IKUM: An Integrated Web Personalization Platform Based on Content Structures and Usage Behavior

Joannis Vlachakis
Fraunhofer IAO,
Nobelstr. 12,
70569 Stuttgart, Germany,
joannis.vlachakis@iao.fhg.de

Magdalini Eirinaki
Department of Informatics,
Athens University of
Economics and Business,
Athens, GREECE,
eirinaki@aueb.gr

Sarabjot Singh Anand
Department of Computing &
Mathematics,
University of Ulster at
Jordanstown, Northern Ireland,
ss.anand@ulster.ac.uk

Abstract

The objective of the I-KnowUMine project (IKUM)[1] is to develop an integrated platform (referred to in the paper as the "IKUM system") that uses state of the art technology and research results from different application domains in order to provide the basis for the development of online services in a wide range of application areas, presenting personalized content, services and applications to users in a structure more suited to their needs. The benefits provided by the IKUM system result mainly from the combination and integration of technology advances in spaces such as Web Mining, Content Management, Personalization and Portals . As a result of this novel combination of these technologies, users of the IKUM system will benefit from the optimal logical structure of information/content provided by the system, allowing them to efficiently execute their processes and to reach their information targets.

1 Introduction

The continuous growth of the World Wide Web in terms of content and usage has heightened the need for new methods in design and development of online services to the end-user. The objective of the I-KnowUMine project is the development of a novel and innovative content delivery platform based on Content, Knowledge and Behavioral data to present personalized content to users in a structure more suited to their needs. Users of the IKUM system benefit from the optimal logical structure of information/content provided by the system, allowing them to efficiently execute their processes/tasks and to reach their information targets. The benefits provided by the IKUM system result mainly from the combination and integration of technology advances in spaces such as Web Mining, Content Management, Personalization and Portals. The system combines knowledge of typical user behaviors with rules and conditions of the underlying content structure and semantics in order to provide the optimal flow of information in content provision applications and services assuring contextual content integrity.

The rest of the paper is organized as follows: In section 2 we define the space of Content Contextualisation Servers (CCS) in the context of the application spaces adjacent to it, technologies from which influence the ability to deliver a system within the CCS space. In section 3 we present related research efforts to IKUM, focusing on those that present an integrated Web personalization system, using usage and content data. The IKUM architecture is described in more detail in Section 4. We present the modules that comprise the system and detail its most innovative features.

2 Content Contextualization Servers

As the IKUM objectives span across a number of application areas, there is a need to define the space within which systems like the IKUM system exist. We refer to this space as the **Content Contextualization Server** space. The Content Contextualization Server is defined as *"Software that delivers content deemed relevant to a users context, taking into account the users behavior and semantic content preferences as defined by the users previous and current access of content."*

As the users context is dynamic, so must be the strategies for structuring the content delivered to the user. Hence, Content Contextualization Servers typically require strong predictive analytics, multiple recommendation strategies, content and knowledge management capabilities and flexible content delivery functions. Thus, content contextualization servers depend on:

- Traditional content management servers to provide components for content authoring and publishing.
- Web Mining components for data collection, pre-processing, analysis and generation of knowledge for use in personalization.

[1] The research work presented in this paper was partially supported by the IST-2000-31077/I-KnowUMine R&D project funded by the European Union

- Multiple personalization components for using the knowledge generated and recommendation of content based on current behavior.
- Portals to provide the deployment platform for content contextualization

Given the definition of the Content Contextualization Server space, the following four spaces are deemed to be adjacent to it: Personalization, Web Mining, Content Management and Portals. These represent market spaces that provide part solutions to the objectives of the I-KnowUMine project, however, they fall short on delivering the complete content contextualization server vision. The following sub-sections very briefly define these spaces.

Personalization tools aim to enable enterprises to adapt their interactions with their customers based on their individual needs. This includes targeting advertising, promoting products, personalizing content presentation on web channels, recommending documents, giving appropriate advice, target e-mailing and custom pricing. One of the most prominent technical and organizational challenges in this context is the provision of dynamic, personalized, collaborative interaction between user (employee, customer etc.) and supplier across all possible interaction channels.

