AHSPeR: Adaptive Hypermedia System oriented toward Personalization of Readings plans

M. Giugni, F. Grimón, J. Fernández, J. Monguet, and A. Guerra

Abstract—The phenomenon of Internet has brought a wide range of communication possibilities and thus a rapid growth of digitized information. Each day the user is overwhelmed with the vast information obtained during the search process, which can hardly identify those that have greater relevance to their information need. With the aim of facing this problem, an authoring tool of adaptive hypermedia system, oriented toward the personalization of readings plans in a learning environment, was developed with a view to assessing its effectiveness and the user’s satisfaction vis-a-vis the proposed adaptation algorithm. This application is based on clustering algorithms and adaptation rules to adjust the user’s model to contents and objects of study. Research is conducted according to the action-research methodology. The system was accepted by 89.5% of the sample assessment using the Technology Acceptance Model TAM.

Keywords—Adaptation algorithm, adaptive hypermedia system, clustering algorithms, personalization.

I. INTRODUCTION

According to [1] Hypermedia systems are widely used as tools to enable users to find information. The characteristics of users have an important role in the Adaptive Hypermedia Learning System (AHLSSs), among them are: preferences, interests, goals and prior knowledge and others [2].

As express in [3] the Adaptive Hypermedia Systems (AHS) are systems that have a user model and a concept model to permit the recovery of personalized content to each user. AHS model is composed of: data collection (content), users and adaptation [4].

In this way, personalization constitutes the mechanisms necessary to automatically customize information content, structure, and presentation to the end-user to reduce information overload [5], [6].

Clustering is a data mining technique, also called segmentation or unsupervised learning, useful for discovering interesting data distributions and patterns in the underlying data. It is a process of grouping physical or abstract objects into classes of similar objects [7]. The essential motivation for using these techniques is that they have demonstrated to work really well for modeling preferences of users [8].

The purpose of this research was to develop an authoring tool, call AHSPeR ( Adaptive Hypermedia System oriented toward Personalization of Readings plans), which provides the user desired information, minimizing the search effort and the amount of irrelevant information. This system categorizes information from the Web, restricts the search space and modifies the user profile, according to their interaction with the system. The professor responsible of the course takes advantage from his/her expertise in teaching and authoring courses while deciding the personalization parameters adapted to his/her course.

AHSPER has been assessed in a scientific community of the computer sciences area at higher education level. In addition, one advantage of this development is the possibility of working with sets of unlabeled documents, i.e. without knowing the correlation between documents and categories to which they belong, a characteristic that makes it valid for any domain.

The paper is organized as follows. Section 2 presents related work. Section 3 and 4 describes the system and clustering algorithm respectively. The system evaluation by the users is in Section 5 and user’s satisfaction degree is in Section 6. Finally, in Section 7 concludes the paper.

II. RELATED WORK

According to [9] both Intelligent Tutoring Systems (ITS) and Adaptive Hypermedia systems (AHS) are normally used for computer-based instruction. However, adaptive hypermedia is better suited for the instruction of concepts whereas intelligent tutoring system generally assists in the use of these concepts to solve problems.

In e-learning, clustering has been used for: finding clusters of students with similar learning characteristics and to promote group-based collaborative learning [10]; discovering patterns reflecting user behavior and for collaboration management to characterize similar behavior groups [11]; grouping students in order to give them differentiated guiding.
according to their skills [12]. In general, clustering analysis is useful to find natural boundaries in datasets and classify objects into subsets with similar patterns in some properties. The meaningful characteristic of cluster hidden in the dataset is useful for analysis.

We have developed an educational adaptive hypermedia system to make it easier students choose where to go from a certain point, i.e. to recommend content based on your profile. Our clustering approach uses information about the contents of documents and about the students (knowledge, interests and log reading), to create custom reading plans. AHSPeR is generic and domain independent, can be customized for particular domains. Initial setup of the system is shown in Fig. 1.

