



Information Retrieval in Virtual Universities

Juha Puustjärvi, Helsinki University of Technology, Finland
Päivi Pöyry, Helsinki University of Technology, Finland

ABSTRACT

Information retrieval in the context of virtual universities deals with the representation, organization, and access to learning objects. The representation and organization of learning objects should provide the learner with an easy access to the learning objects. In this article, we give an overview of the ONES system, and analyze the relevance of two information retrieval models for virtual universities. We argue that keywords based search (i.e., the Boolean model), though well suited for Web searches, is overly coarse for virtual universities. Instead, the vector model, on which our implemented search engine is also based on, seems to be more appropriate as it provides similarity measure (i.e., the learning object having the best match is presented first). We also compare the performance of four algorithms for computing the similarities (matching).

Keywords: algorithms; case study; distance learning; information retrieval; Web-based education

INTRODUCTION

Today people in all professions are faced with increasing demands. Technology develops in an ever-increasing speed, and the roles of people in work, society, and industry are shifting constantly. Keeping up with the pace of change requires continuous education and learning. Traditional campus-universities are trying to answer to this need of lifelong learning by building virtual universities, whilst facing competition from the commercial continuing education providers in the form of e-learning.

E-learning can be defined as information technology enabled and supported form of distance learning, in which the traditional restrictions of classroom learning have disappeared. The main tool of e-learning is a personal computer, and the Internet serves as the principal communication and distribution channel. The learners can participate in online Web-based courses and interact with both the peers, instructors, and the learning materials.

E-learning sets new requirements for universities: they have to build global learning in-

frastructures, course material has to be in digital form, course material has to be distributed, and learners must have access to various virtual universities.

As single virtual universities are independently created, they may provide very heterogeneous functionalities and user interfaces. Ideally, the learner should be able to access all the virtual universities in a similar way (i.e., the heterogeneity of various virtual universities should not burden the learner). How this goal can be achieved is the main topic of the ONES-project. Consequently, the main functions of the ONES system are to hide the distribution of e-learning portals, and to hide the semantic heterogeneity (i.e., problems arising from using same words in different meaning and vice versa).

In order to achieve these goals, the system will deploy many new technologies such as "one-stop portals," Web services, service oriented architecture, RDF-based annotation, ontology editors, and distance measures in searching learning objects.

In this article, we will restrict ourselves on the role of searches in the ONES-system. In particular, we will analyze the applicability of different information retrieval technologies. Our main argument is that the technology based on the Boolean model (Yan & Garcia-Molina, 1994), though well suited for searches in the Web, is not suitable for the emerging virtual universities. Instead, for virtual universities we have to develop methods, which allow learners to be more concerned with retrieving information about a subject than with retrieving data, which satisfy a given query. For example, a learner may be interested in courses dealing with object-oriented programming rather than in the courses where the term "java" or "C++" is stated.

When searching for information about a subject (e.g., object oriented programming) the search engine must somehow interpret the metadata of the learning objects and rank them according to a degree of relevance to the learner's query. The primary goal is to retrieve all the learning objects, which are relevant to a learner's query while retrieving as few non-relevant objects as possible. Unfortunately, characterization of the learner's information need is

not a simple task. Furthermore, the difficulty is not only in expressing the information need but also in knowing how the learning objects should be characterized with the help of the metadata descriptions.

The rest of this article is organized as follows. First, in the second section we give an overview of the architecture of the ONES-system. In the third section we characterize virtual universities. In particular, we will give an overview of the e-learning environment, and specify what the notion of resource-based learning incorporates. Then, in the fourth section, the role of metadata and ontologies in virtual universities is illustrated. In addition, the usability of the Boolean and the vector model in a virtual university is analyzed. Especially, two interpretations of a hierarchical ontology in the context of the vector model, called *weighted leaves* and *multilevel weighting*, are introduced. Then, in the fifth section, the performance of four matching algorithms based on weighted leaves and multilevel weighting principles is compared. Finally, the sixth section concludes the article by summarizing the feasibility of the proposed ideas.

