correspond to the probability of the user being at this location, in this time period. Two *Routine Activity* matrices are considered similar when they both have a high probability of visiting *Reference Locations* in the same time intervals.

Zeinalipour-Yazti et al. [2013] propose a framework called *SmartTrace*$^{+}$ for comparing a trajectory *Q* with the trajectories generated and stored by a multitude of smartphones, returning the set of most similar trajectories. The framework architecture implements a hybrid approach, using a server to manage the smartphones connected to the system and to distribute query requests to all client phones in order to analyze, locally and opportunistically, the similarity between the trajectory that is consulted and the device's own stored trajectory.

Li et al. [2012] propose a framework for on-line identification of groups (clusters) of mobile objects that are traveling together. The authors consider the real trajectory of a mobile object as a continuous function of the temporal domain $\mathbb{T}$ for points in n-dimensional Euclidean space $\mathbb{R}^n$, where n = 2 for most applications. Starting from this definition, the trajectory of an object can be collected by sampling positions over time, in accordance with a particular policy, resulting in a set of points $(t, \vec{r}) \in \mathbb{T}\ X\ \mathbb{R}^n$.

In Abraham and Lal [2012] and Abraham and Lal [2011], two frameworks are proposed for a similarity analysis in sequences of trajectories. The first, known as *TraSimilar*, forms clusters of vehicles with similar trajectories, considering the POI (*Point-Of-Interest*) and TOI (*Time-Of-Interest*) history of each vehicle. The second, *WebTraSim*, applies the trajectory concept to the sequence of web pages visited by a user, thereby creating trajectory similarity metrics that consider the UOI (*URL-Of-Interest*) and TOI (*Time-Of-Interest*) history of each user.

According to Cui et al. [2012], *Time Series Cliques* (TSC) consist of multiple time series naturally related to one another. For example, each player's movements in a football match can be described by a time series. The set of time series that represent the movements of all players in the same game can be seen as TCS. The authors developed a framework for similarity searching in TSC data.

Obweger et al. [2010] propose a model for analyzing the similarity in sequences of events. According to the authors, the method for *Complex Event Processing* (*CEP*) [Luckham, 2001], allows companies to monitor incidents related to their operations in real time and to implement automated decisions for reacting to potential threats to their business. To illustrate application of this model, Obweger et al. [2010] use the example of an online betting company where, after detecting a new method of fraud, analysts can search the historical data for old transactions in which this method was applied.

Niemann et al. [2010] introduce a strategy of similarity calculation for *Objects* and its use to support recommendation based on *Contexts*. Objects are generic recommendation items organized in sequences and the context of an object is formed by objects that have been used before (pre-context) and after (pos-context) during an access session. The usage-based object similarity is calculated comparing the usage context profile (UCP) of objects. This profile is the set of object's different usage contexts, being each usage context the sequence of objects used before and after the considered object in a specific session.

Although the proposal of Niemann et al. [2010] is generic, the experiments were based on sequences of real LOs. These experiments focused on two issues: (1) to prove the initial claim that similarity of usage indicates content similarity; (2) to show how the strategy can support educational content recommendation. The authors describe three methods to support the recommendation of LOs based on usage context profiles. The second one uses the history of LOs access in a session (pre-context) to recommend objects that have highly similar pre-contexts. Niemann et al. [2010] is the most related work found during the literature review. Despite the relevance of LOs to the proposal, we have considered the work as a generic approach of context sequence similarity and it was included in this section.

5.2 Learning Object Recommendation

Research works about recommendation of educational material are becoming increasingly important due to the increasing adoption of e-learning [Cambruzzi et al., 2015]. The Web already provides a huge quantity of materials that can be useful for educational purposes. In this scenario, teachers not only need to examine whether this vast quantity of materials available falls into their pedagogical strategies but, ideally, also check if they comply with the learning profiles of students [Akbulut and Cardak, 2012; Peterson et al., 2009; Felder and Silverman, 1988] and to the teaching context where the learning is occurring [Dey, 2001]. This indicates a need of computational tools to help teachers (and students) to find adequate digital learning content. In this sense, several strategies to recommend learning objects have been developed.

Cazella et al. [2010] propose a model that filters the LOs in a repository according to the skills to be acquired, which are indicated in the user profile. In this model, the teacher selects the objects identifying the list of skills that each LO helps to develop. The students then perform a self-assessment based on the skills identified by the teacher. Lastly, the system suggests LOs based on the skills of each student.

