INTERNET DATABASES IN QUALITY INFORMATION IMPROVEMENT

Cosmin TOMOZEI

University Assistant, Mathematics and Computer Science Department,
Faculty of Science
University of Bacau, Bacau, Romania

E-mail: tomcosmin@yahoo.co.uk

Abstract: *Even though many important companies are reluctant into deploying their databases on the Internet, being too concerned about security, we would like to demonstrate that they shouldn't be worried too much about it, but to try to provide information in real-time to management, boards or people who travel on companies interests.*
However, security is one of the most important factors that should be offered to websites and databases on the Internet. If we consider one of information quality metrics, the time between the sending of the message and the receiving, it can decrease considerably thanks to a secure, normalized and non-redundant database.
Another direction of our study is the interdisciplinary approach, including the cooperation between management science, information technology and quantitative analysis in order to provide a perspective for improving information's quality.

Key words: *data protection; redundancy; data integrity; flexibility; economic efficiency; reliability*

Internet databases, concepts and technologies

In the last decade, databases have been mostly Internet oriented, Oracle or SQL Server providing opportunities for web developers to use them in their applications. Companies, due to globalization also oriented their Information Systems to a distributed approach, available for users, target groups from all over the world due to Internet connections or remote connections. Such an application we developed for the Chamber of Commerce and Industry Bacau with the main purpose to help the management to store and update their activities and offer support in the process of decision making.

Formerly, when a department needed information from the database, it would have been compulsory to ask the IT department to retrieve it from the information system and the answer would have come in a few days, due to the amount of activities in which the department was already involved. Now, anyone can do his own duties via Internet applications, the software giving them specific answers and information in real time. As a result, the gap between the solicitation of data and the time of response has been reduced to almost zero.

The economic efficiency is another important approach because the main objective of any economic entity is to be efficient, profitable and to gain market success. Efficiency and success are strongly connected with a solid, non-redundant, fast moving and secure information system. However, some of the activities cannot be published on the outside network without being protected from any kind of intrusion attempt or unauthorized access. For such a reason, we did our best to maintain protection, security and integrity as main priorities.

In Information Theory, redundancy is defined as the number of bits used to transmit a message minus the number of bits of actual information. Data compression is the most important way to eliminate the unwanted redundancy. On the other hand, Database Theory defines redundancy as the degree of data multiplication, which can be accepted if it doesn't exceed a certain level so as to become uncontrolled.

However, in distributed systems redundancy appears due to data replication, if such a project approach is designed by the analyst.

Replication is the process in which two or more databases communicate transferring data and information. Even though it may look like information cloning the difference between cloning and replication consisting in the fact that a replication server is supervising the data transfer (replication server eg IBM DataPropagator™ for iSeries™, V8.1) .Every database is updated when updates or modifications are made on one replica.

Data Integrity has been also very important in our approach, and it has been taken into consideration by both sides:
- entity integrity, meaning that the primary key columns for each table must contain a unique value.
- referential integrity, each value in a foreign key column must exist as a primary key of the table it references.

Furthermore, the security is also assured by granting access on the database, to read, write, insert, update or delete data depending on the role that the user has. The database administrator, in our case the IT analyst from the Information Technology Department is the one entitled to create user accounts and to offer them specific rights.

Consequently, a unique password and username is provided to every user for working on the database after the process of authentication. For example, the department of Business Information as administrator of the business information system, offering advice to companies about business opportunities, creating and supporting contacts and partnerships between companies as well as offering support for the development of the economic region, is enabled to make daily updates to the databases. The employees of the Department of Business Information are also responsible for the correctness and the accuracy of the data.

Internet database quality metrics

The first step in database design is data modeling which realizes the link between the end-users and the software solution that becomes useful for them. The concordance (agreement) is a metric which compares the results of a specific prototype to the initial objectives. Even though it represents a small fraction of the entire development process, it is very important to spend enough time and resources for data modeling.

It is also very important that this stage of the process should not remain just a theoretical approach but actually to be implemented and followed on the other steps of application development. Software engineers, application developers and analysts are focusing mainly on the code elaboration related quality metrics, process modeling and data analysis not being as much taken into account as the above mentioned factors.

In [PIGECA06][1] there are some proposals of data models. The model should also be validated, audited before continuing the next steps.

Project complexity is defined as

$$E = \sum_{i=1}^{n} (E_i)^C$$

where **n** represents the number of entities of the project, **C**>1.

The complexity of the entity is represented by *'i'*

$$E_i = D_i * F_i$$

D_i - data architecture complexity;

F_i - functional complexity;

$$D_i = Ri(a * FDA_i + b * NFDA_i),$$

Ri - the number of relationships(associations)
FDAi - the number of functional dependencies
NFDAi- the number of non functional dependencies.

Data modeling, relationships and dependencies are all shown in our application in the database diagram 1, which describes the entity-relationship model of our application.

Accuracy measures if the database entities are correct and all the errors have been removed from the model.