Data mining is the technology used to discover non-obvious, potentially useful and previously unknown information from data sources. The potential of **web mining** is in the application of existing and new data mining algorithms to web data, which include server logs, as well as external data on customer, sales, and products etc. The business benefits that web mining affords to e-Business providers include personalization, collaborative filtering, enhanced customer support, product and service strategy definition, product marketing, online usability improvement and fraud detection.

Content management encompasses a set of processes and technologies, enabling the creation and packaging of content (documents, complex media, applets, components, etc.) as part of a dynamic and integrated Web-centric environment.

A **portal** can be defined as a personalizable, browser-based user interface to all appropriate corporate resources from any Internet-capable device. Being holistic business solutions, portals are platforms offering users efficient performance of various business processes across enterprise boundaries. To support efficient process performance, a portal must exhibit comprehensive knowledge concerning customer-specific business processes and provide all contextually relevant information and services to users (for administrators as well as end users) in a customizable manner.

3 Related Work

The use of Web usage mining for supporting Web personalization has recently attracted a lot of interest [AM01, MAB00]. In most of the cases, data mining techniques are used in order to extract useful patterns and rules concerning the users' navigational behavior. Site modifications are then made either by humans, or by a recommendation engine which helps the user navigate through a site. Some of the more integrated systems provide greater functionality, introducing the notion of adaptive Web sites and providing means for dynamically changing a site's structure. Few research projects integrate semantic web site structure knowledge with usage knowledge within the personalization process. An extensive overview of the most representative systems can be found in [EV03]. In this paper we focus on the integration of semantic, structure and usage knowledge to dynamically modify a Web site.

Coenen et al. [CS+00] proposed a framework for self-adaptive Web sites, taking into account the site structure but not the site usage. They underline the distinction between strategic changes, referring to the adaptations that have important influence on the original site structure, and tactical changes, referring to the adaptations that leave the site structure unaffected. The proposed approach is based on the fact that the methods used in Web usage mining produce recommendations including links that don't exist in the original site structure, resulting in the violation of the beliefs of the site designer and the possibility of the visitor getting lost following conceptual but not active links. Therefore, they suggest that any strategic adaptations based on the discovery of frequent item sets, sequences and clusters, should be made offline and the site structure should be revised. On the other hand, as far as the tactical adaptations are concerned, an algorithm for making online recommendations leaving the site structure unaffected is proposed.

Perkowitz et al. [PE98, PE99, PE00] were the first to refer to the notion of adaptive Web sites, defining them to be sites that semi-automatically improve their organization and presentation by learning from visitor access patterns [PE97]. The system, they proposed, semi-automatically modifies a Web site allowing only non-destructive transformations. Therefore, nothing is deleted or altered, instead new index pages containing collections of links to related but currently unlinked pages are added to the Web site. They proposed PageGather, an algorithm that uses a clustering methodology to discover Web pages visited together and to place them in the same group. More recently [PEt00], they proposed IndexFinder, which fuses statistical and logical information to synthesize index pages. In this latter work, they formalize the problem of index page synthesis as a conceptual clustering problem and try to discover coherent and cohesive link sets which can be presented to a human Webmaster as candidate index pages. In the case of IndexFinder information is derived from the site's structure and the page content. Therefore, IndexFinder combines the statistical patterns gleaned form the log file with logical descriptions of the contents of each Web page in order to create index pages.

The most integrated personalization system is WebPersonalizer, proposed by Mobasher et al. [MCS99, MCS00]. WebPersonalizer, the evolution of the WebSIFT prototype system [CTS99, SC+00], provides a framework for mining Web log files to discover knowledge for the provision of recommendations to current users based on their browsing similarities with previous users. It relies solely on anonymous usage data provided by logs and the hypertext structure of a site. After data gathering and pre-processing (converting the usage, content and structure information contained in the various data sources into various data abstractions) [CMS99], data mining techniques such as association rules, sequential pattern discovery, clustering and classification are applied in order to discover interesting usage patterns. The results are aggregated usage profiles. The recommendation engine matches each user's activity against these profiles and provides him/her with a list of recommended hypertext links.