Rocha et al. [2010] developed an LO recommendation module integrated into the AMADEUS virtual environment. The recommendation algorithm represents the LO metadata in vector form, where the terms present in each description correspond to the vector dimensions, and the frequency of terms corresponds to the value of each dimension. The K-nearest Neighbors technique is applied to filter the objects.

Zapata et al. [2013] present a framework called DELPHOS that assists users in performing custom queries in repositories. The framework adopts a hybrid approach, combining *Collaborative approaches* (analyzing the history of the most frequently accessed LOs and the evaluation of users), *Content-based approaches* (calculating the degree of similarity between two LOs) and *Demographic approaches* (calculating the similarity between users).

As a basis for similarity analysis, these three related works [Cazella et al., 2010; Rocha et al., 2010; Zapata et al., 2013] use the full set of object characteristics or the full history of objects accessed by the user. This technique can generate recommendations that do not meet users' needs, when considering their current interests.

Some works propose a context-aware LO recommendation approach. Barbosa et al. [2014] propose a decentralized infrastructure for the development of ubiquitous learning environments, called Global. Global supports an infrastructure to recommend LOs using learners profiles [Wagner et al., 2014] and contexts [Dey, 2001]. LOs recommendation can be implemented by extension of agents or by adding new ones, as demonstrated in Barbosa et al. [2014]. GlobalEdu [Barbosa et al, 2013] and Local [Barbosa et al., 2011] also suggest models to support the ubiquitous learning. Both proposals explore learning opportunities considering learners profiles and context information. The first focuses on large-scale environments and the second is restricted to small-scale environments. Local [Barbosa et al., 2011] was integrated with a trail management system to content distribution using context history information [Silva et al, 2010].

Other research works on LOs recommendation explore *Ontologies* [Rodrígues et al., 2014]. Ontology can be defined as a formal and explicit definition of the conceptual categories existing in some knowledge domain [Gruber, 1993] or as a conceptualization characterized by formal (explicit) properties and specific purposes [Gruber, 2005]. Ontologies are useful to bring semantics to educational contents [Bittencourt et al., 2009] enriching the LOs specification. Based on ontologies, *Software Agents* [Wooldridge and Jennings, 1995; Huhns and Stephens, 1999] can implement intelligent strategies to make decisions and recommendations [Bobadilla et al., 2013; Sokolova and Caballero, 2012].

Han et al. [2010] propose to recommend LOs adapted to the learner needs. An ontology represents the structure of LOs and semantic relationships and concepts. The learner difficulties are combined with concepts and through inference rules, prerequisites are evaluated and the recommendation is implemented. Romero and Godoy [2010] recommend an ontological structure based on standard LOM [2002]. The proposal introduces a conceptual framework for metadata that focuses on the relationships between LOs. Software agents interpret the learner needs and find LOs to be recommended. Yarandi et al. [2012] introduce an ontology for classifying learning materials and user profiles, focusing on providing an adaptive learning environment. The suggested architecture has the *Course Recommendation Mediator*. This module is responsible for selecting LOs according to pedagogical rules and learners' profiles. Educational materials are annotated using terms extracted from two ontologies used to support the learner profile and content information, providing content adaptation. Abech et al. [2016] propose EduAdapt, a model for context-aware LOs adaptation. The model is based on inferences and rules of an ontology. EduAdapt considers learners' context information, such as devices characteristics and learning styles to recommend adapted learning objects.

5.3 Contribution of RecSim

The works described in section 5.1 state different techniques for similarity analysis of context histories in different fields of application. Our approach differs from most of them in its specific application to education.

The proposal of Niemann et al. [2010] is generic, but the evaluation is focused on education, more specifically, on the recommendation of learning objects. RecSim uses a technique similar to that applied in the models proposed by Abraham and Lal [2011] and Abraham and Lal [2012]. Unlike that suggested by Niemann et al. [2010] in their second method to recommend LOs, RecSim uses the DTW similarity function [Berndt and Clifford, 1994] to compare histories of sessions as time series.

In turn, section 5.2 presented specific works in the field of education that recommend learning objects. None of them uses similarity analysis based on DTW technique to guide the recommendation.

