

Databases on the Internet

This paper describes using the Internet services particularly one, World Wide Web(WWW) as a medium for business expanding. The main project goal is finding a solution for storing data in database and reading from it over the Internet. Some problems are found during the project realization and we provide the solutions for them. The test project is developed with C++ and Java but under the scope of analysis is only part in Java.

Keywords: Database, Internet, Java, object orientation.

1. Introduction

This paper originates from a project under the working title "FOI in CARNet/Internet", which is being carried out at the Faculty of Organization and Informatics (FOI). Project description is given in (Kerm97).

Present day business applications development trend leads towards involving new technologies, such as the Internet. The Internet, with all its services outgrew the academic and military limits within which it actually originated, and it becomes a term more and more used in everyday life. WWW has become a dominant Internet service due to its features, such as graphical interface and because it is easy to use. Many companies use these WWW features for their advertisement but more and more for their on-line trade. Successful realization of this idea calls for partial installment of information system on the Internet, which means there will be a data base to store data in and read from.

The paper describes development of a data storing process on the Internet and the arising problems, too. It shows new technologies which not only solve the noticed problems but also create new ones. The new problems have been identified and ways of their avoidance recommended.

2. Distribution of programs on the Internet

Realization of global network, such as the Internet has, inevitably created the need for storing the data and an access to them. It is not important any more where the place of data creating and their input is, or its distance from the computer into which the data are stored. Looking from the edging positions i.e. data creating positions, input, storing and browsing, it is clear that ways of data transmission do not need to be

known to agents from the edging positions. Thus, it can be concluded that the principal mainstay of the whole work should be the transparency that is offered by the operating systems.

Development of program support in this global environment is put additional demands on, which were not present in such an extent in the development for usual local environments. Let us just remember the problems with installing new versions of program support on computers within a company, when only slightly spatially distant. This procedure must be carried out on every used computer. The situation is a little simpler when it comes to uniform computer configurations and equal operating systems. This idealism soon vanishes when it comes to real applications in a business environment. The client/server technology has brought great progress into data administering, approach and other operations in the same category.

Following the logic thought, we are brought to a question: if data are entered decentrally and stored centrally (client/server), why should not the same principle be applied to programs? Is it not simpler to store a new version of a program at one place, i.e. at the server, so that all the clients can have equal access to it? This solves only one part of the problem so there still remains the problem of portability of different architectures and operating systems onto computers. For one and a half year now, the solution to the problem has been coming to light. We are talking about the Java programming language that enables a program written in it to prepare for performing independently of the target platform.

3. Data storing on the Internet

If we look at the history of the Internet (LCC+97) we can clearly see that its development and application follow the usual high technology "products" pattern. The idea is first born within university area, followed by its placement under the patronage of the army and finally it is widely used in everyday life.

Internet is coming into everyday life of a common man more and more, whether we like it or not. Negation of its influence can be a solution, but there still remains the question whether it is the best solution. Global television companies like CNN and NBC have accepted the Internet as means of spreading their influence, but also as an additional way of communicating with its viewers. Regarding the time component, the present forms of communications can be divided into two groups: direct and indirect. Telephone conversation is the best example of direct communication, but has several disadvantages. It has been made clear that its price puts off a large number of users from distant countries. Besides, there is the language barrier in direct conversation on the air. Telefax has given the chance to many people to use more time to compose the text and correct it grammatically. Also, when it is sent, there is no waiting time at the expense of the sender. A letter still has its place in communications, but it has lost the leading role which it used to have ten or more years ago.

The Internet has brought the possibility of communicating through electronic mail. It uses the positive characteristics of telefax, which are further improved in respect of

the price of message sending, and the simplicity of receiving an answer to the message.

Along with the electronic mail WWW is the most exploited service on the Internet. Its usage by a common man is stimulated by the aforementioned advantages that electronic mail has in communications. Although it is only one form of the advantages, there is another one which is directed towards information supplier through WWW server. The price of only one second of television advertisement or a page in the papers is measured in huge amounts. Many companies have reduced the duration or newspaper space for their advertisements, but they have included one new piece of information. They give their WWW address now. Since it was begun by companies in the USA it has the form of: <http://www.name.com>. All who want more information about the company's products or services, can look for further information at any time because WWW server works from 0 - 24 hours. Operation drive expenses are far lower than the expenses of traditional advertisements. In this way several advantages have been gained. The first one refers to permanent information presence and updating, while the other one means that the expenses are covered by the user of the information.

