User Profiling for Content Personalisation in Information Retrieval

Boris Rousseau, Parisch Browne, Paul Malone, Mícheál ÓFoghlú

Telecommunications Software Systems Group,
Waterford Institute of Technology,
Waterford, Ireland,
00353 51 302900

{brousseau, pbrowne, pmalone, mofoghlu}@tssg.org

ABSTRACT

One of the key issues with the overabundance of online information sources is that of finding what is relevant. The key to success for any type of information provider must be the personalisation of content in information retrieval, and this can be achieved through the maintenance of user profiles and the matching of these profiles to content metadata. This paper is concerned with user profiling and its role in content personalisation of information retrieval, and in particular presents a profile model which incorporates user preference information and action history information (representing the user's previous searches). The benefits and costs of such a model are examined and it is argued that the benefits (including personalisation accuracy, computational costs extensibility and flexibility) far outweigh the costs. The matching of profiles to metadata is also discussed as it fulfils an important role in the personalisation process. Although, the user profile model presented is focused on E-Learning, the general platform could be applied to other areas.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *clustering, information filtering, query formulation, relevance feedback, retrieval models, search process, selection process*.

H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *user profiles and alert services*.

General Terms

Management, Design, Standardization, Languages.

Keywords

User profiling, personalisation, GUARDIANS, Information Retrieval, search engine.

SAC'04, March 14-17, 2004, Nicosia, Cyprus

1. INTRODUCTION

In many respects the World Wide Web [5] has become a victim of its own success. With over 40 million web sites [26] on the Internet, the rapid growth in the volume of available information is making it difficult for users to quickly locate pertinent information. Users come from a range of different backgrounds with varied computer literacy and Internet skills, and these users have a wide range of interests and preferences. One particularly important set of Internet users is the business community, who arguably has limited resources in terms of their available time and effort. These business users simply cannot afford to wade through the masses of information, and they certainly would not pay to do it. It is more likely however that they would pay for personalised content, especially if they achieve this by doing the same type of internet keyword search that they are familiar with, and have the personalisation occur through the matching of user preferences to content metadata. Metadata is used to describe the resources available in terms of accessibility, organisation of data, and the properties of the corresponding domains. Metadata is also useful to provide descriptions of non-textual content, for instance, to describe multimedia document properties simplifying document management and retrieval. This matching of user preferences to content metadata should be a background process that requires little or no interaction from the users.

Similar issues exist within large organisations with vast amounts of information at their disposal on intranets. The effectiveness of modern information retrieval [1] is directly linked to the productivity of staff that should be able to easily retrieve pertinent information from the information stores within the organisation. The idea of user profiling can be summarised as "one that renders the user with an experience that is tailored to his/her current situation" [27].

Many disparate standards have existed which enable Information systems solutions within this scenario, but recently the emergence of XML [29], and its related standards, have provided some elegant potential solutions to this content personalisation issue. Key here is the maintenance of a user profile that records the user's interests, preferences and other information, and the use of this profile to aid in the personalisation of information retrieval.

This paper presents a subset of the results of an EU-funded IST project GUARDIANS namely: the modelling, querying and storage of a generic user profile, and the matching of a user profile with content metadata to provide personalised information retrieval.

2. BACKGROUND

A number of EU funded projects (e.g. EASEL [10], GESTALT [12], and GUARDIANS [13]) have recently been advancing the areas of management systems for E-Learning and personalisation of an individual learners programme. GUARDIANS is an IST project whose objective is to specify and implement an open distributed architecture for delivery and management of online personal information services via a range of technologies, and delivered to multiple end user platforms (e.g. PC, iDTV). The information services are expected to be increasingly composed of rich, interactive media, offering enhanced guidance to the user in navigating service offerings and responding to the user's understanding of the material provided. GUARDIANS has built on earlier work from the ACTS GESTALT project, and was completed in November 2002. The work carried out in GESTALT suffered from the immaturity of the technologies of the time, especially the limited support for XML and XML querying, and the lack of distributed platforms interoperability. Additionally, in GESTALT, the user profile was realised through the utilisation of the IEEE Public And Private Information Standard, but the specification was limited in that it lacked data structures and recommendations (profile binding).

