Functionalities of a Content Management System specialized for Digital Library Applications [*]

Giuseppe Amato, Claudio Gennaro, Fausto Rabitti, Pasquale Savino
{giuseppe.amato, claudio.gennaro, fausto.rabitti, pasquale.savino}@isti.cnr.it

ISTI - CNR, Pisa, Italy

Abstract. Given the lack of standard building component, in several cases digital library applications are built from scratch using ad-hoc approaches to implement all required components. On the other hand, our claim is that the development of ad-hoc software modules for each new digital library is not convenient. It is necessary to define and design standard software components in order to support the design and effective implementation of Digital Library Applications. Specifically, in this paper we will focus on the definition of a content management system that offers typical functionalities required by digital library applications.

1 Introduction

The Digital Library (DL) technology emerged during mid 90ties as a combination of different technological results, mainly in the area of Database Management and Information Retrieval, and as an application of these technologies to the management of libraries. The term digital library has been subject of a certain confusion concerning its interpretation and use. Sometime it has been used to refer the accessible digital content of a library. Sometime it has been used to indicate a specific application built to make accessible a specific digital content. Sometime it has been related to a set of software tools that can be used to

[*] This work was partially supported by the ECD project (Extended Content Delivery) [9], funded by the Italian government, by the VICE project (Virtual Communities for Education), also funded by the Italian government, and by DELOS NoE, funded by the European Commission under FP6 (Sixth Framework Programme). We would like to thank Paolo Bolettieri, Franca Debole, Fabrizio Falchi, Francesco Furfari, and Bertrand Le Saux for their valuable contribution to the MILOS implementation.

build a digital library application, in order to access a published digital content. However, we may observe that in the digital library field we may use the same distinction that is made in the database systems field between the *database*, which is the set of data that should be managed, *the database management system*, which is the general purpose software specialized in data management, and *the database application*, which is the application which allows users to access data. Thus, in the following we will distinguish between the *Digital Library*, which is the set of documents – or, more in general, the set of digital objects – that are managed, the *Content Management System (CMS)*, which is a general purpose software module that can be customized in order to provide different functionality in the management of digital objects, and the *Digital Library Application (DLA)*, which is a specific application to access the digital objects belonging to a particular Digital Library.

A first category of DLA, which is nowadays well established, consists in the pure management of the library catalogs in electronic form, with support for the search of the library objects and the identification of their location in the library: no support is provided to the management of digital objects and to their access by content. More recently, new DLs emerged, where the "goods" of the library are stored in digital form and can be searched by content, accessed, and manipulated online. This evolution is quite slow, mainly for economic reasons: it is becoming widely used by publishers that are moving towards an electronic publication, and it is going to be adopted also in the cultural heritage sector, where many museums are offering access to part of their collections online. Usually a DLA offers most of the services provided by a traditional library, which mainly consist of three aspects: (i) library creation, which includes all techniques needed to build the library and index the documents, (ii) library exploration, which includes all techniques used to enable users to select the relevant documents and to visualize them, (iii) library management, which includes all techniques used to support the management of the library in terms of access control, security management, billing, etc.

In the following we first analyze the current status of the digital library technology (Section 2). We discuss the future perspective of the digital library field (Section 3). Accordingly, we define the functionality that a Multimedia Content Management System (MCMS) should provide to support digital library applications (Section 4) and we outline the MILOS system (Section 5), a Multimedia Content Management System specialized for digital library applications. Finally we present several significant DL applications that were implemented by using MILOS, and we show the advantages of the proposed approach in building these specific DL applications, resulting in the simplicity of the implementation and in significant system performance (Section 6). Section 7 concludes.