Accuracy is defined by the triplet *(Entity , Feature, Value), Et=(en,fe,va)*.

En - database entity (which has its origins in the conceived model , such as data table;

Fe - a property of the entity;

Va - a quantitative or qualitative measurement of the entity.

The more accuracy increases the minimum difference between the actual value(*val*) and the correct value (*val'*) is.

$$\textbf{\textit{Err}} = \left| val - val' \right|$$

should minimum if not zero.

That leads us to the idea that

$$\textbf{\textit{Min(ERR)=Min}} \left| \sum_{i=1}^{n} val_i - \sum_{i=1}^{n} val'_i \right| \qquad (*)$$

for each entity.

It is also important to mention that data conversions are very useful but they should be followed by verifications. One of the instruments that makes data transformation easier is the data adapter. If the data provided by the web controls value field is string type , and the field in the table related to that control is int type, data conversion is compulsory to prevent errors.

If not detected in due time, the eventual errors produced may lead to an alteration of the intermediate results and consequently to the alteration of the general results.

Still, detected at the end of the elaboration process, the time of the error making will be very difficult to find out and therefore the correction will be much more difficult.

If not detected at all during the process of testing, the final results will be absolutely wrong and very hard to correct, only after minute and long lasting work.

In order to avoid incorrect taking decision by the managerial staff, it is advisable to check from time to time, using test sets of data if errors have occurred and if so, they overcame a certain amount of significance.
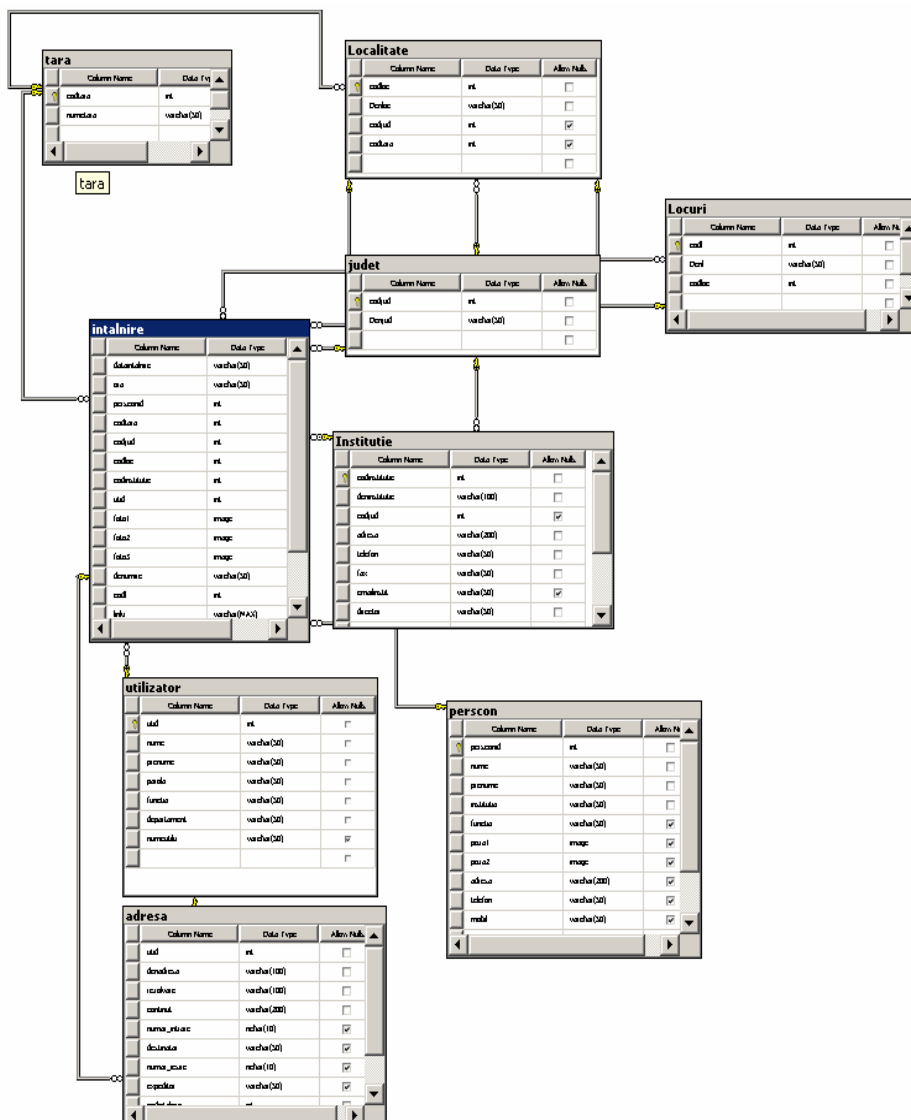

Figure 1. Database entity-relationship diagram

For example, *Convert.ToInt32 (DropDownList1.SElectedItem.Value);* makes the conversion from a data type of string to an integer of 32 bits, avoiding string type variables to be put into calculating formulae, which may lead to a definite alteration of the results as well.