This framework was extended in a more recent work [MD+00, MDa+00, MD+02] to incorporate content profiles into the recommendation process as a way to enhance the effectiveness of personalization actions. The content profiles are further enriched with the use of an Ontology [DM02]. Usage profiles and content profiles are represented as weighted collections of page view records. The content profiles represent different ways in which pages with partly similar content may be grouped together. The overall goal is to create a uniform representation for both content and usage profiles in order to integrate them more easily. The system is divided into two modules, the offline, which is comprised of data preparation and specific Web mining tasks, and the online component, which is a real-time recommendation engine.

The system presented in the paper is most similar in nature to the work carried out in the WebPersonlizer system. The key distinctions of the system proposed, is the method for generating the semantics used to describe the web pages, the use of sequence-pattern based recommendation systems, the use of multiple recommendation systems, the tighter integration with a content management system and the conformance to standards such as PMML and SOAP.

4 System Architecture

Based on the observations in Section 2 the system architecture developed, combines features derived from the following areas: Personalization, Web Mining, Content Management and Portals. The system modules are classified into four main layers: the Content Management Layer, Web Mining Layer, Knowledge Management Layer, and Interaction Layer.

The Content Management Layer incorporates the Content Management Module, the Taxonomy Management Module and the Content Classification Module. The main functionality implemented in this layer are the support for consistent authoring and storage of the content of the web site, its enrichment with semantic information, produced automatically or corrected/provided by a domain expert, support for creating/importing Taxonomies and support for administrative functions such as workflow and user management.

The Web Mining Layer consists of modules for enhancing web log files with semantic information (C-logs), which are used as input to the Web Mining Module, loading the data into a Webhouse and the mining modules. The knowledge generated by the mining modules is represented in PMML [DMG02] and stored within a knowledge base.

The Knowledge Management Layer is responsible for managing the knowledge generated by the Web Mining layer and includes its deployment through various recommendation engines (Recommendation Module).

Apart from these three general layers there is also an Interaction Layer, which includes the Publishing Module and the web server, which will present the corresponding personalized page to every user, by combining possibly "fixed" parts of the web page with parts where the personalized information should be presented.

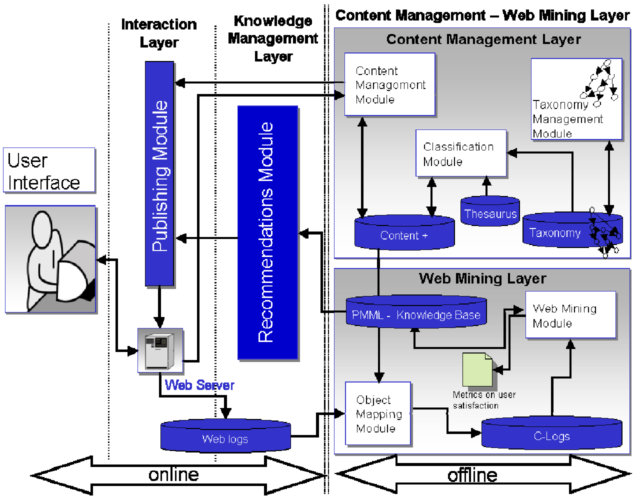

Figure 1: The IKUM system architecture

The layers and modules depicted in the Architecture of Figure 1 may be divided into an on-line and an off-line part. Functions performed off-line include those supported by the Content Management and the Web Mining Layer, while the module which creates the recommendations and the mechanism that publishes the personalized page for the visitors to the web site belong obviously to the on-line part of the IKUM system.

In order to describe how the modules, that implement the fundamental functionality of the system interact, we present a typical scenario of a user visiting a site based on the IKUM system.

The user requests a page from the Web site that uses the IKUM system. This request is received by the web server and passed on to the Content Management Layer. At the same time the web server stores information about the request within the Web server log file. The Content Management Module sends to the Publishing Mechanism the "fixed" part of the requested content, i.e. those sections of the page that do not contain personalized recommendations and should appear the same for all visitors. At the same time, through the combination of the visitors' current clickstream and the knowledge stored in the knowledge base, either using a user profile for the user or using group behavioral knowledge, generated by the web mining modules, the Recommendations Module generates recommendations and sends them to the Publishing Mechanism. The Publishing Module combines the two inputs ("fixed" content + recommendations) and renders the personalized Web page via the web server to the end-user. This process is performed online.