2 Current status of digital library applications

Regrettably, often the Digital Library Applications are monolithic software modules built for a single Digital Library. Existing Digital Library Applications just require documents to be inserted and metadata to be generated before being ready to be operative for document searching and retrieving. Furthermore, the digital library technology is today limited to manage specific types of digital objects and specific metadata description models. This implies that existing DL Applications can be hardly adapted to different application environments and to different metadata description models. Indeed, many DLAs were built having in mind a specific application and, in many cases, a specific document collection. Thus, the result is an ad-hoc solution where all components of the DLA (the data repository, the metadata manager, the search and retrieval components, etc.) are specific to a given application and cannot be easily used in other environments. Digital library applications often offer predefined workflows, metadata schemes, and document formats, that cannot be changed to be adapted to specific application scenarios and end-user requirements and no customizations can be performed for adapting the user interface to the specific scenarios. Many of these systems guarantee inter-operability with other systems, by adopting standard protocols such as OAI, or Z39.50. However, their inter-operability is limited to the exchange (import/export) of data/metadata. In fact, there is no chance of reusing software components, to integrate functionality of other DLAs, or to use digital contents (documents and metadata) compliant to other standards. This is mainly due to the lack of basic building components, tailored to DL application design, which are standard and general purpose.

In this paper we propose an approach similar to that applied in the field of traditional database applications. In fact, database applications are generally built relying on a Database Management System (DBMS), a general purpose software module that offers all functions needed to build many different database applications. We intend to demonstrate that the same can be done in the DL field: it is possible to build a general purpose Multimedia Content Management System (MCMS) which offers functionalities specialized for DL applications (See

Figure 2). Different DL applications can be built on top of such an MCMS, each supporting the management of documents of any data type, described by using different metadata description models, searchable in many different modes. This MCMS should be able to manage not only formatted data, like in databases, but also textual data, using Information Retrieval technology, semi-structured data, typically in XML, mixed-mode data, like structured presentations, and multimedia data, like images and audio/video.

3 Future perspective for digital libraries

Many researchers think that the DL technology could be applied, in the future, well beyond the restricted scope of today applications. To estimate the potential of DL technology, in Figure 1 [16] an information space is considered, with one dimension representing the level in which users and tasks are predefined and known in advance, and the other dimension representing the level in which the data has a known and well defined structure. In this information space, it is possible to distinguish the characteristics of Digital Library applications from typical Web and database applications: Typical Web search engines assume very little about users, tasks, and the data they deal with. Consequently, they occupy a relatively small part of the space. On the other hand, database applications have strong assumptions about users, tasks, and data. The typical interaction with database management systems is usually limited to a few transaction types to be performed very efficiently and data must be specified in advance, using relational schemas. Hence, these applications occupy a small part of the space as well. The rest of the space is considered as potentially belonging to Digital Library applications. In this part of the space, information systems attempt to exploit knowledge about the users, tasks, and domain to improve access, but retain the flexibility that is characteristic of Web-based applications. This is also the potential application area of Enterprise Content Management applications. We believe that only a system that can couple advance Digital Library functionalities with advanced Content Management functionalities could effectively “occupy” the portion of the information space between Web and databases, and then receive as much attention in the commercial world as database and Web applications.

There is a broad consensus that only 10 to 15% of the corporate information is today managed, quite effectively, by database management systems. The rest of this information, like office documents, legal papers, technical references, reg-

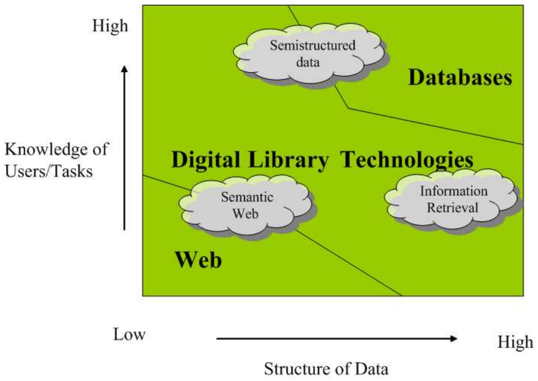


Fig. 1. Scope of Digital Library technology

ulations, marketing material, customer relationship information, etc. should be in the scope of the Content Management System (CMS). In order to effectively manage this kind of data, a CMS should be able to manage not only formatted data, like in databases, but also textual data, using Information Retrieval technology, semi-structured data, typically in XML (W3C standard), mixed-mode data, like structured presentations, and multimedia data, like image and audio/video.