Further development takes us into ordering and buying through Internet. This is an area that belongs to the standard conception of computer supported information system, which uses database with details and quotations of offered items, etc. We have now come to the topic area of this paper, namely, loading data into database and having an access to them.

Successful implementation of the set request cannot be based on the usual static HTML pages that are updated by hand. Now is the time for application of dynamic pages, which are in effect, based on pattern pages, filled with necessary data, upon the client's request and available information in the database. Figure 1. shows the server's diagram, which contains static and dynamic pages.

There are numerous examples which show the use of different techniques of data input and access to them. First attempts are linked with the so called data input forms on the user's side of WWW browser, and CGI script on the WWW server's side, and also with the data record in the usual text files. Along with such a work technology there is also a parallel central information system for habitual work, which relies on a certain database. Criticism of that technology can be directed towards two main problems.

The first one refers to data input by means of forms where there is no possibility of entered data check. For this reason the entered data should be sent to the company's server, where they are processed by CGI script and the commentary is sent in case of a wrong or incomplete data input. It is clear that in this case there is a busy data exchange, waiting for an answer or instruction. It is much more practical for data to be processed at the place of their input before they are proceeded to a server. All errors and defects can immediately be detected and a client can be informed.

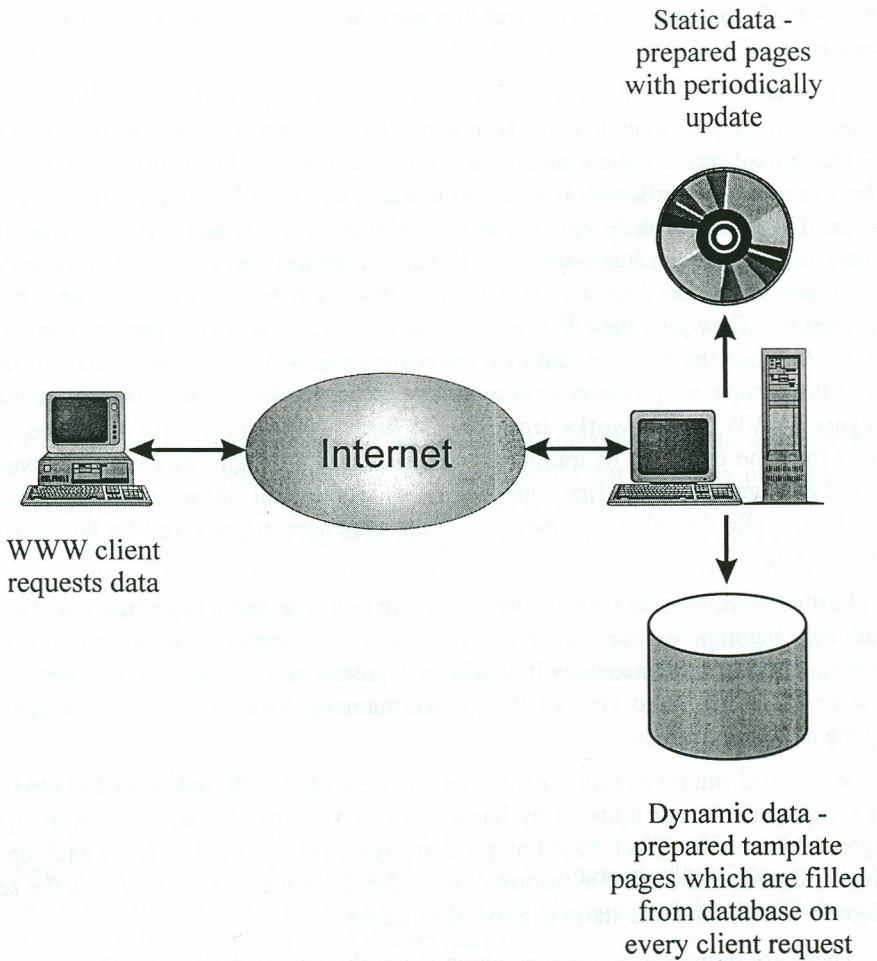


Figure 1. WWW server diagram showing static and dynamic pages

The other problem lies within the data redundancy domain. The double system called for supplementary data conversion procedure during which the data about changes in the Internet shopping will be adjusted to a database format, and updating of the central information system will be carried out. Here it goes for extra resource spending and the possibility of failures due to the additional operations. The simpler technique should be used where data are entered in one format and in one single operation, independent of the source of information, it being a classical procedure or the Internet. Hence the suggested solution where operations for data recording into databases are called and read from.