3. USER PROFILE MODELLING

The problem of user profiling is multi-faceted, and the issues one must address include the choice of information to store and the decision of whether to use existing standards or to create new ones. These questions were answered based on the requirements of the GUARDIANS user profile and an investigation of the existing standards.

3.1 Existing Standards

Investigation of the existing profiling standards revealed that current standards were available and a brief description of these follows.

3.1.1 vCard (version 3)

The vCard specification from the Internet Mail Consortium is a means of Personal Data Interchange (PDI) [16], which automates the traditional business card. It can be used to store vital directory information (name, addresses, telephone numbers, email, URLs), geographic and time zone information, and can include graphics and multimedia (photo, logos, audio clips). The vCard has multiple language support, is standards based and the specification (based on RFC 2425 and RFC 2426) is transport and operating system independent.

3.1.2 IMS Learner Information Package (LIP)

The IMS Learner Information Package (LIP) [17] specification offers a data model that describes characteristics of a user needed for the general purpose of recording and managing learning related history, goals and accomplishments; engaging the user in a learning experience; discovering learning opportunities for user. The main elements are:

- Accessibility: In terms of language, disabilities, and preferences.
- Activity: Any complete learning-related (e.g. self-reported, formal/informal education, training, work experience, and military or civic service).
- Affiliation: Membership of professional organisations.
- Competency: Skills, knowledge, and abilities acquired in the cognitive, affective, and/or psychomotor domains.
- Goal: Learning, career and other objectives and aspirations.
- Identification: Biographic and demographic data relevant to learning.
- Interest: Information describing hobbies and recreational activities.
- Qualifications, Certifications and Licenses.
- Relationship: Relationship between components.
- Securitykey: The set of passwords and security keys assigned to the learner for transactions with learner information systems and services.
- Transcript: A record that is used to provide an institutionally based summary of academic achievement. The structure of this record can take many forms.

3.1.3 IEEE Public And Private Information (PAPI) Specification

PAPI [15] was created to represent student records and its development is moving towards harmonisation with IMS. It specifies data interchange formats, facilitating communication between cooperating systems. User records are divided into personal information and performance information and these are maintained separately. A key feature of the standard is the logical division, separate security, and separate administration of several types of learner information. The current specification splits the learner information into the following areas:

- Learner personal information: name, address, and telephone number (private to learner).
- Learner relations' information: learner's relationship to other users of learning technology systems, such as teachers, instructors, and other learners.
- Learner security information: learner's security credentials, such as: passwords, challenge/responses, private keys and public keys. This is private to the learner (with the exception of public information).
- Learner preference information: describes information that may improve human-computer interactions. This type of information is similar to personal information except that it may be public.

- Learner performance information: relates to the learner's history that is created and used by learning technology components to provide optimum learning experiences. Generally, learner performance information is created and used by automated learning technology systems.
- Learner portfolio information: is a collection of a learner's accomplishments and works that is intended for illustration and justification of his/her abilities and achievements.

3.1.4 Global TV-Anytime Specification

The TV-Anytime Forum [31] is an association of organisations that seeks to develop specifications to enable audio-visual and other services based on mass-market high volume digital storage in consumer platforms. The TV-Anytime Metadata specification employs metadata to describe content, user preferences, consumption habits, for targeting a specific audience. In particular, the consumer metadata section of the specification is interesting, as it describes how to define usage history description schema and user preferences description schema.

3.2 Generic User Profile (GUP)

Armed with this experience, GUARDIANS used the latest XML schema specification to represent the resources available. Emphasis was put on implementing a more generic user profile to incorporate most of today's user requirements in multiple domains. XML is the dominant standard for representing exchangeable data over the World Wide Web. After investigation, the IMS Learner Information Package was chosen as a basis for the user profile implementation.