These kinds of requirements are especially essential for the next generation Digital Library Applications (DLA). Accessing multimedia data is becoming more and more essential for digital libraries since numeric data and text documents account for only 0.003% of the total amount of digital information produced today [16]. Today, there is an extensive research work towards the extension of DL technology to support the management of other media: images, audio, video, etc. [18] [2]. At the same time, the first multimedia DLAs are becoming commercially available [6]. In general, the services offered by a multimedia DLA are similar to the services offered by a textual DLA. However, we must consider that for multimedia documents the problem of indexing (and metadata generation) is extremely more complex than for other types of data, due to the fact that multimedia documents have a complex structure and that it is difficult to automatically extract information from them. Moreover, multimedia documents may also be affected by poor or limited quality, which makes the problem of indexing even more complex and subject to errors.

In the following section we will give more technical motivation toward the design of a content management system for digital libraries.

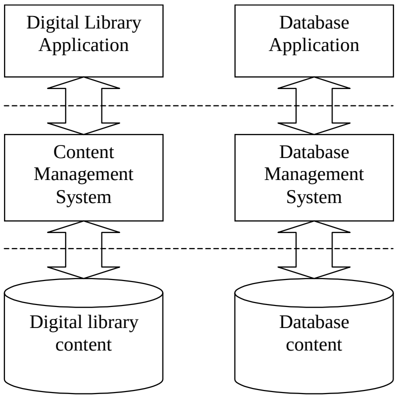

Digital Library Application
Database Application
Content Management System
Database Management System
Digital library content
Database content

Fig. 2. Relationships among layers of DLs and databases

4 Multimedia Content Management Systems for Digital Libraries

Digital library applications are document intensive applications where possibly heterogeneous documents and their metadata have to be managed efficiently and effectively. We believe that the main functionalities required by DL applications can be embedded in a general purpose Multimedia Content Management System (MCMS).

If we make an analogy to the database field, the MCMS is the analogous of a Database Management System (DBMS) in the domain of document intensive applications, as for instance digital libraries, see Figure 2. DBMSs are software tools specialized to support database applications like banking systems, billing systems, etc. MCMSs are software tools specialized to support applications where documents, embodied in different digital media, and their metadata are efficiently and effectively handled.

Note that many digital library projects have mainly aimed at defining general purpose services that should be provided by a digital library. Typical services are for example, repository services, collection services, authentication services, etc. However these definitions were limited to a very high abstraction level. Very little effort has been devoted to define and investigate specific solutions for efficiently realizing these services, or to investigate the existence of technologies, proposed in other fields, that can be used to cope with these issues. In addition, sometime these digital library services were defined as a consequence of requirements of specific digital library applications, while their generality, innovation, and real importance to generic digital libraries was never proved. For example, while it is clear that any digital library should have a repository service that manages documents and/or metadata storage. It is not also obvious that all digital library applications really need complex custom user authentication/authorization services, which can be anyhow *obtained* relying upon services offered either by employed operating systems, application servers, or even database systems.

The minimal requirements of a Multimedia Content Management System are *Flexibility*, in structuring both multimedia documents and their metadata, *Scalability*, and *efficiency*. *Flexibility* is required both at the level of management of basic multimedia documents and at the level of management of their metadata. The flexibility required in representing and accessing metadata can be obtained by adopting XML as standard for specifying any metadata (for example MPEG-7 [3] can be used for multimedia objects, or SCORM [8] for e-Learning objects). Requirements of *scalability* and *efficiency* are essential for the deployment of real systems able to satisfy the operational requirements of a large community of users over a huge amount of multimedia information.