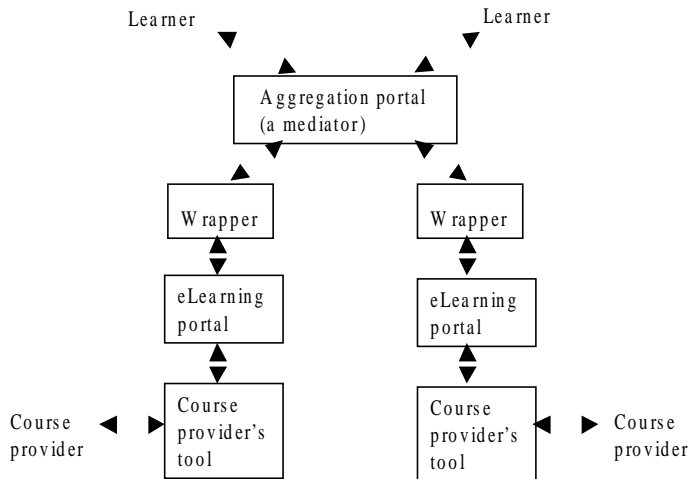
THE ARCHITECTURE OF THE ONES SYSTEM

The name ONES stands for One Stop e-learning Portal. As this name suggests, a salient feature of the system is the aggregation of distance learning information from different learning sources in one portal. The idea of the one-stop portals originated from one-stop shops, and later on it is also adopted in e-government applications. All one-stop applications have the same goal: hide the heterogeneity and distribution of local systems. So, from user's point of view one-stop portal behaves like a centralized system.

The four main components of the ONES-system are (see Figure 1):

- Aggregation portal (mediator),
- Wrappers,
- E-learning portals, and
- Course providers' tools.

Figure 1. ONES-architecture



The *aggregation portal* supports the learners in searching the courses that match to their specific needs. It differs from traditional database interfaces in a way that in addition to the traditional database queries it supports fuzzy queries. *Fuzzy queries* are similarity based, which means that if the similarity between the courses' profiles and the learner's query exceeds a certain threshold, they are said to match. A problem is that the current database management systems do not support fuzzy queries and therefore the ONES-system has to support them.

From technological point of view, the *aggregation portal is a mediator* (Garcia-Molina, Ullman, & Widom, 2000). It supports a virtual view that integrates several learning sources in much the same way as data warehouses do. However, since the mediator does not store any data, the mechanisms of mediators and warehouses are rather different. Since the mediator has no data of its own, it must get the relevant data from its sources and use that data to form the answer to the learner's query. As the data sources (e-learning portals) are independently created it is obvious that they provide heterogeneous interfaces (e.g., they may

provide different kind of functionalities or the same functionalities are provided by different operations).

In order to hide this heterogeneity there is a *wrapper* (Garcia-Molina et al., 2000) between the mediator and each e-learning portal. So a wrapper is a software module that extracts data from local e-learning portals. This implies that the wrapper must be able to accept a variety of queries from the mediator and translate any of them to the terms of local eLearning portal. The wrapper must also communicate the result to the mediator. An important point is that each wrapper provides equal functionality for the mediator. Ideally, each wrapper provides an interface for requesting the metadata of learning objects (i.e., descriptive information of courses, course packages and programs offered by educational institutions, e.g., universities).

From a technological point of view, each e-learning portal is a *Web service* (Vasudevan, 2001). Web services are self-describing modular applications that can be published, located, and invoked across the Web. Once a service is deployed, other applications (e.g., an aggregation portal) can invoke the deployed service. In

general, a Web service can be anything from a simple request to complicated business process.

A course provider can enter data about a course through the *course provider's tool*. The main function of this tool is to provide an interface, which facilitates the creation of the metadata attached to learning objects. Basically, this tool is analogous to the tools that support the content providers of electronic newspapers (Yli-Koivisto & Puustjärvi, 2002) in creating metadata items to news articles. The tool may even generate suggestions of the suitable metadata items, after which the author can make the necessary modifications and enter this information to the system.

CHARACTERISTICS OF VIRTUAL UNIVERSITIES

E-Learning Environment

E-learning can be defined as information technology enabled and supported form of distance learning, in which the traditional restrictions of classroom learning have disappeared (Liu, Chan, Hung, & Lee, 2002). The main tool of e-learning is a personal computer, and the Internet servers as the principal communication and distribution channel. The learners can participate in online Web-based courses and interact with both the peers and instructors and with the learning materials. The teacher-centeredness of traditional learning does not hold for e-learning, where the learning process has become more and more learner centered. The learning process and the resources may be customized according to the individual needs of the learner. At the same time, the role of the teacher becomes that of a facilitator or of a mentor guiding and supporting the individual process of learning (Liu et al., 2002).