The Profiles data model should conform to existing XML standards and specifications. Especially critical to its data and protocol model is its storage using the LDAP protocol (see next section). Moreover, scenario profiles have to be created bearing in mind existing specifications. While each of the standards investigated had its own merits it was felt that none fully addressed the profiling needs of the project. vCard has wide industry support and is easy to understand but it only stored a subset of information GUARDIANS was interested in. IEEE PAPI has advantages in terms of its level of security and also provides an extensible and generic specification but it is limited in terms of implementation examples and readability. Furthermore, the latest PAPI specification is moving towards a harmonization with IMS. IMS LIP provides best practice guides, is extensible, has an implementation from the IMS and is easy to use and read, which makes it the most suitable for the project user profile modelling requirements. IMS LIP, like the other standards, also has its disadvantages. It is relatively complicated, and is lacking in the area of user action history.

Figure 1 shows a tree diagram depicting the main element groupings in the GUP and a brief description of each follows. Note that only elements that differ from the IMS LIP specification are described.

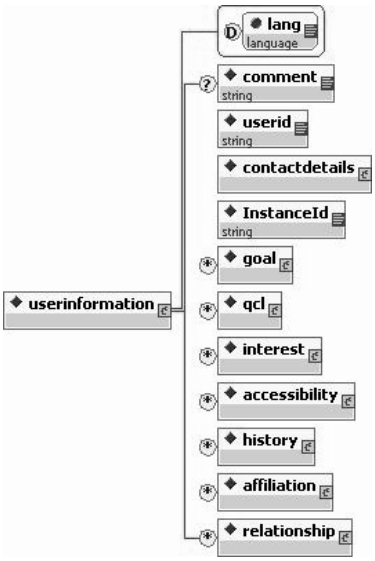

Figure 1: schema tree representation of profile elements

- *Lang:* attribute depicting the language used to represent the User Profile.
- *Comment:* optional comments about the User Profile.
- *Userid:* unambiguously identify a user.
- *Contactdetails:* vCard representing the different contacts of the user (home address, phone number, e-mail address, etc).
- *History*: history of the actions taken by the user based on the global TV-Anytime specification.

Certain LIP based element were removed such as the securitykey element, the reason being that the Lightweight Directory Access Protocol (LDAP, section 4) already has some security defined. Other LIP based elements were simplified. For instance, the identification element and its many children were replaced by a vCard [32] contact details element (see figure 2). Weighting attributes were also added to allow for prioritisation of preferences and could also contribute to the personalisation process.

```
<contactdetails>
<role>Student</role>
<entity>BEGIN:VCARD VERSION:3.0 REV:2002-08-01
CLASS:PUBLIC N:Galileo; BDAY:1975-04-30
ADR;TYPE=postal,parcel:;87 via Augusto, Napoli, ITALY
TEL;TYPE=mobile,voice,pref:+3912844864
EMAIL;TYPE=internet,pref:galileo@guardians.com;
END:VCARD</entity>
</contactdetails>
```

Figure 2: Contact details instance fragment

With regard to the type of useful information to store, an initial starting point was information about the technological platform for the communication. Additionally one can consider some user preferences information (languages, interest, etc). These are useful to guarantee an optimal information delivery path, to improve the user motivation and save time searching for resources.

When designing the schema to represent the user profile, there was a conscious effort made to make it extensible so that preferences could be added in the future without the need to rewrite the schema. This was possible with the three-tiered structure approach for each preference. Each preference comprised type, name and refinement elements such as that described below for a payment method preference.

```
<preference>
<type>payment method</type>
<name id = "t1">credit card</name>
<refinement id = "t1.2" weighting="5">visa</refinement>
</preference>
```

Figure 3: Three-tiered preference structure

This approach allowed representing both domain interest preferences such as for astronomy type, a name of stars and a refinement of sun. Furthermore, this structure allowed specifying different platform capabilities. For instance, the operating system, and supported formats extensions.

Weighting attributes were also added to the refinement elements within the preference structure to allow for prioritisation of preferences. These weightings can be used at either the query building or rating of results found stage. Weightings can range from 0 to 10 with 0 indicating that this preference should not be included in the search and a rating of 10 indicating that this preference is very desirable and must be matched against the content found. The default weighting where the user specifies none is 5.