![Fig. 1 Initial setup of AHSPeR (domain independent)](image)

III. SYSTEM DESCRIPTION

AHSPeR initially stores students’ static information, which is represented in their personal data and research areas. Then, upon an initial assessment, the application identifies the user’s information needs relative to the subjects provided by the system; this information is used to place the student according to the initial knowledge level, which can be beginner, intermediate, advanced or expert. Many studies have found that learners with different levels of prior knowledge benefit differently in hypermedia learning systems, with experts and novices showing different preferences to the use of hypermedia learning systems [13].

User profile describes the student's interests and goals. Initial profile, is generated by the type of assessment diagnostic evaluation, and can be configured according to the domain where the system is used. We have used the simple selection method.

The system can generate the diagnostic evaluation, either by providing a test designed to show the level of user interest in a particular subject, or a simple screening test that identifies the level of student knowledge. The scale of assessment can be configured by the tutor.

As far as contents are concerned, the system groups documents using the K-means clustering algorithm [14]. In this way, clusters consisting of documents that are similar to each other are obtained. This algorithm has been successfully used in other contents personalization [15], [16].

The adaptation algorithm designed for this system considers different aspects including: the student’s performance, knowledge level, number of times the students performs an evaluation, document location in each cluster, among others.

IV. CLUSTERING ALGORITHM

Document clustering algorithms have been used in the information recovery area, because they make it possible to identify typologies or groups where elements are very similar to each other and are very different from those in other groups [17]. K-means algorithm, proposed by [18] is a vicinity-based clustering method that is widely used because it is easy, fast and effective [19].

In the above clustering methods, K-Means is one of the widely used clustering techniques, because it is very efficient and easy to implement. It starts with k arbitrary clusters and partitions a set of objects into k subsets, which is an iterative hill climbing algorithm.

First, an initial selection of k prototypes or centers, which are considered representative of each cluster, is carried out; then each one of the collection elements is assigned to the cluster with the closest prototype. The next step consists of calculating the center of each one of the resulting clusters. The collection documents are again assigned to the closest group. Prior steps are repeated until k centers remain in the same cluster [18].

A. Adaptation Algorithm

The algorithm starts when the user presents the diagnostic evaluation. The result of this evaluation is considered in the first adaptation function \( f(\text{adapt}_i) \) which is detailed in (1).

\[
f(\text{adapt}_i) = \left( \frac{DE_i}{D_{\text{max}}} \right) - \frac{C_d}{C_{\text{max}}} \tag{1}
\]

In (1), \( DE \) represents document \( i \) Euclidean distance, i.e., the distance from the cluster center to the document. \( D_{\text{max}} \), is the largest distance between the cluster center and a document within it. \( C_d \) is the grade obtained by student \( j \) in the diagnostic evaluation. \( C_{\text{max}} \) is the highest grade that a student can get, according to the grading scale defined by the teacher.

This first adaptation function determines the knowledge the student must acquired to attain the learning objectives of a specific topic.

The first factor in the equation represents the weight of the document in relation to the pieces of knowledge to be conveyed. The second element in the equation expresses the diagnostic evaluation/maximum grade ratio to obtain the desired knowledge. A positive result indicates that the document will provide the student with knowledge; a negative value means that the student already possesses the knowledge that this document could provide.

\( f(\text{adapt}_1) \) is used to generate the first reading plan, that is, the listing of documents recommended by the system. Then the student can perform a number \( n \) of evaluations and thereby...
update his/her profile. A variation on the user profile will generate a change of plan reading, however, the system can notify you when you have made a previous reading of a document.

Further reading plans consider the student’s performance, so that the subsequent reading plan updates are governed by a second adaptation function, \( f(adap) \), which is detailed in (2).

