1) IMPACT OF INTERNET ON DATABASE

2) THE XML STANDARD AND ITS DERIVATION

3) XML AND DATABASE

4) XML AND FEDERAL DATA

5) XML AND THE MANAGEMENT OF KNOWLEDGE
INTRODUCTION

The explosion of internet and the progress of technologies give some new approaches of databases used in information system. It has some consequences on the management of information which becomes more complicated. Firms are some good consumers of big and heterogeneous information. Also in order to answer to their needs, some new standards such as XML: extended marked language trying to make and represent information. This standard integrates structural data and multimedia data and more of mixed heterogeneous data it gives the interoperability with all objects which use different data.

So the management of knowledge becomes more automatically with low costs and also that's why XML will have a good future.
1) IMPACT OF INTERNET ON DATABASE.

Databases were developed in 1990 around the object and the relational object. In order to answer to the needs of firms' local area network (LAN), it has become an obligation to couple databases and internet technologies. The development of internet and firms' technologies show three kinds of applications that need the mix between internet technologies and databases.

First, there are the application client-server web, which are built on three states: client->application server->data server. This kind of architecture is a result of a database that must be opened to external clients of firms' LAN and which can access via the web to the system information of the firms and give the possibility to the firm's personnel to access to the system information of their firms.

After the application client-server, there is the development of web sites, which are very numerous and very heterogeneous. The contents of web sites have become very complex, for example, we can find several HTML files with different links and it has become very difficult to update them when it is necessary. That's why databases are a good solution to solve this problem. Databases provide a method of representing information and organize a logical hierarchy between the information.

The last application is the web business. The need for this application is, for example, to show some information to consumers and, more importantly, to memorize the contacts and what consumers like or dislike.

In order to explore the links between the products and to show the hierarchical descriptions and prices, it must use high-speed connections and transactional databases.

The new architectural model of web data, which couples the web and the database, needs to extend HTML to take some data and some requests and to make the web servers more dynamic and to increase the bandwidth between the database and the application.

Today, the architectural model consists of three levels:

1. A level that controls the user's interface
2. A level that includes a web server and adds an application server
3. A level that gives control of the data in the database

Each of them must give the possibility to the application's code to run, to show the data, and to treat the application's logic or to manipulate some complex objects that are stored in the database.

As a consequence of the lack of HTML's extension, new development of data web is going to define a model of exchange that includes the information's content and not the presentation.

So, the XML defines a standard model of data exchange.
2) THE XML STANDARD AND ITS DERIVATION.
3) XML AND DATABASE.

•
•
•
Figure 2: An example of a semi-structural database. Which language can be used for finding the documents in a semi-structural database? The W3C work on it. Some propositions are done: XML-QL of AT&T, XQL of Microsoft. It is a possibility to extend SQL with navigability expression and filter with graph builders, this is the XSQL language. This new query language should have an important impact on the search engine.

Today the relational DB such as Oracle or DB2 are the leaders of the market. The semi-structural object can be integrated as an instance of a particular type in a relational object system. Oracle uses its type <<free text>> (cartridge intermediary) to make it possible for storing documents in XML format and using the interrogation by keyword.

Figure 3: An example of a storage XML document in a DBM. (source from eXMLmedia).

<table>
<thead>
<tr>
<th>root</th>
<th>origin</th>
<th>target</th>
<th>label</th>
<th>ltype</th>
<th>ntype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>name</td>
<td>AGOR</td>
<td>STRING</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>manager</td>
<td>AGOR</td>
<td>INTEGER</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>address</td>
<td>AGOR</td>
<td>NODE</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>number</td>
<td>AGOR</td>
<td>INTEGER</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>6</td>
<td>street</td>
<td>AGOR</td>
<td>STRING</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>drive</td>
<td>ASSOC</td>
<td>NODE</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>8</td>
<td>mark</td>
<td>AGOR</td>
<td>STRING</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>9</td>
<td>color</td>
<td>AGOR</td>
<td>STRING</td>
</tr>
</tbody>
</table>
4) XML AND FEDERAL DATA.

A federal database is a heterogeneous database which is made of different data such as text files, HTML documents, XML, etc. The objective is to give an integrated view to the users of the different data from the enterprise dynamically on request (Technology PULL), or materializing periodically in a data store (Technology PUSH).

Historically, this kind of database is developed around the relational model and SQL. CORBA makes it possible to integrate the applications. However, CORBA hasn’t got a real standard interface with databases and it is not very easy to evolve it.

From a point of view about the architecture of the federal database, there is a standard from DARPA and GIO Widerhold.
8

The central level is a mediator which integrates the data from different sources for composing the responses to the users requests.

The most internal level is made of several adaptators or wrappers which must transform the local data in order to be compatible with the mediator and to filter the data after a request.

The external level is made of facilities, these tools must locate the pertinent data and order it for the applications.

The mediator and its interface are an exchange bus called bus EAI (enterprise application bus). But the object mediators have failed because they are too complex if they must be evolved.

XML is a new exchange standard and it can be a solution to take the place of the above architecture.

Figure 6 shows how it can be used. It is an example of a component which were made by the eXMLmedia enterprise:

5) XML AND THE MANAGEMENT OF KNOWLEDGE.

The technical federal databases make the collect of information from different sources possible but it needs to precise and to integrate more information. It is the domain of application technical data mining which represents the future generation of intelligence tools. Data mining consists to extract a part of information from a big source of information and make it exploitable.
It is interesting to take all the fragments which have the same structure and to store it in a relational table in order to have more free volume and more time when it needs to be extracted from the based documents. Finding the repeating fragments is the domain of the data mining. Another direction of data mining is the automatic tags. A lot of documents from firms need to be transformed into XML files but we must find what kind of important information in the documents should appear. It must find an automatic intelligent tag. Another way is to used intelligent dictionaries which describe the speech domain, and these concepts and their relations.

Conclusion

XML and the semi structural model make easier the integration between web and database. This new kind of database is different that the others because its structure has got more flexibility schema and these are mixed with data. This database takes the links between inter documents and supports hypermedia, also it means the search and the structure of hypertext documents or hypermedia. XML has got a promise future generation. With the evolution of the java language, the XML unifies document formats and java unifies the language of programmation in the three states architecture. However a lot of problems appear around the semi structural database. We have told a little bit of optimization of the storage and of data mining which opens new horizons.
Bibliography:

Online resources
http://www.e-xmlmedia.fr/site_francais/documentation_livres_blancs.htm

Georges, Gardarin Les datawebhouses arrivent

Text resources
Lessons and photocopies from INT.