The knowledge base that provides input to the Knowledge Management Layer in order to produce recommendations, is created according to the following offline procedure: The content of the Web site is stored in the Content Repository (Content+) in the form of objects containing the content and related/associated metadata. The metadata may be added manually by a domain expert or by the use of (semi-) automated classification techniques provided by the Classification Module. Through the use of a domain-specific taxonomy and a thesaurus, the Classification module extracts keywords using text mining techniques and identifies taxonomy categories that characterize every content object that is stored in the Content+ repository. The domain-specific taxonomy is created and administered through the Taxonomy Management Module. The content classification is performed once for every content object and should be repeated only if new content is added, or old content is modified. We describe the content classification method in greater detail in Section 4.1.

Another thing should be taken into consideration: In many Web sites, especially in portal sites, a Web page (portal URI) consists of several different content objects (content URIs). Since the Content+ database contains content URIs, in order to perform Web usage mining, and enhance the usage knowledge with content knowledge stored as metadata properties of the several content objects, a mapping between the portal and the related content URIs is needed. This mapping is performed by the Object Mapping Module that takes input from the Content+ repository and web logs creating an extended version of the Web logs, referred to as C-Logs. The C-Logs are then used by the Web Mining Module on the Web Mining Layer and are processed in order to extract patterns, which are stored in the Knowledge Base, and metrics on user satisfaction.

4.1 Content Classification

When a personalization system relies solely on usage-based patters, information conceptually related to what is finally recommended will not be considered. Additionally in dynamic web environments, specific URIs may not be valid for a long period of time but accesses to them still provide useful information at the level of the content category of interest to a user. Using semantic annotation of the content, the final set of links that are proposed to the end user is expanded containing URIs and general content categories that would not have been recommended to the user otherwise. This is our motivation in designing and including in the general architecture a special module for the classification of the Web site's content into categories belonging to a domain-specific taxonomy. After this procedure, the content is enhanced with semantic information, which is then stored along with behavioral data (clickstreams) in the C-Logs. The process of classifying web site content is as follows. The web content must be parsed to remove all presentation/structure-related tags (Content Parsing). The text must then be converted into a more structured representation (Keyword Extraction) that can be used to classify the document into the various content categories (Keyword Category Mapping) as defined in a domain specific taxonomy. Each of these stages in the process are described in greater detail in the following subsections.

Keyword Extraction

After isolating Web content from structure data (HTML and XML tags, meta-tags etc.), by processing all the content URIs that constitutes the Web site and contained in the Content Repository, a text-mining algorithm is employed in order to extract keywords that characterize each Web page. This procedure can be divided in two sub-stages: (a) document indexing, where content-bearing terms are extracted from the Web page text, and (b) term weighting, where the indexed terms are given weights in order to choose the most important terms for characterizing the text.

Document Indexing: Many of the words in a document may not specifically describe its content. We refer to these words as non-significant words and remove them from consideration when generating the document indexing. The decision of which terms are significant and which ones are not may be defined using term frequency, by using a stop-words list or by some metric that is a combination of both of these, or indeed other measures of term importanceIn our architecture, we use a stop list in order to remove the non-content bearing words.

Term Weighting: Term weighting, extensively used in the vector space model (also referred to as bag-of-words) for document clustering, may be done using several methods, such as raw term frequency, and tf*idf. In the IKUM system, the most representative words that describe a Web document are selected using a combination of terms extracted using:

1) raw term frequency of the Web page
2) raw term frequency of a selected fraction of the reference to Web pages that are pointed to by this page (outlinks)
3) raw term frequency of a selected fraction of the most important Web pages that point to this page (inlinks)

The first method is straightforward, since the most frequent words extracted from the text are included in the proposed keyword set. In the second method, all the links that are contained in the content object under consideration are visited and parsed in order to extract frequent terms. As for the third method, a Web crawler or explicit site structure data is used in order to find Web pages that have links to the content object under consideration. When the link to this page is found, it is parsed along with 100 characters before and after the link, in order to extract keywords. The second and third method enhance the keyword extraction process, since they are based on the assumption that the characterization of the content that other people (content authors) give to the content object may provide more valuable information than the content in the object itself. Related efforts capitalizing on link semantics can be found in [CDG99] and [HNV02].