It was also decided that the IMS LIP specification could be modified and extended to include the TV Anytime [31] concept of user action history. Of particular interest to the project is the consumer metadata section of the TV Anytime metadata specification. It describes how to define usage history description schema and user preferences description schema as well as the representation of audio-visual content.

The user profile used this structure as a basis for defining usage history. These elements were added to store the user actions and thus narrow further the next search request to the relevant resources. Figure 4 shows how the profile is updated according to the user actions. A user accesses the GUARDIANS system through a range of devices and provides some user information. From those information and some keywords a query is formulated by the Search and Mediation Service, which returns some results. At the end of the search process, the results selection is recorded for future search requests.

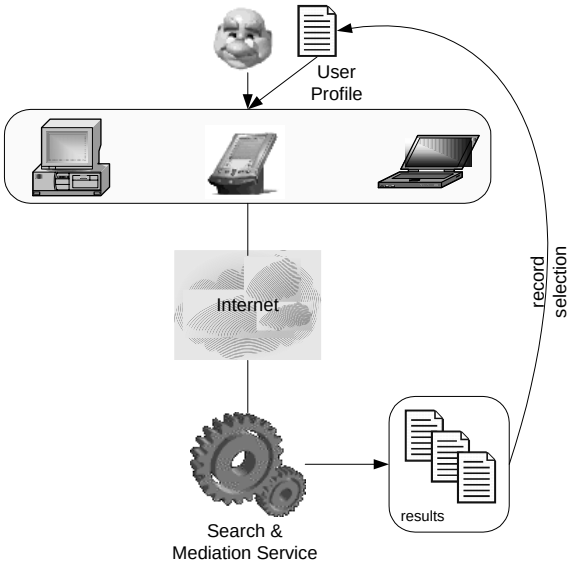

Figure 4: Updating user profile with content selection

4. STORAGE AND RETRIEVAL OF PROFILE

After defining the profile schema, the next issue to consider was the actual storage and retrieval of partial or complete profiles. Three potential solutions were investigated.

- Local client file system storage.
- Central RDBMS Database storage.
- Central LDAP Directory Service.

The first solution is to store the Profile data on a local hard drive. But this would create a problem of storage space in the case of a large number of users. For instance, with a User Profile of 8 Kilobytes, if a million users are registered with the system, 8 Gigabytes of hard disk space will be required. Such solution would require investigating and implementing compression mechanisms, which would ultimately slow down the profile information retrieval process.

Another solution is to store the data in a database system such as Oracle [22] but this raises issues of storing and manipulating XML documents. At the time of implementation, there were no mature, reliable, available XML database management solutions to this problem. Also, this solution can prove to be heavyweight for the overall system and consequently can affect the system's performance.

The final and chosen option is to store the data in a Directory Service such as the Lightweight Directory Access Protocol [33]. A directory is a specialised database optimised for storage, manipulation and retrieval of information in a globally scalable system, supporting sophisticated filtering capabilities and high volume search request. However, unlike database management systems, directory services do not support complicated transactions or rollbacks. A profile in a directory service refers to a person or an organisation having one or more relationships

with the organisation responsible for the maintenance of the directory. For instance, an Internet Service Provider that stores personal details and access rights.

Furthermore, the LDAP protocol already defines some basic classes for persons in an organisation (as it is derived from the X500 Directory Services standards that was developed explicitly for this purpose [2] [8]). The following describes some of the main classes:

- LDAP Person: is a simple class without many references to today's modern information about a person (such as e-mail). It can be used as a common denominator that most LDAP implementations use nowadays, for instance MS Outlook Express and Netscape Address Book.
- OrganisationalPerson: the OrganisationalPerson [34], which has also originated from X.500, but with more attributes, relates to a person's organisational information.
- InetOrgPerson: the InetOrgPerson [24] class is Netscape's proposal for the successor of OrganisationalPerson class. It incorporates new attributes such as e-mail address, user certificates and photos.
- EduPerson: EduPerson Draft Version 1.0 [11] is an auxiliary object class for campus-based directory aimed at facilitating communication among higher education in the USA. It contains all the basic set of data elements for representing individuals in higher education.

