



Al-Khwarizmi Engineering Journal, vol.1, no.1,pp 46-51, (2005)

A New Approach for Designing Multi Information Management System Using XML Technology

Dr. Suha Mohammed Hadi

Information Engineering Dept./Al-Khwarizmi Engineering College/UniversityOf Baghdad

Abstract

XML is being incorporated into the foundation of E-business data applications. This paper addresses the problem of the freeform information that stored in any organization and how XML with using this new approach will make the operation of the search very efficient and time consuming. This paper introduces new solution and methodology that has been developed to capture and manage such unstructured freeform information (multi information) depending on the use of XML schema technologies, neural network idea and object oriented relational database, in order to provide a practical solution for efficiently management multi freeform information system.

Keyword: Free Form Information, XML Technology, Data Modeling, Relational Data Base.

Introduction:

It is estimated that in most businesses over 80% of all information is in unstructured or freeform format. A huge proportion of corporate memory is therefore captured in freeform format. Capturing and managing this information in a computer system offers great potential for improving efficiency and resource utilization increasing innovation and competitiveness by retaining the corporate know-how and memory, helping gain deep insights and understanding into the business environment. This is a key component of the knowledge management discipline and the digital or knowledge economy [1].

The advent of the World Wide Web in the 1990s has resulted the publication

of millions freeform documents in digital form. Such free-form information is difficult to manage [2].

The information is diverse in nature, complex, and subject to frequent changes. Traditionally, forms are used to ensure that relevant information is captured by providing structures and checklists. Forms are especially important when the same type of information is captured repeatedly [3].

In the paper world, we often make copies of the documents so that we can put them in the folders required. This creates its own problems, as comments made on one copy will not be visible in the other copies. Follow-up documents also may not be filed in all the folders. Unfortunately such storage systems do

not support the information usage model. Documents may be classified into many categories and related to each other. In the computer world, the filing system should be virtualized where we have only one copy of the document but visible in as many categories as required.

Freeform information without structure is difficult to manage. And lack of structure dramatically reduces an organization's ability to leverage the value of such content. The simplest and most common form of information storage and management is to store documents as files in folders and use the directory structure as the means of navigation. With more advanced systems, the documents are stored in databases where they may be indexed to provide full-text search facilities [4].

The purpose of our proposed information management system is that to be useful for any organization in various purposes including training and reference, planning, design, innovation and decision-making processes. To meet these requirements we have to look beyond information access and full-text search facilities. Users of the information must also be able to gain insights and deep understanding on the subjects of interest. The facilities provided by this model must assist the users in discovering new information from the unified view of the information available.

1. XML Technology

The Extendable Markup Language (XML) is an open standard for defining data elements on business documents. It defines the structured information those elements contain. XML was originally developed as an application profile of Standard General Markup Language (SGML) to use over the Internet. But the ease of both writing applications that process XML document and creating XML

documents has made XML an instant success for a variety of other application domains, too.

In fact, one area where the benefits of XML have become immediately

apparent is data exchange and data modeling. There are clear benefits from using XML in these areas.

- First, the design of XML is formal and concise, so it's relatively easy to write structured data.
- Second, since XML documents are human-readable, a developer can figure out what the content means by simply inspecting the XML document.
- Third, momentum is building to create standardized XML protocols (also referred to as XML Schemas) for almost any type of business.

So there's an excellent chance that in the near future a majority of transactions will be carried out using XML as the underlying infrastructure protocol. This shift from binary protocols to XML-based protocols resembles the computing industry's move from mainframe-centric terminals and central processing units to the client/server paradigm of distributed computing. Like client/server before it, XML requires new ways of storing data. In the shift from mainframe-centric to client/server computing, a major side effect was the development of relational database servers. Similarly, in the case of the move to XML protocols, there is a noticeable trend toward not only exchanging XML data, but also storing it directly as XML data [5].

2. Designing the New Approach

Before the information can be managed it must be captured first. The bulk of freeform information available

today is unstructured. With XML technology, the structure form can be imposed on most kinds of information. This technology trend is to capture as much of such information as possible.

In order to Use conventional technology, for each form, we require a software program or a program unit to create the user interface, define the data types, validation rules and posting processes. Often, enumerated lists and lookup tables are also required to provide codified information for some fields.