Identifying the characteristics of the *operational model* is essential in order to design a system with suitable performance for the operations considered most critical by the user in the desired application settings. These operational requirements result quite different from operational requirements of typical database systems:

- It should be possible to insert new data and metadata without prior intervention of a database administrator. In databases, no data/metadata can be inserted if its schema has not been previously defined.
- It should be possible to define different metadata corresponding to the same multimedia basic object(s). A metadata description can span over many multimedia objects.
- The access to the content should be based on all its components, that is, the query formulation should combine conditions on formatted data/metadata attributes, textual and multimedia components (via associated features). The query results should be ranked taking into account all components in the answer.
- Searching – the most common operation concurrently requested by many users – must be very efficient. Since it is very convenient to encode metadata in XML, the system must efficiently support standard XML query languages such XPath [14] and XQuery [15], with extensions for querying text and

multimedia components. Note that XML interfaces to relational databases systems are quite inefficient on complex XML objects since they imply a large number of join operations.

- Since update operations are quite rare compared to search operations, it is not necessary to enforce a database-like transactional mechanism, that is quite powerful but at the same unnecessarily rigid. Rather, a concurrency control mechanism to support the editing process of complex multimedia objects and associated metadata is necessary, based on some sort of check-out/check-in protocol.
- The system must be efficient in supporting the continuous insertion of new multimedia objects and associated metadata, i.e. the bulk import. Consider, for example, publishing and broadcasting applications.

We believe that the basic functionalities of a MCMS are related to the issues of *storage and preservation* of digital documents, their *efficient and effective retrieval*, and their *efficient and effective management*. These functionalities should be guaranteed by appropriate management of documents and related metadata, according to the following prerequisites:

1. capability of managing different documents embodied in different media and stored with different strategies;
2. capability of describing documents by way of arbitrary, and possibly heterogeneous, metadata;
3. capability of providing DL applications with custom/personalized views on the metadata schema actually handled.

Point 1) requires that no assumption should be taken on the types of media and encoding used to represent documents, and especially on the specific strategy used to store them. This allows applications to be unaware of the technical details related to multimedia document management. For instance, textual documents can be stored in the file system and served to the users using a normal web server. However, video documents might need to be maintained in a video server that uses various storage devices, as for example digital tapes stored in silos, optical disks, and/or temporary storage space on arrays of hard disks [17]. In addition, video documents might be served exploiting specific real-time continuous media streaming strategies to avoid hiccups during playback. The DL application should be designed independently of these issues, which should be managed transparently by the MCMS. For instance, changes on the storage strategies should be possible without changing the DL application software.

Point 2) states that a content management system should be able to deal with arbitrary metadata. This is required by the fact that different DL applications, according to their specific requirements, might need to use different metadata. Consider that existing archiving organizations have already their own metadata schemas, and hardly want to modify them to be compatible with a specific system. Therefore, a DL management system should be able to support any metadata schema without requiring metadata translation or restrictions on the functionality offered. There are also cases where the same application need to deal with different metadata models at the same time. These different metadata might be needed because the documents have redundant descriptions in terms of different metadata, or because the DL application is dealing with document collection described with heterogeneous metadata. The last case might occur, for instance, in case of integration/merging of archives managed by different organization.

Point 3) makes it possible that the metadata schema seen by the DL application is different from the metadata schemas actually stored in the repository of the content management system. Suppose that an application was built to deal just with a specific metadata schema. The MCMS should be able to serve requests of such an application even if metadata stored in the repository comply to different schemas. Metadata schema independence can be obtained by exploiting techniques of schema mapping. This feature is especially useful in case of heterogeneous metadata available at the same time in the repository: the DL application will refer to just one metadata schema, relying on the multiple schema mapping performed on the fly by the MCMS. In addition, this feature allows different DL application, which require different metadata schemas, to share the same MCMS transparently.

5 MILOS: an example of content management system for multimedia digital libraries

An example of content management system for multimedia digital libraries, which satisfies the requirements and offers the functionalities discussed in previous section, is MILOS (Multimedia dIgital Library Object Server). MILOS [12,10] has been developed by using the Web Service technology, which in many cases (e.g. .NET, EJB, CORBA, etc.) already provides very complex support for “standard” operations such as authentication, authorization management,

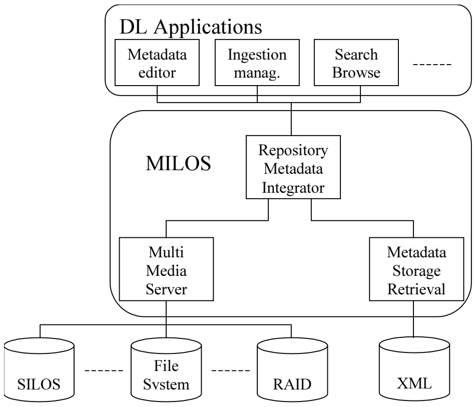


Fig. 3. General Architecture of MILOS

encryption, replication, distribution, load balancing, etc. Thus, we do not further elaborate on these topics, but we will mainly concentrate on the aspects discussed above.