Typical e-learning environments, such as WebCT and Virtual-U, offer the basic elements for delivering e-learning courses: course content delivery tools, synchronous and asynchronous discussion forums and conferencing systems, possibilities for quizzes and polling, workspaces for sharing resources, white boards,

possibilities for evaluation and grading, log-books, possibilities for submitting assignments, and so forth (Liu et al., 2002).

Studying in Virtual Universities

In the recent years, the idea of a virtual university has been becoming more and more popular in many countries all over the world. The enormous development in the field of information and communication technologies has enabled the rise of e-learning and virtual learning environments. As a result, the traditional universities have faced a new challenge emerging from the commercial sector of education. There is a growing need for new kind of learning and teaching as the technology advances rapidly and the skills and competencies required in the working life become more demanding and increasingly dynamic.

Virtual university has been defined as a space where the students are provided with higher education courses with the help of the newest information and communication technology (Niemi, 2002). The degree of utilizing technology in organizing the studies may vary from pure technology-based studies to face-to-face or mixed studies that are supported by learning technologies.

The main channel of communication and delivery of teaching is the Internet (Niemi, 2002; Ryan, Scott, Freeman, & Patel, 2000). Thus, a virtual university can be seen as closely related to e-learning that provides learning opportunities via the Internet. The difference between these two concepts is the level of studies offered; virtual university is aimed to offer higher education studies while e-learning can be used for all educational levels.

A virtual university may be an institution that uses the information and communication technologies for its core activities such as providing learning opportunities, administration, materials development and distribution, delivering teaching and tuition, and providing counseling, advising and examinations. On the other hand, a virtual university may also be a virtual organization created through partnerships between traditional universities and other educational institutes. In addition, the traditional cam-

pus universities may be regarded as virtual universities if they offer learning opportunities via the Internet or combine traditional ways of learning with e-learning (Ryan et al., 2000).

Virtual universities are expected to offer opportunities for life-long learning for audiences otherwise excluded from university studies. The emerging virtual university can be seen very beneficial especially for the industry, when technology-supported learning can be brought to the workplaces and integrated more closely to work. Moreover, virtual university can enhance organizational learning and bring competitive advantage by continuously developing the skills and knowledge of the employees (Teare, Davies, & Sandelands, 1999).

Resource-Based Learning

The Internet is able to store and transmit vast amounts of information in different forms and formats. Therefore the Internet is an ideal support for resource-based learning (RBL) that is one of the corner stones of learning and teaching in the virtual university. RBL has been defined a student-centered way of learning that exploits various specially designed learning materials, interactive media and technologies. RBL can be realized as self-study or as interactive group learning both in distance and in the face-to-face mode (Ryan et al., 2000).

The Internet can be used to enable and support RBL in several ways (Ryan et al., 2000):

- Courses can be delivered via the Internet.
- Resources can be identified and used.
- Internet serves as a communication and conferencing channel.
- Learning activities and assessment can be done in the Net.
- Collaborative work is enabled.
- Student management and support is enabled.

In the next section, we focus on the starting point of RBL, namely on searching learning resources.

INFORMATION RETRIEVAL MODELS

Information retrieval in the context of virtual universities deals with the representation, organization, and access to learning objects. The representation and organization of learning objects should provide the learner with an easy access to the learning objects. The system retrieves all the learning objects, which are relevant to learner while retrieving as few non-relevant learning objects as possible

In this section, we will analyze the usefulness of different information retrieval models (Baeza-Yates & Ribeiro-Neto, 1999) for a virtual university. The used model determines the way the metadata of the learning objects are given as well as the way the learner's queries (information needs) are presented. Before analyzing the information retrieval models we characterize the role of metadata and ontologies in virtual universities.

Metadata and Ontologies

In order to transfer data seamlessly and efficiently in the virtual university, there has to be a standard way for both people and computers to communicate all necessary knowledge, with both people and computer systems (Stojanovic, Staab, & Studer, 2001). One possible solution is to use metadata and an ontology attached to it for describing the learning objects.