The following concepts should be taken into account in the process of Internet Databases application development: **understandability, completeness, flexibility,**

reliability and data protection. The last concept has been dealt with in the first chapter of the article.

Understandability involves that developers should think from user perspectives and try to empathize with his problems, technological fear and rejection of the new. When changes are made, the target group is trained to cooperate with the new application.

Completeness is another important web database quality metric. An application is complete when all the items from the model (such as the tables and columns from the database diagram) correspond to the user requirements. At every step (prototype) it is recommended to study if the stage is complete and if is not there should be made some corrections. In our application, there have been omitted some user requirements such as the storage of official letters that are very important for the business meetings.

A table was added and it has been related to the other tables in the relational model. Problems also appear if the end user requirements are not well defined , which leads the development team to confusion. The target group must be interviewed in the development process to find out that the partial outputs are accurate.

The application is also flexible. **Flexibility** makes an application economically efficient. The number of changes in the future is not supposed to be very big and also changes must be easy to accomplish. The more the application is flexible, the lower cost of changing and upgrading is.

It is also very important for the application to be **reliable** during the life cycle. We hope that our application would maintain all its functions and procedures for a long period of time without any irremediable, beyond repair errors. It is desired that it would work for a long period of time. The index of reliability is

$$I_r = \frac{ndat_c}{ndat}$$

where

$ndat_c$ represents the number of datasets that generated the results;

$ndat$, the total number of datasets. A good application is realized when the indicator has the value greater than 0,78.

The target group consists of the employees of each department of the Chamber; they have been interviewed from the beginning and on the course of the entire development process. The predictive test results have been included in the table 1.

Table 1. The target group's predictive test results

Department	No of users	No of queries	Good responses	Updates	Errors
Business Information	5	35	30	2	2
Fairs and Exhibitions	12	50	46	8	3
Human Resources	3	35	33	4	0
Public relations	5	70	69	3	0

There are 25 users which made 190 queries in one week. Average queries per employee is 190/25=7,6 queries per user. The number of good responses is 178.

The index of reliability is 0,93 meaning that the software application is very good .

Conclusions

This paper related some theoretical themes, studied by many authors in computer science, with the role of describing the quality aspects of web relational databases .

The statistic calculations are important for each software metric on all stages of the project realization because they show the exact level of precision and report the percentage in which the initial objectives have been realized.

It is also useful to relate statistic results from each step to the next step of the project, every step, including analysis and data modeling having the same significance level for the objective's realizations.

Economic efficiency must also be achieved by making a reliable application that would help the managerial board to take appropriate decisions and not to put them in difficulty.

The study is also interdisciplinary, because of the different approaches, economic, technical and social. In the future, the employees will have to answer to questions, about the usability of the programs and the development team will take them into consideration.

References

1. Arsanjani, A. **Empowering the Business Analyst for on Demand Computing,** IBM Systems Journal, Vol.44, nr.1, 2005
2. Block, E., Hiemstra, D., Choenni, S. **Predicting the cost quality trade off for information retrieval queries: Facilitating Databases and query optimization,** www.utwente.nl, 21.11.2006
3. IBM Info center, www.publib.boulder.ibm.com, 28.11.2006
4. Ivan, I., Popa, M., Popescu, Al. **The Aggregation of the Text Entities,** Economic Computation and Economic Cybernetics Studies and Research 38, no. 1-4, 2004, pag.37–50
5. Ivan, I., Saha, P. **Quality Characteristics of the Internet Applications,** Software 2003 - ASE Printing House, 2004
6. Ivan, I., Visoiu, A. **Economic Models Basis,** ASE Publishing House, Bucharest, 2005
7. Platini, M., Genaro, M., Calero, C. **Data Model metrics,** www.uclm.es, 20.11.2006

[1] Codifications of references:

[ARSA05]	Arsanjani, A. **Empowering the Business Analyst for on Demand Computing,** IBM Systems Journal, Vol.44, nr.1, 2005
[BLOHI06]	Block, E., Hiemstra, D., Choenni, S. **Predicting the cost quality trade off for information retrieval queries: Facilitating Databases and query optimization,** www.utwente.nl, 21.11.2006
[IVSA03]	Ivan, I., Saha, P. **Quality Characteristics of the Internet Applications,** Software 2003 - ASE Printing House, 2004
[IVPO04]	Ivan, I., Popa, M., Popescu, Al. **The Aggregation of the Text Entities,** Economic Computation and Economic Cybernetics Studies and Research 38, no. 1-4, 2004, pag.37–50
[IVVI05]	Ivan, I., Visoiu, A. **Economic Models Basis,** ASE Publishing House, Bucharest, 2005
[PIGECA06]	Platini, M., Genaro, M., Calero, C. **Data Model metrics,** www.uclm.es, 20.11.2006
[www1]	**IBM Info center,** www.publib.boulder.ibm.com, 28.11.2006