Additionally, LDAP provides mechanisms for client authentication, protecting the information the server contains. To store a profile in an LDAP directory service, sections of the profile have to be separated into LDAP classes ranging from abstract to more specific.

In the GUARDIANS implementation access to the profiles is provided through a CORBA IDL interface that exposes methods to retrieve, update, search, store and remove the profiles.

5. UTILISATION OF PROFILE IN PERSONALISED SEARCH

From a graphical user interface (out of the scope of this work), optionally containing previous searches, the user defines the keywords he wants to search on. However, unlike a traditional search engine, the search formulation process involves incorporating the user's preferences in the search request. The relevant user profile information for the particular domain in question is processed beforehand to formulate a personalised search query, thus narrowing the search to filter-out irrelevant results. A description of this process follows.

An underlying concept of the GUARDIANS Search Service is that it has no knowledge of the metadata formats on which the repositories are based. Instead it makes use of a canonical language to build schema independent queries and then uses a schema translation function to translate these queries into schema specific queries. The main idea behind this concept is that a word can be represented in many different ways depending on the metadata specification used. The canonical language serves as a middle language rather than providing the transformations between all the various metadata formats.

Hence, if a new repository is required to be integrated with the system, all is needed is a new transformation. Those conversions from various Metadata formats to one or more resource Metadata formats based on XML are achieved using the Basic Semantic Registry [18] tool.

In order to achieve this in an easily configurable manner, a set of XML and XSLT [7] [30] files are maintained whose purpose is to represent the necessary elements in the canonical language used in a query. These files are:

- target.xml: canonical language (CL) extract representing the elements the Search Service wishes to retrieve.

```
<cl>
<commonMetadata>
<descriptionMetadata>
 <descriptionIdentifier><value>id</value></descriptionIdentifier>
 <descriptionTitle><value>title</value></descriptionTitle>
 <descriptionDescription><value>description</value>
 </descriptionDescription>
</descriptionMetadata>
<technicalMetadata>
 <additionalInfo>
  <Title><value>format</value></Title>
  <Value><value>format</value></Value>
 </additionalInfo>
</cl>
```

Figure 5: canonical language targets to be used in the query

- keywords.xml: representing the elements which might contain keywords.

```
<cl>
<commonMetadata>
 <subjectMetadata>
  <keywords>
   <listSource value="classification" />
   <listChoice><value>classification</value></listChoice>
   <listSource value="general" />
   <listChoice><value>general</value></listChoice>
  </keywords>
 </subjectMetadata>
<descriptionMetadata>
  <descriptionTitle><value>title</value></descriptionTitle>
<descriptionDescription><value>description</value></description
Description>
</descriptionMetadata>
</commonMetadata>
</cl>
```

Figure 6: canonical language keywords to be used in the query

- 1.preferences.xsl to n.preferences.xsl: A set of XSLT files, which represent mappings from the user preferences (GUP format) to the canonical language. There is an arbitrary number (n) of these mappings, representing different levels of searches of the repository. n.preferences.xsl is a very restrictive set of mappings which attempts to map all of the user preferences to canonical language elements and as such find metadata which satisfies all the user's preferences. 1.preferences.xsl on the other hand is a very non-restrictive set of mappings. In fact, the '1 ' mapping might not map any user preferences at all or may only map to the users language. The preference files between n and 1 are progressively less restrictive.

```
<xsl:stylesheet version="1.0">
<xsl:output method="xml" standalone="yes" indent="yes"
encoding="UTF-8" omit-xml-declaration="no" />
  <xsl:include href="languserinfo.xsl" />
  <xsl:include href="language.xsl" />
  <xsl:template name="checkWeighting">
    <xsl:if test="@weighting">
      <xsl:attribute name="importance">
        <xsl:value-of select="@weighting" />
      </xsl:attribute>
    </xsl:if>
  </xsl:template>
</xsl:stylesheet>
```