New technologies are trying to eliminate the existing defects. The company NET WAY has offered a product called SQL Surfer which solves the problem of data redundancy by their direct loading into a database. SQL Surfer can be said to be specific because on the browser's side HTML language is extended with new

constructs which define SQL commands requiring their completion. Figure 2. shows the simple example of a copy containing all lines in the table NASTAVNIK (Teacher) with the established HTML links according to personal teachers' pages. IDS Software company offers a similar solution named IDS Server.

```

<SA-SELECT>
  <SA-SQL-STMT>
    SELECT  IME, PREZIME
    FROM    NASTAVNIK
    ORDER BY PREZIME
  </SA-SQL-STMT>
  <SA-INTO $(n_ime), $(n_prezime)>
  <SA-DATA>
    <LI><A HREF="$ (p_cgi) &p_tmpl=nastavnik.html&n_prezime=$(n_prezime) ">
      $(n_prezime) $(n_ime)</A>
    </SA-DATA>
</SA-SELECT>

```

Figure 2. Example of SQL extension of HTML language

HTML language SQL extension interpretation is carried out on the side of the server in the form of CGI script or an executable binary program. Data are repeatedly entered in the planned forms with a very small data input check by means of a built in script language, but very often without a check. By execution of form the request is formed, which is then forwarded to the server. Upon demand interpretation by the CGI script and executive program respectively, it is proceeded towards database, and if necessary it receives databases which are structured in accordance with the HTML page form. The finished page is sent towards a user's browser.

The suggested solution of SQL - Surfer can be applied in those cases when direct data updating and an access to the data are needed. Comparison with the previous solution and its criticism shows that there still remains the problem of data check on the server's side, and all the problems that were mentioned earlier in this context. An openness of HTML language through the possibility of an insight into the original text of every page can pose a source of danger for the database integrity. The example in Figure 2. shows SQL command which gives access to a table in database. Other tables in the database could be found in the same way. If there is a person, and there usually is, who wants to test their destructive potential towards other people's data, the only thing missing is the knowledge of SQL language and poor protection of database on the server.

All forms described so far have one problem in common, and that is poor control, or none at all, of data input by means of forms, and database structure openness towards a user. Recently there have been more and more indications of a new solution

which solves the aforementioned problems. It is the programming language Java with the use of JDBC support for an access to databases. The next chapter of this paper gives a short description of Java and its key features which make it different from other programming languages.

4. Java programming language features

"Java is a simple, object oriented, multithreaded, with built in garbage collecting, safe, robust, architecturally independent, portable, high performance, dynamic language" (DeS97). The appearance of the programming language Java can be pronounced the main event in the area of program support development in the last decade of this millennium. Numerous conferences, seminars, periodicals and books are only a few indicators of this success.

The popularity of this "tyro" lies in the platform independence concept, which originates in the interpreting character of the programming language Java¹. Program written in Java must be translated by Java interpreter. The obtained result is a file in the so called *Java byte code* format, which, unlike other languages such as C++ whose final product is an independent executive version of the program, cannot perform independently, but needs another program that makes it possible. This program is called *Java Virtual Machine* (JVM) and its role is to interpret Java byte code instructions and to carry them out in accordance with the platform on which it is installed. JVM is a Java programming interface designed according to the computer and operating system features. It offers transparency for Java programs. For this very reason every computer with its own operating system should have its own JVM.