Therefore, the scientific contribution of RecSim is the use of a time series similarity analysis technique (specifically, DTW [Berndt and Clifford, 1994]) to perform real-time LOs recommendation during an access session in a content repository. This strategy seeks to consider the current learner interest represented by the choices done during the session.

6 Conclusions and Future Studies

This article presents a model for LOs recommendation based on similarity between the sequences of accesses performed in each user session. RecSim uses a technique known as *Dynamic Time Warping* [Berndt and Clifford, 1994]) to calculate similarity between the sequence of objects consulted during a user's current session and the set of sessions stored in the access logs of a repository. A function was developed to calculate the distance between LOs, inspired by the work of Abraham and Lai [2012].

A prototype was developed and applied in two experiments to evaluate the proposed model. The first experiment evaluated the functionalities aspects of RecSim, demonstrating that the DTW technique can be used to identify sequences of similar LOs. This identification of similarity allows the recommendation of potentially useful LOs to the student, without the need for an explicit record of the relationship between these objects in the repository. The experiment has also shown that the sessions processing rate did not change significantly, depending on the number of LOs in the current session under analysis.

The second experiment demonstrated that RecSim precision presents adequate levels, similar to those observed in the literature described in section 5.2. Moreover, in this experiment it was evidenced that manually created examples of worst-case scenarios were treated correctly by the model. The experiment was conducted with a synthetic database of interactions generated to simulate typical real sessions, suggesting that results in real scenarios can reach user expectations.

Future studies can investigate the use of other techniques of similarity analysis in sequences, in order to compare them against the DTW technique used in this article, as well as making improvements to the calculation function proposed for the distance between objects. In addition, the identification of similar sessions may assist in other applications, such as, the automatic formation of study groups according to the degree of similarity between students or the construction of student profiles according to their history of interests.

Furthermore, as a similarity analysis between two series is a statistical measure, two aspects may be studied and evaluated through experiments. Firstly, the influence of the sample length in the analysis and consequently in the recommendation may be considered. This research effort can be guided by previous studies related to the prediction error functions in time series [Li et al., 2015]. Second, the similarity threshold used to guide the recommendation may be better treated by future studies using methods to adjust loading, as the approach discussed in Li [2005].

The usefulness of a recommended LO is not only related to the relevance of the content, but it is also important that the content has an appropriate level of difficulty for the learner, that is, not too hard and not too easy. This adjusting can be done matching the *Learner Profile* [Wagner et al., 2014] with the LOs' characteristics registered in their metadata.

We researched the adaptation of content recommendation using learners' context histories (also called Trails [Silva et al., 2010]) and also the profile management of learners [Wagner et al., 2014]. In addition, we recently explored the use of *Learning Style* [Akbulut and Cardak, 2012] to adapt the LOs recommendation [Abech et al., 2016]. However, none of these works include a specific discussion on matching the difficulty level of LOs with the ability of learners. RecSim also does not address this issue, because the focus of the proposal is to explore the similarity of sessions. Future work will integrate RecSim with our previous researches on profiles and learning styles. In this sense, LOs adaptation in the recommendation [Abech et al., 2016] will be prioritized, especially considering the adequate knowledge level of learners.

Further evaluations could be performed to better measure efficacy of the proposed model. The models described in related works could be applied to the same data used in the evaluation of RecSim, thereby providing a comparison of efficacy between models. Another aspect to be measured in future evaluations is the impact of the length of compared sequences of LOs to the efficiency of the recommendation.

The technique used in the RecSim for similarity analyses (DTW [Berndt and Clifford, 1994]) does not set a minimum length of sequences to be effective. However, if there is a significant difference between the lengths of sequences (for example, the current session has three LOs and the sequence in the history has thirty LOs), the DTW always will indicate a large distance between them. In these cases, it would be more efficient to find within the larger sequence, a subsequence that best meets the smaller sequence, instead of comparing the two complete sequences. This strategy is explored in many applications as discussed by Meinard Muler in the fourth chapter of his book [Muler, 2007].

RecSim assumes that LOs in a specific session have semantic relationship, because the learner consulted them in a sequence during a specific study. Thus, Recsim does not calculate the difference between the LOs in the session. However, this calculation may be an improvement in the model. It could indicate the degree of confidence in a given sequence stored in the repository's sessions history, based on the value of "*internal similarity*" between the session LOs. For example, a session that the user consulted the LOs in a coherent sequence would have a greater degree of *internal similarity* than one session that the student was just accessing LOs randomly. If these values are calculated and stored for each session, then the process of recommendation could use them as indexes, first comparing the current session with the most reliable sessions and even discarding those sessions that have not reached a minimum level of *internal similarity* (threshold). Future studies will consider potential improvements to the recommendation using the *internal similarity* of LOs in a session.