By using this technology, the information has to be stored in relational database tables, which will run into thousands as the information in each complex form will have to be posted into several normalized relational tables. All this must be designed with some considerations that may influence the development of such data capture systems.

Our new approach schema forms is designed to address the issues mentioned above. It is XML driven and based on XML schema, with these schema forms, the data entry forms are generated from XML schema documents that define the data types, limits, enumerated lists and other constraints. Effectively, Schema-Forms can handle an unlimited number of forms based on the user-defined XML schema. A diagrammatic representation of the new approach design process can be shown in Fig. 1.

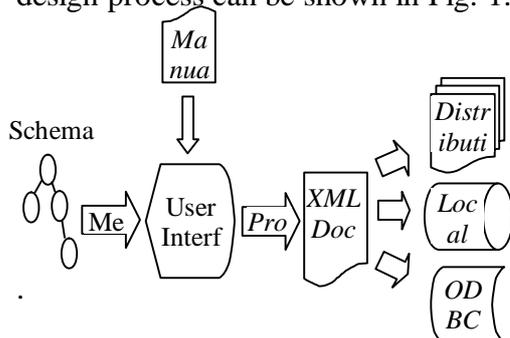


Fig. (1): The Basic function design for the schema form

The output from the data entry starts as an XML document that can be distributed to a distribution list, stored in a local database, posted to an Open

Database Connectivity (ODBC) compliant relational database. The distribution and posting can be conditional based on the contents of the document, this is a facility made possible by XML technology.

The XML schema documents are user-defined; they can be defined with any other XML Schema tools. The process in our design is a user-driven, because we can define the schema and the distribution and posting instructions, which make the process fast, efficient and cost effective. As the XML schema is in pure ASCII, they can also be easily distributed. The whole process can be done offline and together with the local database storage, supports the long transaction requirements described above.

3. The New Information Model Architecture

In order to perform the new proposed information management system we recognize that these documents should be stored as an information network where any document can be connected to other documents. Instead of trying to handle the network conventionally, this new information model has been designed as an inverse of the network whereby any node (knowledge object) can be selected as the centre of focus, and lists represent related objects based on relationships.

The tenets and the architecture of this multi information model are as the following:

1. Information is represented as objects. An object represent anything that can be given a name, e.g., a concept, a person, an organization;
2. Every object can be the centre of focus with all related objects made visible i.e., there can be

as many perspectives as there are objects. This object can also be viewed as a tree with the branches expanded to any level (subject to system resources) as explained in Fig. 2.

3. All objects are equal. An object can have more than one name and more than one description. Any list of objects can be filtered, clustered or expanded based on the relationships defined in the network. The lists also represent additional perspectives. So that they can be dynamically rearranged without losing context.

4. The common set search feature allows users to find common sets from a list of objects based on the set Theory. Objects can be related to each other as groups and members (i.e. related in hierarchical structures or generic relationship).

5. Objects can be associated with each other, the connectivity search feature allows user to find the connections between any two objects through multiple levels. To make this search more meaningful, selected relationships and objects can be excluded from the search. XML-based document search allows users

to find text within specific nodes of documents.

6. Objects can also be classified into themes; the data query allows selected information to be extracted for display purposes or to other formats for processing which can be also referred as layers that can be shown in Fig (3).