MILOS is composed of three main components as depicted in Figure 3: the Metadata Storage and Retrieval (MSR) component, the Multi Media Server (MMS) component, and the Repository Metadata Integrator (RMI) component. All these components are implemented as Web Services and interact by using SOAP. The MSR manages the metadata of the DL. It relies on our technology for native XML databases, and offers the functionality illustrated at point 2) above. The MMS manages the multimedia documents used by the DL applications. MMS offers the functionality of point 1) above. The RMI implements the service logic of the repository providing developers of DL applications with a uniform and integrated way of accessing MMS and MRS. In addition, it supports the mapping of different metadata schemas as described at point 3) above. All these components were built choosing solutions able to guarantee the requirements of flexibility, scalability, and efficiency, as discussed in the next sections.

6 Designing and building digital library application with MILOS

In order to verify and demonstrate the flexibility and efficiency of MILOS in managing different heterogeneous DL applications, we took four data sets used by four different existing DLs and we built from scratch the corresponding DL application on top of MILOS. The data sets that we considered consist of documents and metadata of very different nature: the Reuters data set [4], the ACM Sigmod Record dataset [5], the DBLP data set [1], and the ECHO data set [11]. These data sets and the corresponding MILOS powered DL applications are described in next subsections.

The DL applications that we built use the same MILOS installation and all data sets were stored together. The functionality of MILOS allows individual applications to selectively access just data and metadata of their interest or to perform cross-library search. Each DL application consists of a specific search and browsing interface (built according to the data managed) and a bulk import tool. The search and browse interfaces were built as web applications using Java Server Pages (JSP). The bulk import tool was a simple java application. On average, the effort required to build each application from scratch was one week of work of a single skilled person. This, we believe, is really a little effort compared to the cost that would have been required to build from scratch a DL, without general purpose tools, or the cost that would have been required to translate and adapt the data and metadata to cope with the requirements and restrictions of an existing DL system.

Notice that we built from scratch the browse and retrieval interface. However, we are currently working at designing and developing tools for the automatic generation of the browsing and retrieval interface in correspondence of data and metadata managed by MILOS. This will contribute to a further reduction of the cost of building DL applications.

All applications resulted to be very efficient. We installed the system, the applications, and the data on a single computer equipped with a Pentium 1.8 GHz and 1 Gb of RAM, running Windows 2000 server. We have used JAX-RPC as SOAP application server to run MILOS. We asked 30 persons to use the DL at the same time from remote workstations, and to execute a predefined search intensive job. On average the response time of the system was below 1 second. Notice also that for more intensive use of the system, the underlying Web Service technologies offer plenty of solutions to guarantee scalability exploiting techniques of replication, load balancing, resource/connection pooling etc.

The screen shots of the DL applications can be seen in Figure 4.

6.1 Reuters data set

The Reuters data set [4] contains text news agencies and the corresponding metadata. There are two types of metadata: Reuters specific metadata including titles, authors, topic categories, and extended Dublin Core metadata.

The Reuters data set contains 810,000 news agencies (2.6 Gb) where text and metadata are both encoded in XML. We associated the full text index and the automatic topic classifier to the elements containing the body, the title, and the headline of the news. Other value indexes where associated with elements corresponding to frequently searched metadata, like locations, dates, countries.

The search interface allows the user to perform integrated text, category, and exact match search. The bulk import tool simply takes all XML file, corresponding to the news, contained in a specified directory, associates them with a unique URN, and inserts them into MILOS. The bulk import of the Reuters data set (with the corresponding indexing of data and metadata) took 20 hours.