Keyword-Category Mapping

After the aforementioned process, the most highly ranked words that are extracted using the three methods are selected as representatives of its content. However, since all Web documents should be uniformly characterized by a limited, domain specific vocabulary, the keywords that were extracted in the previous stage should be mapped to the concepts defined in the taxonomy. This mapping is performed by using a thesaurus and a domain-specific taxonomy that has been defined in a machine-readable format (RDF) using the Taxonomy Management Module. If the keyword belongs to the taxonomy, then it is included as it is. Otherwise, the system finds the “closest” category word to the keyword through the mechanisms provided by the thesaurus (WordNet [WN]). For more details on mapping between keywords and categories see [HNV02]. It should be noted that in the case that the extracted keywords are in a language other than English, these words are translated to English before further processing using a dictionary. At the end of this stage, each content URI is characterized by a set of categories that are part of this taxonomy.

4.2 Sequence Tree Based Recommendation

Sequence patterns generated by Sequence pattern discovery algorithms such as Capri [BB+00] provide insights into navigational behavior of visitors to a web site. Matching these patterns to current behavior provide the basis for recommendation generation with the confidence measure associated with the matching sequence rule providing the measure of confidence in the recommendations. Capri generates PMML as well as a more compact, tree-based representation of the discovered sequences. As both these knowledge representations have associated XML representations, one representation can be easily transformed into the other using transformation languages such as XSLT. The sequence based recommendation engine used by the IKUM system takes as input sequence patterns represented in the tree format. However, it is worth stressing on the fact that as a PMML document representing sequences generated by a different sequence discovery algorithm can be easily transformed into this representation, thus this input format by no means restricts the use of Capri for generating the knowledge.

The recommendation engine takes as input the current user clickstream and matches all sub-sequences contained within the clickstream to the sequences contained in the sequence tree, recommending the ‘n’ content objects that have the highest confidence values associated with them, given the set of matching sub-sequences.

In addition to the sequence tree, two user-defined parameters bias the recommendation generation by associating a weighting with each sub-sequence matched, based on the recency of the matching sub-sequence within the clickstream and the length of the sub-sequence matched. These parameters take the form of discrete functions, $f(s.l)$ and $f(s.r)$, that map to the interval [0,1], the length and recency of the sub-sequences respectively. A function, $f(s.r)$, that assigns weights to a sub-sequence that is directly proportional to the recency of the sub-sequence, would bias recommendations to those that are generated by sub-sequences that were more recent within the clickstream, allowing the recommendations to reflect the changing context of a user within the visit. A function, $f(s.l)$, that assigns a weight to a sub-sequence that is inversely proportional to its length, biases recommendation to those generated by shorter sub-sequences.

The recommendations generated are represented in an XML document referred to as a recommendation pack. An example recommendation pack is shown in the below.

4.3 Using Multiple Recommenders

The architecture of the IKUM system uses a message driven approach to engaging multiple recommendation engines. Each recommendation engine is implemented as a message driven bean as defined within the EJB2.0 specification. Each of the recommendation engines deployed generates a recommendation pack that consisting of recommendations and an associated perceived value (attribute “score” within the recommendation element) of the recommendations. The knowledge used in the recommendation pack generation depends on the underlying approach used by the recommendation engine and can range from content filtering that use the semantic tagging generated by the IKUM system as defined in Section 4.1, through to behavior based group profiles (based on Segmentation knowledge) or sequence tree based recommendation engines as outlined in Section 4.2. The IKUM system administrator may currently provide relative weightings to the recommendation engines to create a single consolidated recommendation pack for a web site visitor. More complex mediation strategies are an interesting, open research question.


```
<?xml version= “1.0” encoding= “UTF-8”?>
<recommendation-pack process-time= “551”>
<recommendation description= “description” score= “0.83”
title= “Vice President Sales” value=
“/people/management/vps.htm”>
    <engine name= “Binary Segmentation Advisor”>
    <engine name= “Precompiled Sequence Advisor”>
</recommendation>
<recommendation description= “description” score= “0.74”
title= “Chief Executive Officer” value=
“/people/management/ceo.htm”>
    <engine name= “Precompiled Sequence Advisor”>
</recommendation>
<recommendation description= “description” score= “0.73”
title= “Chief Technology Officer” value=
“/people/management/cto.htm”>
    <engine name= “Frequency Segmentation Advisor”>
<engine name= “Precompiled Sequence Advisor”>
</recommendation>
</recommendation-pack>
```