The main focus of our work is the use of the DTW technique as a tool for determining similarity between sessions. The calculation of the distance between LOs is used to determine each grid point within the warping window, as discussed in section 3.2. The accuracy of this calculation is essential for the quality of the final result.

Currently, RecSim applies a simple solution, i.e., a comparison of subject areas as done in Abrahm and Lai [2012]. We assume that LOs metadata are filled in a simple pattern (subject areas) and that there is no problem with ambiguity (LOs are described with a unique subject area). These two assumptions are basic to the equation 2. Thus, if they do not happen, it would not be possible to complete the grid. We adopted this simple solution because the distance calculation was not our main research concern.

We are aware that these limitations are strong obstacles to a real application of the RecSim. However, the method use for this calculation can be changed without compromising the DTW use to calculate the similarity of sessions. Future studies can explore the use of semantic comparison of metadata described in ontologies or comparing files content. RecSim only needs that the applied method returns a consistent value to be placed at each grid point.

Finally, the model could be implemented as an experimental module in an application and then subjected to evaluation by a group of students in a distance-learning environment. In this sense, the work could be enriched with an evaluation of the technology acceptance. Upon execution, the users would answer an assessment questionnaire based on the *Technology Acceptance Model* (TAM). This model measures the satisfaction through perceived usefulness and perceived ease of use.

The TAM model has been considered a standard to evaluate the acceptance of new technologies [Marangunić and Granić, 2013]. Furthermore, a long-term experiment with students would support a more complete assessment of the effectiveness of the model and would bear further studies based on user behavior, such as the influence of the first LO to guide the recommended content access path.

Acknowledgements

This work was financed by Capes/Brazil (http://www.capes.gov.br) and CNPq/Brazil (http://www.cnpq.br). We would like to acknowledge the support of both institutions. In addition, we would like to thank Unisinos (http://www.unisinos.br) and Feevale (http://www.feevale.br) for embracing this research. Finally, we would like to express our sincere thanks to the reviewers for their valuable contributions to the final quality of this article.

References

[Abech et al., 2016] Abech, M., Costa, C.A., Barbosa, J.L.V., Rigo, S.J., Righi,R.R.: A model for learning objects adaptation in light of mobile and context-aware computing; Personal and Ubiquitous Computing, 20, 2, 167-184, 2016. DOI= 10.1007/s00779-016-0902-3.

[Abraham and Lal, 2011] Abraham, S., Lal, P. S.: Spatio-temporal similarity of web user session trajectories and applications in dark web research; in Proc. 6th Pacific Asia conference on Intelligence and security informatics (PAISI), 2011, 1-14. DOI=10.1007/978-3-642-22039-5_1.

[Abraham and Lal, 2012] Abraham, S., Lal, P. S.: Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations; Transportation Research Part C: Emerging Technologies, 23, 109-123, 2012. DOI= 10.1016/j.trc.2011.12.008.

[Akbulut and Cardak, 2012] Akbulut, Y., Cardak, C. S.: Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011; Computers & Education, 58, 2, 835-842, 2012. DOI= 10.1016/j.compedu.2011.10.008.

[Al-Khalifa, 2008] Al-Khalifa, H. S.: Building an Arabic learning object repository with an ad hoc recommendation engine; in Proc 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS), 2008, 390-394. DOI= 10.1145/1497308.1497378.

[Barbosa et al., 2011] Barbosa, J. L. V., Hahn, R., Barbosa, D. N. F., Saccol, A.: A Ubiquitous Learning Model Focused on Learner Integration; International Journal of Learning Technology, 6, 1, 62-83, 2011. DOI=10.1504/IJLT.2011.040150.

[Barbosa et al., 2013] Barbosa, D. N. F., Barbosa, J. L. V., Bassani, P. B. S., Rosa, J., Martins, M., Nino, C.: Content management in a ubiquitous learning environment; International Journal of Computer Applications in Technology, 46, 1, 24-35, 2013. DOI=10.1504/IJCAT.2013.051385.