The term *metadata* has variable interpretations depending upon the circumstances in which it is used. For example, in the context of documents the common forms of metadata include the author(s), the source of publication, the length of document, and so forth. This kind of metadata is commonly called *descriptive metadata*. For example, the metadata elements of the Dublin Core (Pöyry, Pelto-Aho, & Puustjärvi, 2002) represent descriptive metadata.

Educational metadata is needed for improving the retrieval of learning objects, for supporting the management of collections of learning objects, and for supporting the decision process of the learners looking for educa-

tional resources. LOM seems to be the most powerful and most widely used metadata standard for educational information systems (Holzinger, Kleinberger, & Müller, 2001; Lamminaho, 2000). More generally, educational metadata can be used by educational institutes and professionals as well as by learners in order to describe (e.g., the content, structures, and relationships of the learning objects and to search for educational objects) (Lamminaho, 2000; Stojanovic et al., 2001).

Educational metadata may describe any class of educational objects, such as study courses. The pedagogical features of the course, the contents, special target groups, and the technical requirements of the study course can be described with the help of a metadata schema (Lamminaho, 2000). More generally, educational metadata can be used to describe, for example, the content, structures, and relationships of the learning objects (Stojanovic et al. 2001). Educational metadata can be utilized by educational and pedagogical professionals, by the institutions offering education, and by the students searching for education. Well-designed and sufficient metadata aid the decision making process of the students and help the educational institutions to provide suitable information about their educational supply (Lamminaho, 2000). Educational metadata is very much semantic metadata, but a thorough metadata schema must include also at least structural metadata in order to be able to describe the learning objects efficiently.

The idea of using standardized metadata schemas is being able to develop universally applicable tools dealing with the metadata descriptions of the learning objects. In order to create metadata records containing the resource descriptions specific tools are needed for creating the metadata according to the standards (Kassanke, El-Saddik, & Steinacker, 2001). Metadata is also useful when guiding non-experienced users through a large collection of learning resources (Strijker, 2001). Moreover, metadata is seen as value-added information that is used to arrange, describe, track or otherwise enhance the access to the object content. At the moment metadata becoming increasingly

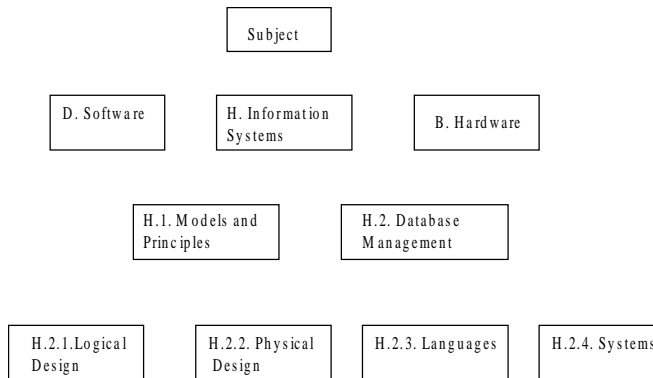
important when digital government and e-commerce are emerging. Metadata enables increased accessibility, expanded use of objects, multi-versioning, and system improvement. The granularity of metadata, which refers to the level of details in the description, is an important question when developing a metadata set (Gilliland-Swetland, 2000).

A salient feature of *descriptive* metadata is that it is external to the meaning of the document, (i.e., it describes the creation of the document rather than the content of the document). The metadata describing the content of the document is commonly called *semantic metadata*. For example, the keywords attached to many scientific articles represent semantic metadata (Jokela, 2001).

An *ontology* provides a general vocabulary of a certain domain (Fridman & McGuinness, 2001), and it can be defined as “an explicit specification of a conceptualisation” (Gruber, 1993). In essence, an ontology gives the semantics to the metadata. Ontologies are *formal, explicit, and shared* specifications of some conceptualizations. Formal means that the ontology should be machine readable, and explicit refers to having defined the types of concepts and the constraints on their use are explicitly defined. Shared refers to the fact that an ontology must reach a consensus (Fensel, 2001). Ontologies together with metadata enhance efficient access to information by offering possibilities to organize and categorize the content of the information system in question. In this context an ontology is defined as a means to formalize and to specify a common terminology for a defined area of interest (Turpeinen, 2000).