4.4 Integration with a Content Management System

The Content Management Layer consists of several components such as the Content Management Module, the Classification Module and the Taxonomy Management Module and plays a central role in the IKUM architecture. As the main component of this layer one of the key requirements of the Content Management Module is the provision of tools for managing and administrating web content. Further key requirements consist of the provision and support of functionalities and interfaces for seamless integration between the Content Management Module and the other components within the Content Management Layer as well as the Publishing Module, which belongs to the Interaction Layer. As described in section 4.1, the Content Classification Module uses a thesaurus and a taxonomy to classify the content into semantic categories. Therefore the Taxonomy Management Module provides tools for creating and managing taxonomies. Furthermore it offers the possibility of importing/exporting taxonomies in a standard format allowing the exchange of already available taxonomies, which could be used for certain domains. Within the IKUM system this standard is RDF (Resource Description Framework). The domain expert is able to modify the imported taxonomy in the same way as he would modify a taxonomy created from scratch using the GUI provided by the Taxonomy Management Module.

4.5 Conformance to Standards

The IKUM system is architected with ease-of-integration in mind. As a result, the recommendation engines use XML/PMML input formats. The conformance to PMML enables the use of any data mining vendors tools that support the PMML standard, as the knowledge generation engine.

Furthermore, the taxonomy used can be imported from RDF and hence the user does not have to use the Taxonomy Management module of the content management layer, specifically, to develop the taxonomy used by the content classification module.

The content management module provides a SOAP interface, as does the recommendation module, enabling these modules to be used by any content management or alternative service that controls the interaction with the user.

5. Conclusions and Future Work

The paper described a system being built as part of the IKnowUMine project. The objective is to build a system that incorporates semantics as well as the navigational behavior of users of portals to provide a personalized service. The observation that the context of the user, when using a portal, can vary from one visit to the next, or indeed within the same visit, requires the development of flexible mechanisms for personalization that can adapt to the changing context of the user. Keeping this in mind, an architecture for a system that closely integrates web mining, personalization, content management and portal functionality to deliver personalized content is presented.

The innovations of the presented system with respect to automated content tagging, use of sequence knowledge for recommendation generation, support for multiple recommendation engines, close integration with a content management system and the conformance to industry standards such as EJB2.0, PMML and SOAP are presented in greater detail. Initial tests on a Greek construction portal site show promise for the system while highlighting difficulties related to the availability of robust multi-lingual methods for classifying content.

Future work will focus on the evaluation of the system in controlled tests to measure the lift in user experience using typical database marketing techniques. Additionally new research areas highlighted by the work, to date, include the lack of robust evaluation methodologies for recommendation engines and mediation strategies when using multiple recommendation engines. The authors are actively researching these research topics.

References

[AM01] S. S. Anand, B. Mobasher (ed.). Working Noted of the IJCAI-01 Workshop on Intelligent Techniques in Web Personlisation, August, 2001.

[BB+00] A.G. Büchner, M. Baumgarten, S.S. Anand, M.D. Mulvenna, J.G. Hughes. Navigation Pattern Discovery from Internet Data, B. Masand, M. Spiliopoulou (eds.) Advances in Web Usage Analysis and User Profiling, Lecturer Notes in Computer Science, Springer-Verlag, 2000.