6.2 ACM Sigmod Record and DBLP data sets

Both the ACM Sigmod Record data-set [5] and the DBLP data-set [1] consists of metadata corresponding to the description of scientific publications in the computer science domain. The ACM Sigmod record is relatively small. It is composed of 46 XML files (1Mb), while the DBLP data-set is composed of just one large (187Mb) XML file. Their structure is completely different even if they contain information describing similar objects.

For these two datasets we built just one DL application from which both data are accessed. We exploited the mapping functionality of MILOS for having the requests of the application correctly translated for the two schemas. We associated a full text index to the elements containing the titles of the articles, and we associated other value indexes to other frequently searched elements, such as the authors, the dates, the years, etc.

The search and browse interface, allows users to search for articles by various combinations full text and exact/partial match of elements. In addition it allows users to browse results by navigating trough links (and implicitly submitting new queries to MILOS) related to the authors, journals, conferences, etc.

The bulk import tool simply takes the XML files corresponding to the metadata and directly inserts them in the system, after having associated them with unique URNs. The bulk import of these data sets, including their indexing, required about 5 hours.

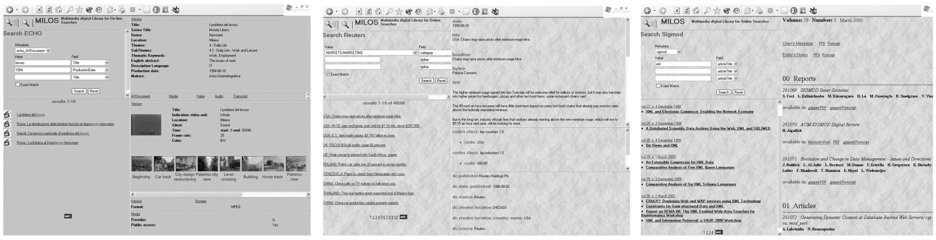
6.3 ECHO data set

The ECHO data set [11] includes historical audio/visual documents and the corresponding metadata. ECHO is a significant example of the capability of MILOS to support the management of arbitrary metadata schemas. The metadata model adopted in ECHO, based on IFLA/FRBR [7] model, is rather complex and strongly structured. It is used for representing the audio-visual content of the archive and includes among others, the description of videos in English and in the original language, specific metadata fields such as Title, Producer, year, etc., the boundaries of scenes detected (associated with a textual descriptions), the audio segmentation (distinguishing among noise, music, speech, etc.), the Speech Transcripts, and visual features for supporting similarity search on key-frames.

The collection is composed of about 8,000 documents for 50 hours of video described by 43,000 XML files (36 Mb). Each scene detected is associated with a JPEG encoded key frame for a total of 21GB of MPEG-1 and JPEG files. Full text indexes where associated to textual descriptive fields, similarity search index where associated with elements containing MPEG-7 image (key frames) features, and other value indexes where associated with frequently searched elements.

The search and retrieval interface allows users to find videos by combining full text, image similarity, and exact/partial match search. Users can browse among scenes, and corresponding metadata. We simply used the file system as a repository for the videos. However, as previously discussed, moving to more complex multimedia storage services is straightforward and transparent, from the system and application perspective. The original ECHO DL application, as resulted from the ECHO project [2], was built using a relational database, and translating all metadata in a relational schema. Even simple searches required several (up to 10 or more) seconds to be processed. With MILOS we had a dramatic improvement of performance, being able to serve requests in less than one second even with several users accessing the system.

The bulk import tool takes the videos and the key frame images, and inserts them in the system. Each one of these document is associated with an unique identifier, which is also added to the ECHO metadata to correctly refer the documents from the corresponding metadata. Then the updated metadata are also inserted in the system. The bulk import of data and metadata and their indexing required almost one hour.