In order to standardize semantic metadata specific ontologies are introduced in many disciplines. Typically, such ontologies are hierarchical taxonomies of terms describing certain topics. For example, the ACM Computing Classification System is a hierarchy (a tree) in which the nodes represent the classes of the taxonomy. In Figure 2, a subset of that hierarchy is represented.

Figure 2. A subset of the ACM Computing Classification System



The Boolean Model

Applying the Boolean model in searches requires that each learning object is augmented by a set of metadata items such as keywords or classification identifiers (e.g., the searches in the CUBER system (Pöyry et al., 2002; Pöyry & Puustjärvi, 2003) are based on the Boolean model). A learner can then query learning objects by Boolean expressions comprising of operands and operations. The operands are the used keywords and the operators are typically “and,” “or,” and “not.” For example, by using ACM Computing Classification system (Figure 2) the keywords attached to a learning object might be D, H.1, and H.2.2 (corresponding the keywords *Software*, *Models and Principles*, and *Physical Design*). Now, if a learner presents the query “D and (B or H.1)” (i.e., learning objects having the keyword “Software” and at least one of the keywords “Hardware” and “Models and Principles”), then the previous learning object will match that query.

The Boolean model is intuitive and clear. Moreover, it can be efficiently implemented even in the case of huge amount of objects. For example, many Web search engines are based on this model. However, using that model in a virtual university gives rise to following drawbacks:

- First, the model is based on a binary decision criterion, meaning that each learning object is predicted to be relevant or non-relevant. In reality, it is obvious that the resulting learning objects fit more or less to the query (i.e., some kind of grading should be possible).
- Second, expressing the requirements of learning objects by a Boolean expression may be difficult.
- Third, a typical problem concerning search engines based on the Boolean model is that either the result of the query includes too many or too few learning objects.

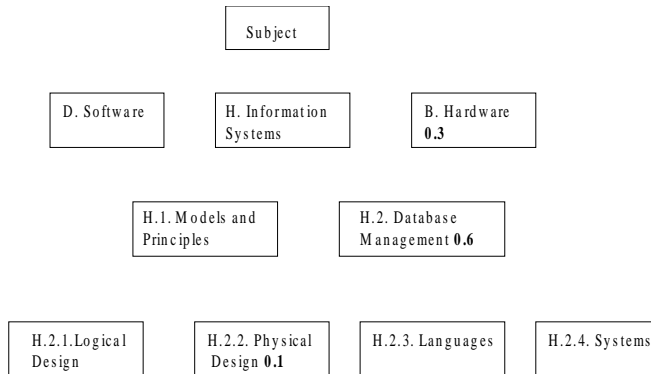
In the next section, we consider a more advanced model, which avoids many of the drawbacks just described.

The Vector Model

The vector model differs from the Boolean model in that weights can be assigned to each metadata item of a document as well as to the keywords of the query. The idea behind this model is that we can more accurately specify the queries and the contents of the documents (e.g., learning objects).

Assuming that the standard metadata items (e.g., the classes in Figure 2) specify a vector space (i.e., each item (keyword) in the

Figure 3. A metadata specification of a learning object



hierarchy represents a dimension in the vector space), we can represent each document and query as a vector in that vector space. Then we can process the query by computing the distance of the query vector and the document vectors. This kind of computing requires that the sum of the weights of each document and query equals to a predefined constant. For convenience, the used constant is usually one.

As the result of the query the documents are sorted in the order determined by the similarity (i.e., the document having the best match with the query is presented first). The number of the documents in the result should be restricted by requiring a certain degree of similarity.

Using the vector model in a virtual university requires that the course provider assign the metadata items and their weights into each learning object. The metadata items to be used are selected from the used domain ontology. Depending on the used course provider's interface this can be done in various ways. For example, as in our prototype system, there may be an ontology structure on which the course provider inserts the weights. In Figure 3, the ontology structure of the Figure 2 is augmented by setting weights on the nodes "B.H.2," and "H.2.2." Note that the node having no weight means that its weight is actually zero. Hence,

the profile of the learning object can be presented by a vector in 9-dimensional vector space as follows: $[0 \times D, 0 \times H, 0.3 \times B, 0 \times H.1, 0.6 \times H.2, 0 \times H.2.1, 0.1 \times H.2.2, 0 \times H.2.3, 0 \times H.2.4]$. That is, the profile is a point in an orthogonal 9-dimensional vector space.