[Barbosa et al., 2014] Barbosa, J. L. V., Barbosa, D. N. F., Oliveira, J. M., Rabello, S. A.: A Decentralized Infrastructure for Ubiquitous Learning Environments; Journal of Universal Computer Science, 20, 12, 1649-1669, 2014. DOI=10.3217/jucs-020-12-1649.

[Bellavista et al., 2012] Bellavista, P., Corradi, A., Fanelli, M., Foschini, L.: A Survey of Context Data Distribution for Mobile Ubiquitous Systems; ACM Computing Surveys, 44, 4, 24-45, 2012. DOI=10.1145/2333112.2333119.

[Berndt and Clifford, 1994] Berndt, D., Clifford, J.: Using dynamic time warping to find patterns in time series. in Proc AAAI Workshop on Knowledge Discovery in Databases, 1994, 359-370.

[Bittencourt et al., 2009] Bittencourt, I., Costa, E., Silva, M., Soares, E.: A computational model for developing semantic web-based educational systems; Knowledge-Based Systems, 22, 4, 302–315, 2009. DOI=10.1016/j.knosys.2009.02.012.

[Bobadilla et al., 2013] Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey, Knowledge-Based Systems, 46, 109-132, 2013. DOI=

10.1016/j.knosys.2013.03.012.

[Brown et al., 1997] Brown, P. J., Bovey, J. D., Chen, X.: Context-aware applications: from the laboratory to the market-place; IEEE Personal Communications, 4, 5, 58-64, 1997. DOI= 10.1109/98.626984.

[Cambruzzi et al., 2015] Cambruzzi, W., Rigo, S. J., Barbosa, J. L. V.: Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach, Journal of Universal Computer Science, 21, 1, 23–47, 2015. DOI=10.3217/jucs-020-12-1649.

[Cazella et al., 2010] Cazella, S. C., Reategui, E. B., Behar, P.: Recommendation of Learning Objects Applying Collaborative Filtering and Competencies; in Proc. IFIP Advances in Information and Communication Technology, 2010, 35-43. DOI=
10.1007/978-3-642-15378-5_4.

[Cha, 2007] Cha, S. H.: Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions; International Journal of Mathematical Models and Methods in Applied Sciences, 1, 4, 300-307, 2007.

[Cui et al., 2012] Cui, B., Zhao, Z., Tok, W. H.: A Framework for Similarity Search of Time Series Cliques with Natural Relations; IEEE Transactions on Knowledge and Data Engineering, 24, 3, 385-398, 2012. DOI= 10.1109/TKDE.2010.270.

[Dey, 2001] Dey, A. K.: Understanding and Using Context; Personal and Ubiquitous Computing, 5, 1, 4-7, 2001. DOI=10.1007/s007790170019.

[Dey et al., 2001] Dey, A. K., Abowd, G. D., Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications; Human-Computer Interaction, 16, 2, 97-166, 2001. DOI=10.1207/S15327051HCI16234_02.

[Clarke and Driver, 2004] Clarke, S, Driver, C.: Context-aware trails [mobile computing; Computer, 37, 8 97-99, 2004. DOI=10.1109/MC.2004.89.

[Driver and Clarke, 2008] Driver, C., Clarke, S.: An application framework for mobile, context-aware trails; Pervasive and Mobile Computing, 4, 5, 719-736, 2008.

DOI=10.1016/j.pmcj.2008.04.009.

[Esling and Agon, 2012] Esling, P., Agon C.: Time-series data mining; ACM Computing Surveys, 45, 1, 1-12, 2012. DOI=10.1145/2379776.2379788.

[Felder and Silverman, 1988] Felder, R. M., Silverman, L. K.: Learning and teaching styles in engineering education; Engineering education, 78, 7, 674-681, 1988

[Fu, 2011] Fu, T. C.: A review on time series data mining; Engineering Applications of Artificial Intelligence, 24, 1, 164-181, 2011. DOI= 10.1016/j.engappai.2010.09.007.

[Ghosh and Dekhil, 2009] Ghosh, R., Dekhil, M.: Discovering user profiles; In Proc. 18th international conference on World wide web (WWW), 2009, 1233-1234.

[Gruber, 1993] Gruber, T. R.: A translation approach to portable ontology specifications; Knowledge Acquisition, 5, 2, 199–220, 1993. DOI=10.1006/knac.1993.1008.

[Gruber, 1995] Gruber, T. R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43, 5-6, 907–928, 1995. DOI= 10.1006/ijhc.1995.1081.

