Meta data and user profiles as base for REA agents

Matevž Pogačnik
Jurij Tasič
Faculty of Electrical Engineering
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
matevz.pogacnik@fe.uni-lj.si

Abstract

Meta-data and user profiles are emerging standards, which should simplify work of agents. Remote Education Application (REA) is one of such applications, which can benefit from the use of agents. In the article first the situation in fields of meta-data and user profiles is discussed. It is followed by a suggestion of possible use of agents in REA, based on meta-data and user profiles. Three major tasks for agents in REA are identified

1 Introduction

Increasing number of documents available on Internet is making organised search almost impossible. People have different preferences and interests, which makes each search unique. In order to make these searches more user-friendly new tools have emerged, called agents. Agents are not used only to perform searches, but to make applications more personal and efficient. One of such applications is Remote Education Application (REA), with focus on presenting lessons in an efficient and interesting way.

To create appropriate environment for work of agents, some requests have to be met. First, documents must be equipped with structured and detailed descriptions called meta-data, and second, users have to get profiles, so agents can work based on models, describing users. In past years quite some work has been done in this area, which will eventually lead to standards.

2 Meta data

Finding a particular item on the World Wide Web is similar to standing in the room full of unorganised books, only now the number of choices is exponential. The potential of the Web for providing rich materials and experiences is often abandoned as people find themselves relying solely on strategies of hit and miss. Commonly used are search engines, which in general use four types of search [3].

Full-text search engines (also referred to as free-text search engines) analyse the contents of files on referenced sites in such a way, as to allow users to search for any string of text they wish to find. Catalogue-based search engines (also known as index search engines) use some form of classification system, to help users to identify files, that have been labelled by human agents as providing useful information on a particular subject. Meta-search engines (also known as multi-search engines) allow users to search for the same keywords using more than one search engine, either sequentially or simultaneously. Specialist search engines are specifically designed to provide responses relevant to specific areas of knowledge [3]. Typically search engines seek to make the maximum possible number of 'hits', and then rely on users scanning through lists of possible sources to identify those relevant to their search. Number of irrelevant 'hits' among them is usually much higher then of relevant ones, which makes use of automated tools such as agents practically impossible. Therefore the World Wide Web requires a common methodology and language for describing online learning and other resources, that will facilitate purposeful and effective searching. Narrowing a search in a bookstore or library begins by identifying particular characteristics of the desired book. The multiple categories of the card catalogue system (e.g. subject, author, title) provide several different paths to the same object. These descriptive categories are meta-data about the object. Online materials need a similar system of meta-data.

The Instructional Management Systems Project (IMS), an Educom NLII initiative, is developing a specification and software for managing online learning resources. Learning resources not only include educational content but also people, educational service companies, tools, and activities.

According to their suggestion, meta-data is organised into categories, or fields. Each field represents a characteristic of the learning resource, for example, the resource's title, its unique identification number, summary, language etc. Each field has a value and some

fields may have multiple values. For example, the abstract field may have only one value, a brief summary of the material, or the field may have multiple values, if the same abstract is provided in several different languages. More important fields are mandatory, others are optional There are multiple schemas used for description of educational resources, depending on its type and purpose, but all derive from the Base meta-data schema. Base meta-data schema with no optional elements is presented below.

```
General:
    Identifier: DIdentifier
    Title: DLangString
Characteristics
    Language: DLocale
    Description: DLangString
    Discipline:
        TaxonPath:
            Source: DSource
            Taxon: DEntry
Life cycle:
    Version: DDecimal
    Create:
        Date: DDate
        Contribute:
            Role: DVocabulary
            Person: DPerson
    Publish
        Organisation: DOrganization
        Date: DDate
Meta-metadata:
    Create:
        Date: DDate
        Person: DPerson
        Organisation: DOrganisation
    Scheme: DString
Technical:
    Format: DFormat
    Locspec: DLocspec
    OSRequirements:
        Operating System: DVocabulary
Rights management:
    Role: DVocabulary
    Conditions:
        Price:
            Monetary unit: DVocabulary (if amount <>
0)
            Amount: DDecimal
```

In order to add meta-data to web pages (and resources displayed within web pages), IMS recommends embedding inline the meta-data as XML/RDF. For backward compatibility with browsers that do not support XML/RDF, IMS recommends using the HTML link tag as suggested by the World Wide Web Consortium[2] i.e. '<link rel="meta" href="mydocMeta-data">', where mydocMeta-data is a file containing meta data about its HTML page.

3 User model

User profiles are mobile, user-controlled collections of personal and educational data including personal information, performance information, and preference information. This data represents a rich resource that a user can draw on to facilitate, customise, and manage his or her learning experience(s).