Implementation of JVM can be found in two versions: browsing and independent. The first one is based on performing within different WWW browsers² and is characterized by broader application, for which reason this version will be considered in the later part of this paper. Java application must be specially prepared for such a performance in the applet form, which is recognized by the browser as a part of HTML language, and is then proceeded to JVM for execution. The independent version is similar to programs written in other programming languages, but a program containing Java is still necessary. Regarding the Java program structure, the key difference between the two kinds lies in the methods to be included. Browsing version behaves as a part of a browser, a window within which it is performed, while

¹ It is not for Java, that we met for the first time the interpreting character of object oriented language. Smalltalk as the first object oriented language, is also an interpreter and there is also platform portability within its producer. By translating (or interpreting) Java program, a proper file is created for every class (class name along with the class extension), while Smalltalk records all the interpreted classes into one single program memory image. In the course of the Java program execution JVM loads only those classes that are needed for the course of performance. Others are not necessary in this course or performance but may be necessary with the other one. This seems to be a particularly important feature especially when Java program is taken over from a location within the Internet. Reduction of the amount of data transmission enables greater permeability for transmission of other data. Smalltalk is based on the image, which means that the whole of it must be at disposal before its performance starts.

² Netscape Navigator v.2.0, Microsoft Internet Explorer v3.0, Sun Hot Java v1.0.

independent version appears as its own work space and does not need to communicate to other windows of the same parent.

The other reason why Java is so popular is based on its object oriented features which come from the basic programming language C++. It is usual for a foundation and its superstructure to be related as a set and an overset³. In this case we can say that this relation is not dominant because there is a large number of properties which are not Java supported. The Java authors point out that some features of C++ language e.g. multiple inheritance, are not often applicable and were not worth installing. With other features the main reason was protection of environment where Java is performed. Computer viruses have shown that the use of unauthorized programs can end in disastrous consequences.

The Internet is a medium where Java programs can be taken from, represents an ideal environment for distribution of "evil" programs. Everybody who has felt the charm of "lost" disks, files or programs, might feel such a strong aversion towards Java that it may prevail over the advantages it offers. The use of a pointer with C/C++ is considered to be one of the main methods of work. The different dangers which are brought along with them (access to memory location outside the program or the famous memory "leaking") only confirm the law of fire and bad master. This is why Java⁴ does not have explicit pointers of memory locations and does not support operations with pointers. Instead of pointers there are references which carry out linking to objects. A supplement in the form of garbage collector⁵ makes sure that memory locations of those objects to which leads no reference, are free. Finally, the security of Java programs performance depends on the security given by JVM, within which, in effect, it is performed.

The beginning of the use of Java brought certain dynamics in the Internet medium. In this case dynamics means control of operations with a mouse, picture, table and graph generating, animation, including sound and sound records in the applications, network services etc. This is probably the positive influence of program examples delivered by the Sun company along with Java interpreter. Installing an access to databases came a little later.

Connecting the access to databases is the best sign of Java programming language maturity. Animation and sound make individuals' and companies' personal WWW pages more attractive. In the business area visual experience of WWW pages gains a slight advantage over other competitors. If there is a possibility of computer communication with the company and its database of products and/or services, then this can be a decisive factor which will prevail in its favour and not of its competitors.

³ This is not always the rule but if we divide the features into groups then the said relation applies to majority of groups.

⁴ Java does not support multiple inheritance, but interface application opens similar possibilities.

⁵ Engl. garbage collector.

5. Java and databases

The market today offers a great number of databases. In business applications domain, the dominant place is taken by relational databases. Standardized use of databases through SQL language made simpler the existing diversity in the situation where there is no generally accepted standard. Some cheaper databases are oriented towards smaller quantity of data input and do not support all recommendations of relational data model. It is to be expected that more expensive databases offer wider possibilities.

The union of diversity of databases, programming languages and operationg systems clearly shows the situation where designers and programmers could find themselves during their work on a project. In the early stages of program support development many decisions have to be made. What happens if some of them are wrong? Microsoft company has suggested an open approach to databases, ODBC⁶, which separates access to data from knowing their organization.

The introduction of ODBC the work with databases in the local network area is made simpler. In the Internet medium the main means of work is the selected browser. They (for the time being) do not have built in capabilities of work with the databases using SQL language, so, consequently, it is not possible to use ODBC. Producers of system software support suggested the solution in the form of JDBC⁷. "JDBC is Java API⁸ for performing SQL commands. It consists of a set of classes and interfaces that are written in Java programming language. JDBC offers a standard API for tool/database programmers, enabling application writing for databases by using pure Java API".(Jav97).