The gain of attaching metadata description for learning objects is that we can use mathematical distance measures in computing learners' queries. Further, computing the distance requires that the descriptions (vectors) be specified in an orthogonal vector space. In other words, the nodes in the hierarchy that are used in profile vectors must be independent. In practice this means that we have to follow one or the other of the following interpretations:

- **Multilevel weighting interpretation:** The leaves and the nodes of the ontology hierarchy represent independent concepts.
- **Weighted leaves interpretation:** The parent node represents the union of its siblings. In other words, each sibling represents a subset of its parent. Yet the siblings represent independent concepts.

The intuition behind *multilevel weighting* is that we can express the level of a leaning object (as well of a query) by altering the weights on a node and its siblings. To illustrate this let

us consider the weighting of the course “Physical design in database management systems.” Now, it is obvious that the weights should be given on the node H.2 (Database management) and its siblings H.2.2 (Physical design) and H.2.4 (Systems). Assuming that approximately half of the course deals with databases in general and the other part deals with physical design and database management systems, then giving weight 0.4 to H.2 (Database management), 0.3 to H.2.2 (Physical design) and weight 0.3 to H.2.4 (Systems) could be an appropriate assignment. On the other hand, if the course is very specific then the weight of H.2 could be zero.

If we follow the *weighted leaves* interpretation, then in determining the profile of a learning object weights are set only on the leaf nodes of the hierarchy. Consequently, the profiles of the learning objects are specified by vectors in an orthogonal vector space, which is determined by the leaf nodes of the hierarchy. To illustrate this approach let us consider the weighting of the course “Physical design in database management systems.” In this case, all the weights are given on the nodes H.2.2 (Physical design) and H.2.4 (Systems) independently of the level of the course.

PROCESSING LEARNER'S QUERIES

The learner presents queries in the same way as the content provider determines the weights of the learning object; both these are presented by vectors. Hence the query presents an ideal profile of the learning objects that satisfy the learner's requirements. For example, assuming that the multilevel weighting interpretation of the ontology is used, and a learner wants to find basic courses concerning database management. In this case the learner will set rather heavy weight on H.2 (database Management) and lighter weights on H.2.1 (Logical Design), H.2.2 (Physical Design) and H.2.3 (Languages). In contrast, if a student is looking more advanced courses on database management then the student will give a lighter weight on H.2 and heavier weights on its siblings.

As the learners interact with the system by submitting queries it is reasonable to require that the response times should be only a few seconds. We investigated the effects of different matching algorithms and the amount of stored learning objects on response times. The test environment was equipped with Pentium II processor and 192 MB memory. The computers were running the Sun Solaris 5.8 operating system. We implemented and tested four matching algorithms (i.e., algorithms) which compute the distance measures of learning objects and learners' queries. We next give a short description of the algorithms.

The *Cosine matching algorithm* (Baeza-Yates et al., 1999) calculates the cosine measure between the query (a vector) and the documents profiles. As a matter of fact the algorithm does not compute distance measures but rather approximates distance measures by computing the angles of the query vector and the vectors representing documents, such as the learning objects.

The *Euclidean matching algorithm* (Friedman, Bentley, & Finkel, 1977) calculates the Euclidean distance from the query profile to all learning objects' profiles. The *Manhattan distance algorithm* (Bentley, Weide, & Yao, 1980) calculates a so called “city block-distance.” The name comes from the fact that this measure in two dimensions tells how many blocks in a city one would have to walk between two points.

Our developed *Fuzzy matching algorithm* attempts to achieve more efficient matching procedure than the “exact” matching algorithms. The improved efficiency is achieved by performing the actual matching on a pre-selected subset of all learning objects. The predefined subset of the documents' profiles is determined by choosing the three biggest weights from the query and then computing the subset based on these weights. Then only the profiles, the weights of which are within a specified tolerance interval are selected for the final query processing. Therefore the result set is not guaranteed to contain all the profiles that are closest to the matching profile. However, the close-

Table 1. Matching times for the algorithms

	1 000	5 000	10 000
Cosine	0.80	0.93	1.07
Euclidean	0.83	0.96	1.16
Manhattan	0.83	0.98	1.23
Fuzzy	0.61	0.73	0.89

ness values of the profiles in the actual result set are exact, since they are calculated using the Euclidean measure.