Figure 7: mapping of preferences using XSL

The XSLT above is 1.preferences.xsl, the lesser restrictive transformation mapping only language preferences and its weightings. The XSLT preference files are iterated through to build a progressively less restrictive set of user preferences against which the metadata is checked. Using the three files it is possible to create three canonical language XML documents representing the target, preferences and keywords of the search. In general terms the search query can be represented as:

SELECT *target* FROM *repository* WHERE *preferences* AND *keywords*

Where *target, preference* and *keywords* are canonical language XML documents, and *repository* is the location of the metadata. The metadata is store using the Learning Object Metadata or Dublin Core standards specifications [19]

The Search Service queries the Directory Service to discover the content provider's repositories in this domain and using this information first translates the three canonical language documents into XML documents of the repositories' schema types. The information retrieved from the Directory Service also contains information about the location (a URL) of the web services used to query these repositories and the query language (SQL [14], XPath [9] or XQuery [6]) used by these web services. Once armed with this information the Search Service can go about the job of building a query to search these repositories for learning objects [21].

For ranking based on keywords, two main algorithms have been researched (although not in details due to the lack of resources): the TFIDF [23] and the vector spread activation. TFIDF is a traditional information retrieval measure, which rate the document according to the frequency of occurrence of query terms. The vector spread activation first assign a relevance scoring using TFIDF and then the score of a document is propagated to the documents it references.

The results discovered are then rated using a simplistic mechanism as follow:

1: a base score is given for all results (since it has been found, matching a minimum of one keyword).

2: a score is given for each keyword found in the title.

3: a score is given for each keyword found in the description.

4: a bonus score is given if the title or description contains more than one keyword.

5: a score is given for the preference level at which the result was found.

A greater number of preferences matched, means a better, more personalised result for the user.

6. EVALUATION AND TESTS

In order to perform an evaluation of the accuracy of the search with a user profile, it was necessary to develop a test client displaying the results of a search. It was developed in Java servlet and runs on the Apache Tomcat servlet container (version 3.2). This client interface gets keywords and user profiles as inputs and calls the Search Service to execute a search request on a GUARDIANS set of 243 documents (limitation due to the universities resources' IPR issues) based on NETTUNO [20] and Open University [2] resources. After the execution the number and the relevance of results are recorded.

Each series of test requests a search based on a set of two user profiles. A comparison is made between a search request making use of a profile and one without personalisation.

For this purpose, searches were performed with the following keywords, respectively: universe, sun, electronics and magnetic, Java and C++ and programming, mathematics. Also, two user profiles were available, in short: (1) an English speaker with text as his displayable format, (2) an Italian speaker with QuickTime as his preferred format.

The results can be summarised into the following chart:

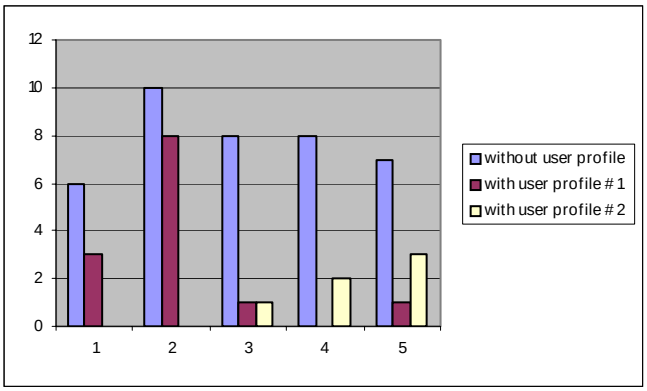

Figure 8: personalisation evaluation

In this graphical representation, the accuracy is much better when the search request is personalised. The fact is that without the use of a user profile, only keywords are taken into account for the search request whereas for personalisation, preferences are also included. In searches number 1, 2 and 4, the user does not get any result for two simple reasons: either no result was relevant to his particular language preference or his preferred format.