[CDG99] S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, Mining the Link Structure of the World Wide Web, IEEE Computer (1999) Vol.32 No.6

[CMS99] Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, Data preparation for mining world wide Web browsing patterns, Knowledge and Information Systems, February 1999/Vol.1, No. 1

[CS+00] Filip Coenen, Gilbert Swinnen, Koen Vanhoof, Geert Wets, A Framework for Self Adaptive Websites: Tactical versus Strategic Changes, in proceedings of WEBKDD'2000 "Web Mining for E-Commerce – Challenges and Opportunities", 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2000, Boston, MA

[CTS99] Robert Cooley, Pang-Ning Tan, Jaideep Srivastava, WebSIFT: The Web Site Information Filter System, in Proceedings of the Web Usage Analysis and User Profiling Workshop (WEBKDD'99), August 1999

[DM02] H. Dai and B. Mobasher. Using Ontologies to Discover Domain-Level Web Usage Profiles. In proceedings of the Second Workshop on Semantic Web Mining, at the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02), Helsinki, Finland, August 2002.

[DMG02] The PMML2.0 Specification, www.dmg.org.

[EV03] Magdalini Eirinaki, Michalis Vazirgiannis, Web Mining for Web Personalization, survey paper to appear in ACM Transactions on Internet Technology, vol.3 no.1

[HNV02] M. Halkidi, B. Nguyen, I. Varlamis, M. Vazirgiannis, THESUS: Organizing Web Document Collections Based on Semantics and Clustering, Technical Report (2002), ftp://ftp.inria.fr/INRIA/Projects/verso/VersoReport-214.ps.gz

[MCS99] Bamshad Mobasher, Robert Cooley, Jaideep Srivastava, Creating Adaptive Web Sites Through Usage-Based Clustering of URLs, in Proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99), November 1999

[MCS00] Bamshad Mobasher, Robert Cooley, Jaideep Srivastava, Automatic Personalization Based on Web Usage Mining, Communications of the ACM, August 2000/Vol. 43, No. 8, pp. 142-151

[MD+00] B. Mobasher, H. Dai, T. Luo, Y. Sung, J. Zhu, Discovery of Aggregate Usage Profiles for Web Personalization, in Proceedings of the Web Mining for E-Commerce Workshop (WebKDD'2000), August 2000, Boston.

[MDa+00] B. Mobasher, H. Dai, T. Luo, Y. Sung, J. Zhu, Integrating Web Usage and Content Mining for More Effective Personalization, in Proceedings of the International Conference on E-Commerce and Web Technologies (ECWeb2000), September 2000, Greenwich, UK

[MD+02] B. Mobasher, H. Dai, T. Luo, M. Nakagawa. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. In Data Mining and Knowledge Discovery, Kluwer Publishing, Vol. 6, No. 1, pp. 61-82, January 2002.

[MAB00] M.D. Mulvenna, S. S. Anand and A. Büchner (ed.). Special Issue on Personalization, Communications of the ACM, August, 2000

[PE97] Mike Perkowitz, Oren Etzioni, Adaptive Web Sites: An AI Challenge, in Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, Nagoya, Japan, 1997

[PE98] Mike Perkowitz, Oren Etzioni, Adaptive Web Sites: Automatically Synthesizing Web Pages, in Proceedings of the 15th National Conf. on Artificial Intelligence, Madison, Wisconsin, July 1998

[PE99] Mike Perkowitz, Oren Etzioni, Adaptive Web Sites: Conceptual Cluster Mining, in Sixteenth International Joint Conference on Articial Intelligence (IJCAI99), Stockholm, Sweden, 1999

[PE00] Mike Perkowitz, Oren Etzioni, Towards Adaptive Web Sites: Conceptual Framework and Case Study, in Proceedings of WWW8, 1999

[PEt00] Mike Perkowitz, Oren Etzioni, Adaptive Web Sites, Communications of the ACM, August 2000/Vol. 43, No. 8, pp. 152-158

[SC+00] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, SIGKDD Explorations, January 2000/Vol. 1, Issue 2, pp. 12-23

[WN] WordNet, A lexical database for the English language, http://www.cogsci.princeton.edu/~wn/