Parameters considered in (2) are: \( D_E = \) document i Euclidean distance. \( D_{max} = \) the highest distance between the cluster center and a document within it. \( C_d = \) grade obtained by the student in the diagnostic evaluation. \( NC = \) the student’s knowledge level (1, beginner; 2, intermediate; 3, expert; 4, advanced). \( C_p = \) prior grade obtained by the student in a self-evaluation. \( C_{max} = \) the highest grade a student can obtain according to the grading scale defined by the teacher.

\[
P = \text{penalization percentage per number of attempts. } S = \text{guessing parameter (guessing likelihood), considering studies developed by [20].}
\]

\[
f(adap)_{ij} = \left( \frac{DE_i}{D_{max}} - \frac{CD \times NC}{C_{max}} + \frac{1-S_{max}}{4} \times P + \frac{C_p}{C_{max}} \right)
\]

The algorithm for the adaptation process is as follows:

**Adaptation Algorithm**

1. Identify the research area (topic)
2. Generate vector V with vectors ordered in ascending order according to their distance to the center of each cluster.
3. If (the plan is generated for the first time) and (the student performed the diagnostic evaluation) then Generate vector V* applying adap f(adap) adaptation function for each position of vector V.
   
   else
   
   If (the student self evaluated the topic)
   
   Generate vector V* applying f(adap) adaptation function for each position of vector V.
4. Calculate the average value of the vector V* obtained in the step above (average of contributions)
5. Using the average of contributions (previous step) documents around this value at ± 50% are evaluated. This value is looked for in vector E that represents the supplement to the student’s knowledge.
6. Position z, obtained in step 5 is multiplied by the size of vector V of documents and divided into the size of the scale. This provides position w of the first document to be shown.
7. The reading plan is completed with documents that are equidistant to that point.

**End of the Algorithm**

V. **SYSTEM ASSESSMENT**

AHSPeR was developed with the JAVA programming language, using MySQL Server.

The system can store different courses, which, in turn, will be sorted by topics or study areas. The topics will comprise a series of documents organized into clusters and are the main information source. The system is also able to record the events produced by its users; this will make it possible to include new variables to the personalization and recommendation rules in future research works.

A group of eight students of the last year of Computer Sciences at the Universidad de Carabobo, Venezuela, who are presently preparing their special degree projects, has been considered in this first assessment.

The sample consisted of six men and two women aged between 21 and 24 years of age. All they have used Moodle [21] as learning management systems and conducting research in the area of information system.

In empirical study, users are provided with a reading plan based on their knowledge on the topic; therefore, once the diagnostic evaluations were made.

At first the system was fed with a group of 130 documents that deal with different topics concerning the study area. Once the clustering algorithm was applied, 3 clusters were obtained: cluster A (Topic: Research Methodology) with 36 documents; cluster B (Topic: Software testing) with 68; and cluster C (Topic: Human Computer Interaction) with 26 documents.

Once the profile of each user was determined and a reading plan was generated for each one of them based on their individual profile and interests, it was observed that a reading plan with documents from cluster A, according to the interest pinpointed through the system interface, was generated for users 1, 5, 6 and 8.

The documents in cluster “Research Methodology” were organized according to their Euclidean distances to the center of the cluster (see Table I). Documents with a shorter distance to the center are closely related to the objective of the study topic; that is, those documents that are near the center are more related to the topic and those located at the cluster borders will be related to other topics. For example, a document relating to “quantitative research” will be located closer to the center, while one that deals with issues of “triangulation” will be located at the borders. These latter are recommendable for advanced students who master the study topic better.

**TABLE I**

**EUCLIDEAN DISTANCE OF SOME OF THE DOCUMENTS TO THE CENTER OF CLUSTER A**

<table>
<thead>
<tr>
<th>Doc</th>
<th>Distance</th>
<th>Doc</th>
<th>Distance</th>
<th>Doc</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5.44636</td>
<td>19</td>
<td>5.67848</td>
<td>28</td>
<td>7.11980</td>
</tr>
<tr>
<td>11</td>
<td>5.45829</td>
<td>20</td>
<td>5.81181</td>
<td>29</td>
<td>7.82890</td>
</tr>
<tr>
<td>12</td>
<td>5.46038</td>
<td>21</td>
<td>5.86193</td>
<td>30</td>
<td>8.46334</td>
</tr>
<tr>
<td>13</td>
<td>5.47376</td>
<td>22</td>
<td>5.96814</td>
<td>31</td>
<td>9.09812</td>
</tr>
<tr>
<td>14</td>
<td>5.48794</td>
<td>23</td>
<td>5.99178</td>
<td>32</td>
<td>9.65262</td>
</tr>
<tr>
<td>15</td>
<td>5.48856</td>
<td>24</td>
<td>6.06223</td>
<td>33</td>
<td>10.33098</td>
</tr>
<tr>
<td>16</td>
<td>5.49306</td>
<td>25</td>
<td>6.33050</td>
<td>34</td>
<td>11.78507</td>
</tr>
<tr>
<td>17</td>
<td>5.49987</td>
<td>26</td>
<td>6.41346</td>
<td>35</td>
<td>47.14277</td>
</tr>
<tr>
<td>18</td>
<td>5.51889</td>
<td>27</td>
<td>6.42010</td>
<td>36</td>
<td>56.71149</td>
</tr>
</tbody>
</table>