Metadata: echo_AVDocument
Value: lavoro — Field: Title
Value: 1954 — Field: ProductionDate
Value: — Field: Title
☐ Exact Match

results 1-10

- I problemi del lavoro
- Roma: La distribuzione della befana fascista al dopolavoro ferroviario
- Napoli: Congresso nazionale di medicina del lavoro.
- Roma - La Befana al Dopolavoro ferroviario

Version
Title: I problemi del lavoro
Series Title: Mondo Libero
Genre: Newsreel
Location: Milano
Themes: 4 - Daily Life
SubThemes: 4.1 - Daily Life - Work and Leisure
Thematic Keywords: Work, Employment
English abstract: The issues of work
Description Language: IT
Production date: 1954-06-18
Makers: Astra Cinematografica

Version
Title: I problemi del lavoro
Indication video unit: Whole
Location: Milano
Silent: Sound
Time: start: 0 end: 92040
Frame rate: 25
Color: BW

Format: MPEG

Media
Provider: IL
Public access: Yes

Value: MARKETS/MARKETING — Field: category
Value: — Field: byline
Value: — Field: byline
☑ Exact Match

results 1-10 of 40506

- USA: Chains may raise prices after minimum wage hike.
- USA: NYSE says exchange seat sold for $1.16 mln, down $287,500.
- USA: U.S. farm trade surplus $1.741 billion in June.
- UK: FOCUS-BSkyB profits surge 66 percent.
- UK: Virgin group to expand into South Africa - paper.
- POLAND: Polish car sales rise 33 percent in seven months.
- VENEZUELA: Pepsi to vanish from Venezuelan mkt soon.
- CHINA: China calls on TV makers to halt price cuts.
- THAILAND: Thai rice traders greet expected end of Mexico ban.
- CHINA: China car production surplus prompts exports.

1 2 3 4 5 6 7 8 9 10

date:
1996-08-20

title:
USA: Chains may raise prices after minimum wage hike.

headline:
Chains may raise prices after minimum wage hike.

byline:
Patricia Commins

text:

The higher minimum wage signed into law Tuesday will be welcome relief for millions of workers, but it may also translate into higher prices for hamburgers, pizzas and other fast-food items, some restaurant chains said.

The 90-cent-an-hour increase will have little short-term impact on many fast-food chains that already pay workers rates above the federally mandated minimum.

But in the long run, industry officials fear that workers already earning above the new minimum wage, which will rise to $5.15 an hour next year, will be looking for more.

codes class: bip:countries:1.0

- code: USA

codes class: bip:industries:1.0

- code: I66100

dc.publisher.Reuters Holdings Plc

dc.date. published: 1996-08-20

dc.source.Reuters

dc.creator.location.CHICAGO

dc.creator.location. country. name: USA

dc.source.Reuters

Metadata: sigmod
Value: xml — Field: articleTitle
Value: — Field: articleTitle
Value: — Field: articleTitle
☐ Exact Match

vol 27 n. 4 December 1998
- XML and Electronic Commerce: Enabling the Network Economy

vol 28 n. 3 September 1999
- A Distributed Scientific Data Archive Using the Web, XML and SQL/MED

vol 28 n. 4 December 1999
- On Views and XML

vol 29 n. 1 March 2000
- An Extensible Compressor for XML Data
- Comparative Analysis of Five XML Query Languages

vol 29 n. 3 September 2000
- Comparative Analysis of Six XML Schema Languages

vol 30 n. 1 March 2001
- EIHA?!?: Deploying Web and WAP services using XML Technology
- Constraints for Semi-structured Data and XML
- Report on XEWA-00: The XML Enabled Wide-Area Searches for Bioinformatics Workshop
- XML and Information Retrieval: a SIGIR 2000 Workshop