A user model will typically cover different aspects of every user like capabilities, skills, knowledge, etc. which may refer to different application topics. User models should be organised in so called domains, according to thematical relevance of user model contents. They can be represented with tree structures. At this stage FIPA still hasn't standardised the aspects of structure of user models and at least perhaps the minimal set of user model contents. [5] This has been attempted in activities like P3P, standards like VCARD and Open Profiling Standard (OPS).

The later is mutual effort of Netscape Communications Corporation, Firefly Network, Inc., and Verisign, Inc., which announced the OPS, a proposed standard that enables personalization of Internet services while protecting users' privacy. This new proposed standard will provide Internet site developers with a uniform architecture for leveraging Personal Profile information to offer individuals tailored content, goods and services that match their personal preferences. Individuals will have a Personal Profile that contains their personal information. This profile will be stored on their personal computer. The first time that an individual visits a website that supports OPS, the website will request information from the Personal Profile.

.OPS allows for the trusted exchange of information of any sort and is fully extensible. There is a small number of "well known sections" contained in Personal Profiles. The first is a Unique Identifier that's assigned to the Personal Profile when it's first created. The second is a Unique Identifier that is unique to each service visited, and only available to that service. The third is basic demographic information (Country, Zip Code, Age and Gender) that's of use to a broad range of websites. The fourth is contact information (based on the vCard standard), such as name, address, zip or postal code, country of residence, telephone number, fax number, electronic mail address, etc. There will also be the possibility for creating sections for commerce information (such as Credit Card numbers, eCash, etc.) and site-specific information, such as detailed personal preferences (favorite books, magazines and music) that are of value to users in the context of one or a small group of websites. OPS places individuals in full control of their personal information and they can choose to release all, some or none of their information to websites that request it[4].

It seems that at present moment there is no standardised

user profile model (yet), used for educational purposes. What can be used are general suggestions e.g. OPS, with added items for each specific application.
Whatever the user model shall be, it should allow inspection and altering of her/his user model in the system. However, the user model is not always in the form where the user could interpret it from various parameters and rules. Also in this case, the user should be able to predict the effects of the adaptivity, and if s/he is not able to fine tune the model, s/he should at least be able to turn the model off. If preferences are hard to identify from user behaviour, like user's knowledge and habits, we can easily ask the user about her/his preferences. It is more difficult for the users to describe their knowledge or their habits. An ideal user profile should include information on how the user looks, what does s/he know, what is s/he interested in and what are her/his habits. However, the few commercial successes have either had a very simple knowledge of the user or have created their adaptations based on what the other users do, rather than complex models of individual users. [1]

4 Use of agents in REA

Learning means and environment are two very important factors when we are talking about efficient learning. It is not only important what is the content of a lesson, but also how it is organised and presented to a student. REA was designed to provide students with an interesting and helpful user interface on one hand and with a rich variety of possible lessons on the other. The first means it enables active ´surfing´ through lessons and possibility of presenting them in many different ways (e.g. textual documents, video and sound clips, interactive applets, vrmls, animations etc). The later means it is using massive storage device, which can record tens of gigabytes of lessons. The massive storage device is a linear tape from which parts of lessons are copied to hard disk and accessed from there. But these important characteristics alone are not enough. Some intelligence is necessary to manage all the functions in best possible way. In order to make it more efficient, REA is going to be equipped with agents.
Three major agent tasks can be identified at this moment for use in REA.

- presentation of lessons' structure in the user interface
- creation of lessons from material saved on massive storage device
- smart caching of lessons on hard disk

Intelligence in the user interface may violate many of the good usability principles developed for direct manipulation systems. Those principles include giving the user control over the system, making the system predictable so that it always gives the same response given the same input and making the system transparent so that the user can understand some of its inner workings. These principles increase the efficiency of agent system and prevent from unwanted agent behaviour.
User interface agents can follow the spirit of the user interface design guidelines by [1]:

- Control: the user should be able to turn the user interface agent off and to undo its actions when possible.
- Predictability: the user should be aware of the rules, which the agent follows.
- Transparency: although it is usually unnecessary and too difficult to understand the characteristics of the agent in details, the user should at least know what information the agent monitors, what output it provides and what are the basic rules of reasoning and operation.