7. CONCLUSION: FINDINGS & FUTURE WORK

The GUARDIANS project has recognised the need for extensions to the currently used profile schemas and has implemented such an extended profile schema based on IMS Learner Information Package (LIP). This model was extended to also incorporate users' action history, which was missing from the LIP specification. Further additions to the schema include simplification of the current specification, weightings attribute and the introduction of a preference element.

All those changes made the user profile extensible and generic enough to satisfy most of today's E-Learning Management Systems requirements. This profile work has aided in the creation of an integrated system that allows for content personalisation in information retrieval. GUARDIANS project results have demonstrated that personalisation directs the user to more relevant content. As IMS LIP was not recommending any specific technology for storing the User Profile, LDAP Directory Server was chosen because of its hierarchical structure, which suited the profile specification.

LDAP provides an object oriented and hierarchical database where data is required to be structured in classes. LDAP offers a lightweight hierarchical database with the possibility of defining different layers of security. LDAP is more suitable than any database since it already defines some basic classes for persons in an organisation. In GUARDIANS, the profiles back end is Sun's iPlanet [25] Directory Server. This provides a platform and operating system independent environment to support LDAP. In terms of performance, iPlanet Directory Server supports thousands of user queries per second. Furthermore, it supports both Java technology-based and C client applications. LDAP allows the specification of generic classes, which permits reusability of components information.

Through the development of the GUARDIANS system, suggestions and recommendations were made either from other partners or from IMS feedbacks. After mentioning the GUP to CETIS-JISC [28] – monitoring IMS in the UK – it seemed to represent a major input to the current personalisation achievements in E-Learning. GUARDIANS is one of the few project to have successfully implemented a distributed educational system based on the IMS standards.

As the focus of this paper was E-Learning personalisation, information retrieval methods have not been researched. However, to apply this work on personalisation to other areas, a substantial study of information retrieval methods, rating/ranking mechanisms [1] will have to be researched. It is expected that the authors will be involved into future research in the area.

Dynamic information such as profiling agents is the way forward. For instance, an agent should be able to discover the user's platform preferences or get more information from previous searches. This issue was encountered while populating the user profile with some device information. Hence, a typical user would not know about its device(s) capabilities and will require an automatic process discovering it. A good balance between static and dynamic profile information should be researched further to present the user with the profile that requires the minimum efforts and at the same time the profiling service should not misinform the profile. A form of interaction between the profiling mechanism and the user is therefore crucial to get some feedbacks to the system.

The abstract representation of data on the World Wide Web, also known as the Semantic Web [4], is gradually taking over the current HTML based Web. In the near future, the development in the area will lead in significant new possibilities for machine to enhance and process the data that they merely display at present. Hence, programs will be able to exchange information with others and automate services. Such functionalities together with the flexibility of XML will be promoted using agents, that will help users tracking down resources and match them against their criteria and preferences.

8. ACKNOWLEDGEMENTS

The authors would like to thank the IST GUARDIANS (Gateway for User Access to Remote Distributed Information and Network Services) project consortium for the extensive discussions and also for pointing out the relevant literature on E-Learning and user profiling.

9. REFERENCES

[1] Baeza-Yates, R., Ribeiro-Neto, B. "Modern Information Retrieval", Addison-Wesley, ACM Press, May 1999. ISBN: 020139829X.

[2] Barker, P. “Experiences in providing a White Pages directory service based on the X.500 standard”, Computer Networks and ISDN Systems (1994) 26, pp. 1365-1374.

[3] BBC, the Open University open science Final Frontier: our universe, available at: http://www.open2.net/science/finalfrontier/youarehere/youarehere_index.htm

[4] Berners-Lee T., Hendler J. and Lassila O. “The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities”. S Scientific American, May 17, 2001. Available at: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

[5] Berners-Lee, T., Cailliau, R., Graff, J., and Pollerman, B., “World Wide Web: The Information Universe”, Electronic Networking: Research, Applications and Policy, Volume 1, Issue 2, 1992.