Once the procedure is carried out to determine z and w
values (See Adaptation Algorithm, step 5), which are fundamental to calculate the location of the documents that comprise the reading plan, data reflected in Table II was obtained.

<table>
<thead>
<tr>
<th></th>
<th>e1</th>
<th>e5</th>
<th>e6</th>
<th>e8</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>w</td>
<td>36</td>
<td>26</td>
<td>23</td>
<td>16</td>
</tr>
</tbody>
</table>

An analysis of this table shows that when 5 documents are to be recommended (parameter introduced by the teacher), student e1 (w = 36) was recommended documents d34, d35 and d36. Fig. 2 present the contribution of these documents to student e1 (diagnostic grade of 4.5).

The student has knowledge of 0.9 (straight line in Fig. 2), according to the second part of equation 1 (normalized value) and the documents that are closer to the line have a greater contribution to the student. The first documents are at the basic level and that they are very far from the knowledge each student has about the topic; that is, they are documents for level-1 students (beginner); whereas those that are at the cluster borders are in accordance with the level of student e1 (advanced). The equation considers the nearest neighbors to the document d36, according to the example (See Adaptation Algorithm, steps 5 and 6).

![Fig. 2 Relation between knowledge of student e1 and the contribution of each document](image)

VI. USERS’ SATISFACTION DEGREE

Students, with their information needs, are the main information source in this experiment. In this case, the Technological Acceptance Method (TAM) developed by [22] has been used. This method measures the quality of information systems and is used to forecast acceptance and the use of new technologies.

This model takes into account perceived usefulness and ease of use. The former one implies the belief that the use of technology boosts productivity; whereas the latter refers to the additional effort that the application of a given system entails. According to [23] TAM has been widely applied to studies of technology use, is considered the most parsimonious and powerful theory to describe the user acceptance of information systems [24].

TAM is the theoretical system most applied to assess ICTs acceptance within the scope of information systems in corporate and educational environments [25].

At the end of the experiment, the three questions on perceived usefulness show that 89.5% of respondents considered that their productivity was increased. This finding reflects a positive perception about the usefulness of the whole system.

With respect to the results obtained for easiness of use of the tool, when data in columns “High degree of agreement” and “Agreement” are summed, it can be observed that the entire sample considers that both learning as well as using the tool was easy.

It can be observed that data is in agreement with [26], [27], who point out that the easiness of use in an information system must positively influence its perception of usefulness.

VII. CONCLUSION

The best results could be related, in part, to the sensible decision to include the Euclidean distance among documents and grading of the evaluations of the students’ knowledge in the adaptation function. This suggests that it is important to properly select the elements contained in the adaptation function.

The result of this research work benefits the community of application developers in the area of information recovery in adaptive hypermedia systems, who desire to put this experience into practice.

The application makes it possible to monitor students, closely watching their evolution within a research area. This renders corrective actions, which must be taken at a given time in order to improve the students’ performance, easier.

Next steps will include the addition of other groups in order to compare the studied aspects and correlations among variables. Also it would be interesting to evaluate the system effect on teacher performance. We recommend in future researches to consider additional aspects in the survey to include more details about the software tools developed.

Finally, a study based on a set of hypotheses to learn about aspects of effectiveness and satisfaction.

The authors consider this paper as an initial model for integrating data mining techniques and process authoring for content adaptation in learning environments. A future study should focus on experiment with other clustering methodologies. So, other variables within the adaptation function may also be included.

REFERENCES