Another problem is creation of user profile. It is clear that most of agents' actions will rely on acquired user profile. The questions that arise are which aspects of user profile to capture, what information does user's behaviour give us on his/her actual intentions and preferences and finally what use can be made of any such information captured. No standards have been made in this field, except for some very general suggestions. It is clear though, that the user profile should be:

- dynamical; its contents should be updated through observation of users actions
- hierarchical; contents of user profile should be organised according to relations among its attributes and fields in the profile should be weighted according to their importance (relevance)
- descriptive; it should contain all major attributes like user's skills and knowledge, its preferences and disliking

4.1 User interface

The user interface in REA includes a tree presentation of lesson's structure. All chapters and subchapters are there, as they follow in the lesson. It is clear that like chapters in a book there are some of higher importance and some of lower. Some of them are obligatory and some are just informational. Last but not least, some are more difficult then others. Lessons are designed for all students and must therefore cover many levels of knowledge. Still students' knowledge level and interests differ and require different presentations of the same lesson. This seems as an appropriate task for agents. They could change presentations of lessons by hiding difficult lessons from students with lower level of

knowledge and giving them easier homework. On the other hand they could present students with higher knowledge level with extra lessons, give them additional tasks etc. These actions should be based on student's user profile, its past actions and of course on difficulty structure of the lesson. Some demonstrational lessons that have been prepared are already rated according to the difficulty level, but to enable efficient functioning of agents, use of more complex meta-data is required [1]

4.2 Creating lesson structures

Similar concept to adaptation of lesson's structure is its creation. A lot of additional learning material to each lesson can be found on Internet. We can imagine a situation where teacher would suggest checking related material and REA system would automatically[2] save it to the massive storage device. A constructing agent would then create structure of a lesson, based on meta-data of saved materials. It would group related documents together, and make chapters and subchapters of it. This kind of operation might require specific types of meta-data that would normally not be part of widely used meta-data standards. Such meta-data should understandably label the nature of each document e.g. introductional, overwiev, detailed etc. It seems that standards are going to be flexible enough to allow addition of new labels. These labels could be added by the teacher, who suggested them. If not, the agent should still be able to make a structured lesson of the documents, although probably not so efficiently.

4.3 Smart caching

Use of the linear tape with large capacity is very appropriate for multimedia applications such as REA. They often use large files like video files and sound clips and since one linear tape has a capacity of ten average hard disks, they seem a logical choice. But there is another side to this story. Access time to those files on tape is very long, not only compared to hard disks or CD-ROM's, but also compared to user's patience. On average it takes 2 minutes to 'randomly' access a file. Therefore hard disks are used as cache for files used in REA. When the disk is full, unneeded files are removed from it and requested files are copied from tape. But how do we know which files are going to be accessed again with least probability, so we can remove them from disk? What if the user decides to use (read, watch, listen to...) part of the lesson which has just been removed from disk? In such case these files will have to be restored and others removed, which is a clear loss of time. Looks like a task for another agent. Based on user's previous actions (accessed lessons) structure of the lesson and user profile, caching agent should decide which parts of lesson (files) should be on disk in given moment. Copying and removal of files should be performed in advance, so the user would have the impression that entire lesson is on disk. Agent learning and adaptation should be done 'on-line' and improve in time.

[1] Possible use of meta-data would be according to IMS standard.

[2] This operation could be as well performed by an agent, for example after receiving teacher's e-mail, containing appropriate URL.

5 Conclusion

The best way to make a huge, unorganised collection of data useful is to structure and label it. This is the current tendency in organisation of Internet and other documents. Meta-data standards are evolving not only for general resources, but also for specialised types such as educational documents, tools etc. A good example is IMS initiative. Similar activities have been noticed in the area of user profiles, but apart from general suggestions such as OPS, nothing useful is yet to be found. Agents are the tools for which this entire infrastructure is intended for, and at the moment they are evolving and appearing in a wide variety of areas.
REA is an application, which should benefit from all areas mentioned. Its specific requirements can be met with some changes to suggested meta-data and user profile standards, but no major modifications should be needed. Further development in discussed areas is expected with great anxiety.

References

[1] Eija Kaasinen, Usability Issues in Agent Applications: What Should the Designer be Aware of, USINACTS, 1998

[2] Resource Description Framework (RDF), Model and Syntax Specification, http://www.w3.org/TR/WD-rdf-syntax/., 1998

[3] Searching techniques, European Commission DGXIII/E, 1998, http://www2.echo.lu/oii/en/search.html

[4] The Open profiling Standard (OPS), 1997, http://developer.netscape.com/ops/ops.html

[5] Human Agent Interaction, FIPA 98 Draft Specification: Part 8, http://drogo.cselt.stet.it/fipa/

[6] Thomas D. Wason, University of North Carolina, Chapel Hill, USA, IMS Meta-Data Structure, July 1998, http://www.imsproject.org/tehnical/metadata/library/did/171.html