[6] Boag, S., Chamberlin, D., Fernandez, M.F., Florescu, D., Robie, J., Simeon, J., and Stefanescu, M., eds 2002. “XQuery 1.0: An XML Query Language”. W3C Working Draft, http://www.w3.org/TR/2002/WD-xquery-20020430/

[7] Bradley, N. “The XSL Companion”. Addison-Wesley Pub Co, September 7, 2000, ISBN: 0201674874.

[8] Chadwick, D. “Understanding the X.500 Directory”, Chapman & Hall, London (1994).

[9] Clark, J., and DeRose, S. eds 1999. “XML Path Language (XPath) version 1.0”. W3C Recommendation. http://www.w3.org/TR/xpath

[10] Educator Access to Services in the Electronic Landscape (EASEL) project Web site: http://www.fdgroup.co.uk/easel/

[11] EduPerson object class home page. Available at: http://www.educause.edu/eduperson/

[12] Gateway for User Access to Remote Distributed Information And Network Services (GUARDIANS) project Web site: http://www.ist-guardians.tv/

[13] Getting Education Systems Talking Across Leading-edge Technologies (GESTALT) project Web site: http://www.fdgroup.com/gestalt/

[14] Groff J.R., Weinberg P. N. “SQL: The Complete Reference”, Mc Graw-Hill Professionnal publishing, March 1999. ISBN: 0072118458.

[15] IEEE Public and Private Information Draft 8 specification, available at: http://www.edutool.com/papi

[16] Internet Mail Consortium Personal Data Interchange. Available at: http://www.imc.org/pdi/

[17] IMS Learner Information Package specification, available at: http://www.imsglobal.org/profiles/index.cfm

[18] ISO 11179 Parts 1-6, Specification and Standardization of Data Elements from the ISO/IEC 11179.

[19] Luis E., Anido Manuel, J., Fernández Manuel, Caeiro, Juan M., Santos, Judith S. Rodríguez and Martín Llamas “Educational metadata and brokerage for learning resources”, Computers & Education, May 2002, Volume 38, Issue 4, Pages 351-374.

[20] NETTUNO resources samples, available at: http://nettuno.unina.it/guardians/resources.html

[21] Palacios L. (2002), “Innovative Gateway to Learning Objects”, International Conference for Information and Communication Technologies in Education. Badajoz, Spain.

[22] Rousseau, B., Leray, E., ÓFoghlú, M. “Metadata & Information Management issues in XML-based Mediation”. Proceeding of the Fifth International Conference on Enterprise Information Systems, Angers (France), pages 375-379.

[23] Salton, G. and MCGill, M. J. (1983), “Introduction to Modern Information Retrieval”. McGraw-Hill, New York, NY.

[24] Smith, M., “Definition of the inetOrgPerson LDAP Object Class”. (2000). The Internet Society Request for Comments, 2798.

[25] SUN’s iPlanet web site (now Sun One). Available at: http://wwws.sun.com/software/product_family/iplanet.html

[26] Survey from http://www.netcraft.com/ dated from the 30th of June 2003.

[27] Suryanarayana, L., Hjelm, J., “Profiles for Situated Web”. WWW 2002, May 7-11, 2002, Honolulu, Hawaii, USA. ACM 1-58113-449-5/02/0005.

[28] The Centre for Educational Technology Interoperability Standards web site: http://www.cetis.ac.uk

[29] The World Wide Web Consortium eXtensible Markup Language specification, available at: http://www.w3c.org/XML

[30] The World Wide Web Consortium XML Stylesheet Transformations web site: http://www.w3.org/TR/xslt

[31] TV Anytime Forum Web site: http://www.tv-anytime.org/

[32] vCard overview. Available at: http://www.imc.org/pdi/vcardoverview.html

[33] Wahl, M., Howes, T., Kille, S. “RFC 2251: Lightweight Directory Access Protocol (v3)”. (1997). From the World Wide Web: http://www.ietf.org/rfc/rfc2251.txt

[34] Wahl, M., “A Summary of the X.500 User Schema for use with LDAPv3, 1997, RFC 2256”. (1996). From the World Wide Web: http://www.ietf.org/rfc/rfc2256.txt