The computing time for matching of each algorithm is presented in Table 1. The test was performed for different amount (1000, 5000 and 10 000) of learning objects. Basically, the differences of Euclidean, Cosine and Manhattan algorithms were rather small (less than 10%). Fuzzy matching algorithm required least computing time (about 20% less than others). However, the test proves that all the algorithms are quick enough in the test environment as the response times are less than 1.2 seconds. If the number of the learning objects or the dimensions of the vector space (i.e., the used attributes in the profile) increases, then it obvious that the Fuzzy Matching algorithm will be more superior to the other algorithms. In our test environment the vector space comprised of 15 dimensions (i.e., each profile could have at most 15 attributes). In practice, the number of attributes cannot increase significantly as otherwise the determining the weights for learning objects would overly burden the coarse creators. In addition, as the system is developed for universities it is not obvious that number of learning objects can be very huge (e.g., over 10,000).

CONCLUSION

Virtual university has been defined as a space where the students are provided with higher education courses with the help of the newest information and communication technology (Niemi, 2002). The degree of utilizing

technology in organizing the studies may vary from pure technology-based studies to face-to-face or mixed studies that are supported by learning technologies.

A virtual university may be an institution that uses the information and communication technologies for its core activities such as providing learning opportunities, administration, materials development and distribution, delivering teaching and tuition, and providing counseling, advising and examinations. On the other hand, a virtual university may also be a virtual organization created through partnerships between traditional universities and other educational institutes. In addition, the traditional campus universities may be regarded as virtual universities if they offer learning opportunities via the Internet or combine traditional ways of learning with e-learning (Ryan et al., 2000).

E-learning sets new requirements for universities: they have to build global learning infrastructures, course material has to be offered also in digital form, course material have to be distributed via the Internet and learners must have access to various virtual universities. A problem is that the current virtual university portals provide heterogeneous functionalities, which in turn hampers the learner in accessing various virtual universities.

The main goal of the ONES-project is to investigate the ways of integrating various virtual universities in a way that such an aggregated virtual university would be as easily accessible for a learner as a single virtual university. Achieving such a goal requires mutual understanding of the used technology and standardized descriptions of the learning ob-

jects. Furthermore, searching from various virtual universities requires mutual understanding of the information retrieval model to be used.

We argue that keywords-based search (i.e., the Boolean model), though well suited for general Web searches, is unsuitable for the virtual universities' purposes. Instead, the vector model (on which our implemented search engine is also based on) seems to be more appropriate as it provides a similarity measure (i.e., the learning object having the best match is presented first. We also introduced two interpretations for the hierarchical ontologies, which allow increasing the power of the used metadata descriptions. And finally, we also compare the performance of four algorithms for computing the similarities of the profiles. It turned out that our developed Fuzzy Matching algorithm requires less computing time as the other "exact matching" algorithms represented in the literature.