1 2 3 4

Volume: 29 **Number:** 1 March 2000

00 Reports

291069 SIGMOD Sister Societies
S. Ceri L. Kalinichenko M. Kitsuregawa H. Lu M. Ozsoyoglu R. Snodgrass V. Via

available in: gzipped PDF gzipped Postscript

291070 ACM-SIGMOD Digital Review
H. Jagadish

available in: Microsoft Word PDF gzipped Postscript

291071 Evolution and Change in Data Management - Issues and Directions
J. Roddick L. Al-Jadir L. Bertossi M. Dumas F. Estrella H. Gregersen K. Hornsby Lufter F. Mandreoli T. Mannisto E. Mayol L. Wedemeijer

available in: gzipped PDF gzipped Postscript

01 Articles

291072 Generating Dynamic Content at Database-Backed Web Servers: cgi-vs. mod_perl
A. Labrinidis N. Roussopoulos

Fig. 4. The ECHO (Left), Reuters (Center), and Sigmod/DBLP (Right) retrieval interfaces implemented in MILOS

7 Conclusion

This paper proposed an approach to build Multimedia Content Management Systems for digital library applications. The solutions proposed can be used to obtain a system that is flexible in the management of documents with different types of content and descriptions, and that is efficient and scalable in the storage and content based retrieval of these documents. In particular, we described the approach adopted to support the management of different metadata descriptions of multimedia documents in the same repository. This goes towards the solution of the challenging problems of interoperability among different metadata descriptions. The proposed solution, based on the use of a mapping mechanism among the metadata fields of the different models, has been practically experimented by using the MILOS system to archive documents belonging to four different and heterogeneous collections which contain news agencies, scientific papers, and audio-video documentaries. The archiving of these documents was straightforward and it only required the creation of the mapping file and the development of the user interfaces to archive and to search the documents.

Evolutions of this activity are foreseen in several directions: on one side we are working to improve the retrieval capabilities of the Metadata Storage and Retrieval component; on the other side, we are working with partners of the ECD [9] project on the automatization of the mapping between different metadata schemas, by using thesaurus and cross-language vocabularies [13].

References

1. DBLP computer science bibliography. http://www.informatik.uni-trier.de/ ley/db/.

2. Echo: European CHronicles On-line. http://pc-erato2.iei.pi.cnr.it/echo/.
3. Motion picture experts group. http://mpeg.cselt.it.
4. Reuters corpus. http://about.reuters.com/researchandstandards/corpus/.
5. Sigmod record, xml edition. http://www.acm.org/sigs/sigmod/record/xml/.
6. Virage web site. http://www.virage.com/.
7. IFLA study on the functional requirements for bibliographic records, 1998. http://www.ifla.org/VII/s13/frbr/frbr.pdf.
8. Shareable content object reference model initiative (scorm), the xml cover pages, October 2001. http://xml.coverpages.org/scorm.html.
9. ECD - Enhanced Content Delivery, 2002. http://ecd.isti.cnr.it/.
10. Milos, 2004. http://milos.isti.cnr.it/.
11. G. Amato, D. Castelli, and S. Pisani. A metadata model for historical documentary films. In J. L. Borbinha and T. Baker, editors, *Proc. of the 4th European Conference ECDL*, pages 328–331. Springer, 2000.
12. G. Amato, C. Gennaro, F. Rabitti, and P. Savino. Milos: A Multimedia Content Management System for Digital Library Applications. In R. Heery and L. Lyon, editors, *ECDL 2004, Bath, UK, September 13-15, 2004*, volume 3232 of *LNCS*, pages 14–25, 2004.
13. D. Beneventano, S. Bergamaschi, S. Castano, V. D. Antonellis, A. Ferrara, F. Guerra, F. Mandreoli, G. C. Ornetti, and M. Vincini. Semantic integration and query optimization of heterogeneous data sources. In *OOIS Workshops*, pages 154–165, 2002.
14. W. W. W. Consortium. XML path language (XPath), version 1.0, W3C. Recommendation, November 1999.
15. W. W. W. Consortium. XQuery 1.0: An XML query language. W3C Working Draft, November 2002. http://www.w3.org/TR/xquery.
16. DELOS. Digital libraries: Future directions for a european research programme. Report, June 2001. http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf.
17. D. J. Gemmell, H. M. Vin, D. D. Kandlur, P. V. Rangan, and L. A. Rowe. Multimedia storage servers: A tutorial. *IEEE Computer*, 28(5):40–49, May 1995.
18. H. Wactlar. The informedia digital video library. http://www.informedia.cs.cmu.edu/.