JDBC is performed by means of an appropriate driver within the category of a system program support. In order to avoid the diversity of drivers the Sun Microsystem company has suggested the reference "JDBC compliant", which, in effect, means a set of standard JDBC functionality that users can rely on. Each JDBC drive⁹ should at least comply with ANSI SQL -2 Entry Level (IDS97).

All JDBC drivers are not built in the same way. This is why they are put into four categories:

1. JDBC - ODBC bridge - on the side of the server an ODBC driver program is needed. JDBC - ODBC bridge carries the user's requests to ODBC driver, and forwards the received data to the user. The binary ODBC code has to be loaded on the user's side, but very often the user's database code, too.

⁶ ODBC - Open Database Connectivity.

⁷ JDBC is a protected brand name, not an acronym. Despite this, JDBC is often thought to mean "Java Database connectivity".

⁸ Application programming Interface.

⁹ Engl. driver

2. Native -API partly Java driver - interprets JDBC calls into clients' API for appropriate database. In this case it is necessary to load some binary code on the client's side.
3. Network- protocol full- Java driver - carries JDBC calls into a network protocol which is independent of database and which is further carried into the target database protocol on the server. This particular form of driver is capable of linking pure Java clients with many different databases because it depends on the database supplier. Generally speaking, it is the most flexible JDBC variety.
4. Native - protocol full- Java driver - interprets the JDBC calls into a network protocol directly used by the database. This makes possible the direct calls from the client's side towards database on the server, which is a practical solution of access through the Internet.

At the moment there are several JDBC drivers on the market. JavaSoft as an affiliated company of Sun Microsystems, has offered JDBC driver of the first category, and IDS company has produced JDBC driver of the third category. The tested examples are based on the IDS JDBC drivers.

Figure 3. shows a situation where there are two users of database. The user on the left side already has the necessary programs on the computer, while the one on the right side does not have them, so they have to be obtained from the server.

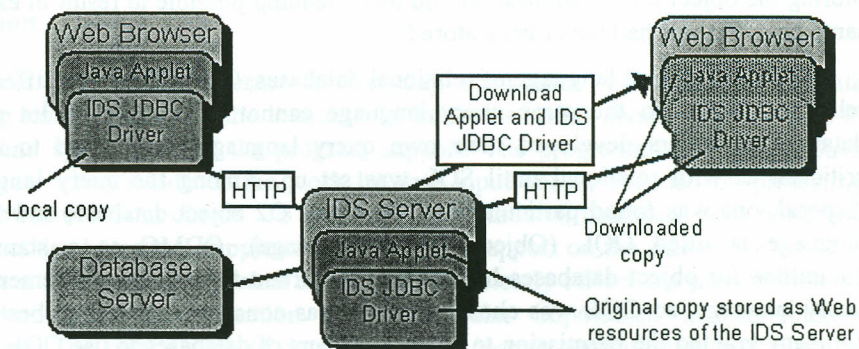


Figure 3. Example of IDS JDBC driver in network environment (IDS97)

The use of database in a program written in Java, through JDBC driver must be in accordance with the prescribed sequence of operations. The first operation presents an initiation of JDBC driver. Then follows establishing the connection with the required database through JDBC driver. This step consists of several suboperations. First, we have the identification of a driver by means of its own signature (jdbc:ids://). After that follows determination of the required database for the connection (www.foi.hr:12). At the end there may be an identification of a user according to the database by means of

the user's name (&uid=guest), and identity confirmation by means of a password (&pwd=foi). All of these instructions are defined in one sentence:
jdbc:ids://www.foi.hr:12/conn?dsn='SQL Server ' &uid=guest&pwd=foi

Other operations require obtaining the data from database using certain SQL commands. The main difference is in the obtained data interpretation by the server. Relational database consists of tables where the data are written in rows. Each row can be regarded as one record, and columns in a table as a field. Fields are the smallest parts of a database which can be used. In the case of Java usage there is no direct access on the column level, but the columns are obtained by line splitting into its columns. In other words, we can say that there is no direct assignment on the field level.

Object orientation influence is spread over the database area, too. Object databases are no more second class semi-processed products, which cannot be used in the commercial application domain. The market has several object databases e.g. ObjectStore, O2 and Poet. The market value of a reference "object oriented" compelled the relational bases producers to start installing some object oriented properties in some of their databases. It is expected that the majority of existing databases will take this step in the near future if they have not done so far.