[Han et al., 2010] Han, Q., Gao, F., Wang, H.: Ontology-based learning object recommendation for cognitive considerations; in Proc. World Congress on Intelligent Control and Automation, 2010, 2746–2750. DOI=10.1109/WCICA.2010.5554857.

[Hong et al., 2009] Hong, J., Suh, E., Kim, J., Kim, S.: Context-aware system for proactive personalized service based on context history; Expert Systems with Applications, 36, 4, 7448–7457, 2009. DOI=10.1016/j.eswa.2008.09.002.

[Huhns and Stephens, 1999] Huhns, M., Stephens, L.: Multiagent Systems and Societies of Agents; in Weiss, G. (eds), Multiagent systems: a modern approach to distributed artificial intelligence, MIT Press, Cambridge, 1999.

[Knappmeyer et al., 2013] Knappmeyer, M., Kiani, S.L., Reetz, E.S., Baker, N., Tonjes, R.: Survey of Context Provisioning Middleware; IEEE Communications Surveys & Tutorials, 15, 3, 1492-1519, 2013. DOI=10.1109/SURV.2013.010413.00207.

[ILOD, 2015] International Bank of Educational Objects (Brazilian Education Ministry), 2015, http://objetoseducacionais2.mec.gov.br.

[Li, 2005] Li, M.: An iteration method to adjusting random loading for a laboratory fatigue test; International Journal of Fatigue, 27, 7, 783-789, 2005. DOI=10.1016/j.ijfatigue.2005.01.011.

[Li et al., 2012] Li, X., Ceikute, V., Jensen, C., Tan, K.: Effective Online Group Discovery in Trajectory Databases; IEEE Transactions on Knowledge and Data Engineering, 25, 12, 2752-2766, 2012. DOI=10.1109/TKDE.2012.193.

[Li et al., 2015] Li, M., Li, Y., Leng, J. Power-type Functions of Prediction Error of Sea Level Time Series; Entropy, 17, 7, 4809-4837, 2015. DOI=10.3390/e17074809.

[LOM, 2002] 1484.12.1-2002 - IEEE Standard for Learning Object Metadata, 2002. DOI=10.1109/IEEESTD.2002.94128.

[Luckham, 2001] Luckham, D. C.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems; Addison-Wesley, Boston, 2001.

[Lv et al., 2013] Lv, M., Chen, L., Chen G.: Mining user similarity based on routine activities; Information Sciences, 236, 17-32, 2013. DOI=10.1016/j.ins.2013.02.050.

[Marangunić and Granić, 2014] Marangunić, N., Granić, A.: Technology acceptance model: a literature review from 1986 to 2013; Universal Access in the Information Society, 14, 1, 81-95,, 2014. DOI=10.1007/s10209-014-0348-1.

[Matuschek et al., 2008] Matuschek, M., Schlüter, T., Conrad. S.: Measuring text similarity with dynamic time warping; in Proc. International Symposium on Database Engineering and Applications (IDEAS), 2008, 263-267. DOI=10.1145/1451940.1451977.

[Moore et al., 2010] Moore, P., Hu, B., Jackson, M., Wan, J.: Intelligent context for personalized mobile learning; in Caballe, S., Xhafa, F., Daradoumis, T., Juan, A. A. (eds), Architectures for

Distributed and Complex M-Learning Systems: Applying Intelligent Technologies, IGI Global, Hersey, 2010. DOI=10.4018/978-1-60566-882-6.ch012.

[Muler, 2007] Muler, M.: Dynamic Time Warping; in Muler, M. (ed), Information Retrieval for Music and Motion, Springer-Verlag, Berlin, 2007. DOI=10.1007/978-3-540-74048-3_4.

[Niemann et al., 2010] Niemann, K., Scheffel, M., Friedrich, M., Kirshenmann, U., Schmitz, H.-C., Wolpers, M.: Usage-based Object Similarity, Journal of Universal Computer Science, 16, 6, 2272–2290, 2010. DOI=10.3217/jucs-016-16-2272.

[Obweger et al., 2010] Obweger, H., Suntinger, M., Schiefer, J., Raidi, G.: Similarity searching in sequences of complex events; in Proc. Fourth International Conference on Research Challenges in Information Science (RCIS), 2010, 631-640. DOI= 10.1109/RCIS.2010.5507284.