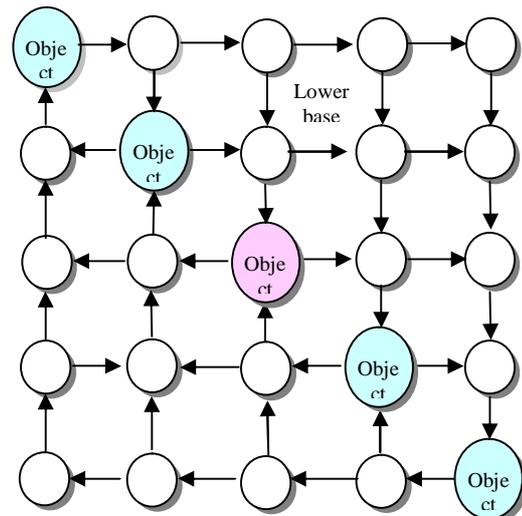


Fig. (2): The Main interfaces between objects

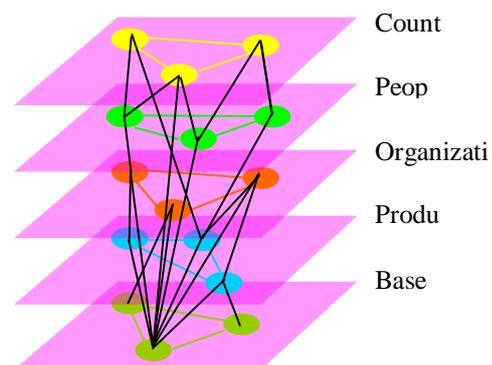


Fig. (3): The thematic map concepts

The concept or the knowledge of the new proposed model represented as a network of information object that is defined in XML and based on relational database technology. A schema document (say B) can reference another schema (say A). In this model, the system will allow the

definition of document mapping i.e. retrieval of specific nodes from schema "A" documents to schema "B" documents. In our model while viewing a document of schema "A", there will be an option for the user to launch Schema-Forms with a new entry form based on schema "B".

The information from the viewed document will be transferred to the entry form based on the nodes mapping defined. Since the posting of documents in Schema-Forms can be conditionally defined, the schema "B" document can be added as another document to the schema "A" document and update the relationships of the object containing the schema "A" document. Alternatively, the schema "B" document can be posted as a new object related to the Schema "A" document. These powerful user-definable model features allow the development of generic document workflow applications rapidly using XML Schema-Forms for a variety of purposes.

4. Conclusions

Developers today are challenged to integrate their applications with other systems over a network. Oftentimes these other systems are deployed on disparate platforms and are written using disparate programming languages. XML has quickly become the defacto standard for exchanging data between systems; possible solutions for consuming XML documents including using low-level generic parsing applications have been developed.

Using XML Object Link offers the ability to easily handle XML documents that can readily be exchanged with other systems, including those based on different languages. Information Technology is the key to the knowledge management and the knowledge management has the potential to help the organizations

innovate, make better decisions and improve their competitiveness.

A unified approach in managing information has been designed which capture divers information in a single database that the user can navigate, explore, expand, filter and analyze. Making every thing visible because information is only useful if we can see what is really there.

The XML-driven solution model that we proposed provides a viable platform for the capture, management, distribution and publishing of enterprise information. XML is an emerging technology that committed for enterprise information management as they become available. This approach has stood the test of time. The new multi information management solutions can be used on desktops and within local area networks.

References

1. XML Style sheet Language Transformation (<http://www.w3.org/TR/xslt>), 12-3-2004.
2. Extensible Markup Language (<http://www.w3.org/TR/REC-XML>), 05-12-2001.
3. Human computer factors (users and Information systems), Andy Smith, (1997), Washington DC, USA, pp 14, 87, 117.
4. Model Driven Architecture, Joaquin Miller and Jishnu Mukerji, *Modeling design journal*, Architecture Board ORMSCI, 9-July- 2001, UK.
5. Oracle/ XML SQL Utility, <http://www.oracle.com/technology/tech/xml>, 22-01-2003.

طريقة جديدة لتصميم نظام لادارة المعلومات المتعددة باستخدام تكنولوجيا ال XML

د.سها محمد هادي

قسم هندسة المعلومات/كلية هندسة الخوارزمي/جامعة بغداد

الخلاصة:

ان تكنولوجيا ال XML تم اشراكها في البناء الاساسي لكثير من تطبيقات قواعد البيانات الخاصة بالاعمال الالكترونية حيث يتناول هذا البحث مشكلة المعلومات ذات التكوين الحر التي تخزن في اي منظمة او مؤسسة وكيفية استخدام ال XML مع هذه الطريقة المقترحة والتي سوف تجعل عملية البحث عن المعلومات ذات كفاءة اكثر وبأقل وقت ممكن.

ان هذا البحث يعرض طريقة حل جديدة مطورة لامسالك ولادارة مثل هذه المعلومات الغير مهيكلة والحررة (اي المتعددة) والمعتمدة على استخدام تكنولوجيا سكيما ال XML ومبدأ الشبكات العصبية وقواعد البيانات الشبئية لغرض توفير حل عملي لادارة كفاءة لانظمة المعلومات المتعددة الحررة .