Object database connects the semantics of object oriented programming language with data management and query conventional database features. This simplifies the job of managing large amounts of data and the model of relation between them. Storing the object into a database should make loading possible to result in exactly the same object as if it had never been stored.

SQL is a standard language of relational databases. Object model is different from relational model so the same query language cannot be applied. A lot of object database producers developed their own query language which led to the same criticism as with relational until SQL was set up. Among the query languages at disposal, one was found particularly good. It was O2 object database, and the query language is titled OQL (Object Query Language). ODMG as a standardizing committee for object databases held it very important to reach an agreement on one query language, so OQL was chosen since it was considered to be the best one. O2 company granted the permission to other producers of databases to use OQL for their databases. ODMG continues to work further on standardization of OQL language properties. Among others, there is documentation referenced as ODMG-93, etc. (Poe97).

Access to object databases can also be carried out by means of ODBC, which means that object database can be used as a base for data storing of WWW server. Object database testing of examples was done on database Poet.

6. Problem of database security on the Internet

Databases which are used inside local computer networks are not particularly in "danger". First of all, it is clearly defined who has an access to a certain computer in the network. Network operating system also provides a sound protection from unwelcome users. An openness of the Internet can generate reduced security of a database which contains information about doing business through the Internet.

Database that will be at disposal for Internet users should be particularly well designed regarding the authorization of access and operation permission. This is why some simple databases like Access, dBase and the like, should not be chosen for this purpose because their protection mechanisms are not powerful enough. MS SQL Server and Oracle are a good choice because with them several database users can be determined and to each of them a set of operations can be assigned. It is obvious that a client of WWW application should be assigned a database user who has a limited number of sets of operations. This set should by no means have the operation such as DROP TABLE, which erases tables from the database.

The stored procedures calling (Jep97) represents one of the successful elements of database security growth. WWW client using HTML page with SQL extensions is free to have an insight into the original page text, hence into SQL commands. It is up to him whether he will make a mischief or not. In case of the stored procedures calls being used, the possibility of an insight into database structure is reduced.

Conclusion

Business use of the Internet is everyday occurrence. Databases make data processing simpler. Merging the Internet and databases brings about the new needs in program and system part of the program support. The programming language Java facilitates the client's side of the work by data input check, but also by protecting database structure from clients. To link Java and relational databases, and more and more object databases, too, supplementary program support on the side of the client is needed. It is JDBC driver. Several kinds of JDBC clearly reflect the present state of development. Database protection is of particular importance in the Internet environment. The right choice of database and its correct configuration will result in peaceful dreams for everybody who is in charge of its exploitation.

References

1. (DeS97) DeSoto, A. *Using the Beans Development Kit 1.0, A Tutorial*. JavaSoft, Sun Microsystems, 1997. <http://www.javasoft.com>, 10. April 1997
2. Leiner, B.M., Cerf, V.G., Clark, D.D., et al. *The Past and Future History of the Internet*. Communications of the ACM, vol.40.no2., 102-108. 1997.
3. (IDS97) IDS Software Documentation for IDS Server <http://www.idssoftware.com>, 4. May 1997

4. (Jep97) Jepson, B. *Java Database Programming*, John Wiley, 1997.
5. (Ker97) Kermek, D. *Metrics of Visual Programming in Object Oriented Environment*, CASE9, Opatija 1997.
6. (Net97) NetAway documentation for SQL - Server, <http://www.netaway.com>, 07. May 1997.
7. (Poe97) Poet documentation for Poet, <http://www.poet.com>, 20. April 1997.

Received: 1997-08-15

Kermek D. Baze podataka na Internetu

Sažetak

U članku je opisana primjena jednog Internet servisa WWW kao medija za proširenje poslovanja. Glavni cilj projekta je da se ponude rješenja za pohranjivanje podataka u bazu podataka i čitanje iz nje preko Interneta. Uočeni su neki problemi tijekom realizacije projekta, te su predložena rješenja. Testni projekt razvijen je u programskim jezicima C++ i Java, pri čemu je analiziran dio u Javi.

Ključne riječi: Baza podataka, Internet, Java, objekt-orientacija.