[Polsani, 2003] Polsani, P.R.: Use and Abuse of Reusable Learning Objects; Journal of Digital Information, 4, 3, 1-10, 2003.

[Peterson et al., 2009] Peterson, E. R., Rayner, S. G., Armstrong, S. J.: Researching the psychology of cognitive style and learning style: Is there really a future?; Learning and Individual Differences, 19, 4, 518-523, 2009. DOI=10.1016/j.lindif.2009.06.003.

[Resnick and Varian, 1997] Resnick, P., Varian, H. R.: Recommender systems; Communications of the ACM, 40, 3, 56-58, 1997. DOI=10.1145/245108.245121.

[Rocha et al., 2010] Rocha, R., Franco, R., Barros, R. S., Dantas, E. R. G., Azevedo, R. R., Dias, G. A., Gomes, A. S.: Amadeus recommends: a recommendation system for learning objects; In Proc. 16th Workshop of Informatics in School (WIE-SBC), 2010, 1457-1460.

[Rodrígues et al., 2014] Rodríguez, N. D., Cuéllar, M. P., Lilius, J., Calvo-Flores, M. D.: A survey on ontologies for human behavior recognition; ACM Computing Surveys, 46, 4, 1-33, 2014. DOI=10.1145/2523819.

[Romero and Godoy, 2010] Romero, L., Godoy, R.: An ontology for semantic definition of learning objects; in Proc. 5th Iberian Conference on Information Systems and Technologies (CISTI), 2010, 1–7.

[Schilit and Theimer, 1994] Schilit, B. N., Theimer, M. M.: Disseminating active map information to mobile hosts; IEEE Network, 8, 5, 22-32, 1994.DOI= 10.1109/65.313011.

[Silva et al., 2010] Silva, J. M., Rosa, J. H., Barbosa, J. L. V., Barbosa, D. N. F., Palazzo, L. M.: Content Distribution in Trail-aware Environments; Journal of the Brazilian Computer Society, 16, 3, 163–176, 2010. DOI=10.1007/s13173-010-0015-1.

[Sokolova and Caballero, 2012] Sokolova, M., Caballero, A. F.: A Review on Frameworks for Decision Support Systems. in Sokolova, M. and Caballero, A. F. (eds), Decision Making in Complex Systems, Springer-Verlag, Berlin, 2012. DOI=10.1007/978-3-642-25544-1_2.

[Spence et al., 2005] Spence, M., Driver, C., Clarke S.: Sharing Context History in Mobile, Context-Aware Trails-Based Applications; In Proc. International Workshop on Exploiting Context Histories in Smart Environments (ECHISE), 2005, 1-5.

[Wagner et al., 2014] Wagner, A., Barbosa, J.L.V., Barbosa, D.N.F.: A model for profile management applied to ubiquitous learning environments; Expert Systems with Applications, 41, 4, 2023-2034, 2014. DOI=10.1016/j.eswa.2013.08.098.

[Weng et al., 2009] Weng, S., Lin, B., Chen, W.: Using contextual information and multidimensional approach for recommendation; Expert Systems with applications, 36, 2, 1268–1279, 2009. DOI=10.1016/j.eswa.2007.11.056.

[Wooldridge and Jennings, 1995] Wooldridge, M., Jennings, N.: Intelligent agents: Theory and practice; Knowledge Engineering Review, 10, 115-152, 1995.

[Zeinalipour-Yazti et al., 2013] Zeinalipour-Yazti, D., Laoudias, C., Costa, C., Vlachos, M., Andreou, M. I., Gunopulos, D.: Crowdsourced Trace Similarity with Smartphones; IEEE Transactions on Knowledge and Data Engineering, 25, 6, 1240-1253, 2013. DOI= 10.1109/TKDE.2012.55.

[Yarandi et al., 2012] Yarandi, M., Jahankhani, H., Tawil, A.: An ontology-based adaptive mobile learning system based on learners' abilities; in Proc. IEEE Global Engineering Education Conference (EDUCON), 2012, 1–3. DOI=10.1109/EDUCON.2012.6201115.

[Zapata et al., 2013] Zapata, A., Menéndez, V., Prieto, M., Romero. C: A framework for recommendation in learning object repositories: An example of application in civil engineering; Advances in Engineering Software, 56, 1-14, 2013. DOI= 10.1016/j.advengsoft.2012.10.005.