REFERENCES

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: Addison Wesley.
- Bentley, J., Weide, B., & Yao, A. (1980). Optimal expected-time algorithms for closest point problem. *ACM Transactions on Mathematical Software*, 6(4), 563-580.
- Fensel, D. (2001). *Ontologies: Silver bullet for knowledge management and electronic commerce*. Berlin: Springer Verlag.
- Fridman, N. N., & McGuinness, D. L. (2001, March). *Ontology development 101: A guide to creating your first ontology* (Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, Stanford Medical Informatics Technical Report SMI-2001-0880).
- Friedman, J., Bentley, J., & Finkel, R. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3(3), 209-226.
- Garcia-Molina, H., Ullman, J., & Widom, J. (2000). *Database system implementation*. New Jersey: Prentice Hall.
- Gilliland-Swetland, A. J. (2000). *Introduction to metadata, setting the stage*. Retrieved December 20, 2004, from <http://www.getty.edu/research/institute/standards/intrometadata/>
- Gruber, T. R. (1993, March). Toward principles for the design of ontologies used for knowledge sharing. In *Padua Workshop on Formal Ontology* (p. 23).
- Holzinger, A., Kleinberger, T., & Müller, P. (2001). Multimedia learning systems based on IEEE Learning Object Metadata (LOM). In *Proceedings of ED-MEDIA 2001*, Tampere, Finland.
- Jokela, S. (2001). Metadata enhanced content management in media companies. In *Acta Polytechnica Scandinavica. Mathematics and computing series no. 114*. Doctoral thesis, Helsinki University of Technology.
- Kassanke, S., El-Saddik, A., & Steinacker, A. (2001). Learning objects metadata and tools in the area of operations research. In *Proceedings of ED-MEDIA 2001*, Tampere, Finland.
- Lamminaho, V. (2000). *Metadata specification: Forms, menus for description of courses and all other objects*. CUBER project, Deliverable D3.1.
- Liu, J., Chan, S., Hung, A., & Lee, R. (2002). *Facilitators and inhibitors of e-learning*. In L. C. Jain, R. J. Howlett, N. S. Chalkaranje, & G. Tonfoni (Eds.), *Virtual environments for teaching and learning, Series on innovative intelligence* (Vol. 1, pp. 75-109). World Scientific.
- Niemi, H. (2002). *Empowering learners in the virtual university*. In H. Niemi, & P. Ruohotie (Eds.), *Theoretical understandings for learning in the virtual university*. University of Tampere, Research Center for Vocational Education and Training.
- Pöyry, P., Pelto-Aho, K., & Puustjärvi, J. (2002). The role of meta data in the CUBER system. In *Proceedings of the Annual Conference of the SAICSIT 2002* (pp. 172-178).
- Pöyry, P., & Puustjärvi, J. (2003). CUBER: A personalised curriculum builder. In *Proceedings of the 3rd IEEE International Confer-*

- ence on Advanced Learning Technologies*, Athens, Greece (pp. 326-327).
- Ryan, S., Scott, B., Freeman, H., & Patel, D. (2000). *The virtual university. The Internet and resource-based learning*. London: Kogan Page.
- Stojanovic, L., Staab, S., & Studer, R. (2001). E-learning based on the Semantic Web. In *Proceedings of WebNet2001 — World Conference on the WWW and Internet*, Orlando, FL.
- Strijker, A. (2001). Using metadata for re-using material and providing user support tools. In *Proceedings of ED-MEDIA 2001*, Tampere, Finland.
- Teare, R., Davies, D., & Sandelands, E. (1999). *The virtual university — An action paradigm and process for workplace learning*. Cassell.
- Turpeinen, M. (2000). Customizing news content for individuals and communities. In *Acta Polytechnica Scandinavica. Mathematics and computing series no. 103*. Doctoral thesis, Helsinki University of Technology.
- Vasudevan, V. (2001). *A Web service primer*. Retrieved December 20, 2004, from <http://www.xml/lpt/a/2001/04/04/Webservices/index.html>
- Yan, T., & Garcia-Molina, H. (1994). Index structures for selective dissemination of information under the Boolean Model. *ACM Transactions on Database Systems*, 19(2), 332-364.
- Yli-Koivisto, J., & Puustjärvi, J. (2002). CoMet: An electronic newspaper prototype. Workshop on XML in Digital Media. In *Proceedings of the 8th International Conference on Distributed Multimedia Systems (DMS'2002)* (pp. 703-707).

J. Puustjärvi obtained his BSc and MSc in computer science in 1985 and 1990, respectively, and his PhD in computer science in 1999, all from the University of Helsinki, Finland. Currently he is a professor of information society technologies at the Technical University of Lappeenranta. He is also a docent of e-business technologies at the Technical University of Helsinki, and a docent of computer science at the University of Helsinki. His research interests include e-learning, e-business, knowledge management, Semantic Web and databases.

P. Pöyry obtained her BA (Educ) and MA (Educ) in 2001 and 2002 from the University of Helsinki, Finland. Ms. Pöyry obtained her LicSc (Tech) in 2004 from the Helsinki University of Technology, where she is a doctoral student. Currently she works as a researcher and prepares her PhD thesis at the Helsinki University of Technology in the Software Business and Engineering Institute as a member of the Information Ergonomics Research Group. Her research interests include e-learning, knowledge management, CSCW, and usability research.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.