DATA CERTIFICATION

(Research in Progress)

Ion IVAN,
Professor, Ph.D.
Academy of Economic Studies, Bucharest
e-mail: ivan@fx.ro

Paul POCATILU,
Assistant
Academy of Economic Studies, Bucharest
e-mail: paul@fx.ro

Teodor MIHAI,
Software Engineer
Storm Technology Ltd, Galway, Ireland
e-mail: teddy@storm.ie

Costel STANCA,
Lecturer
Constantza Maritime University e-mail:
cstanca@neptun.imc.ro

ABSTRACT *The widespread of computer applications mainly covers three aspects the increasing number of users for a given software product, the quality of the results obtained from processing initial data, and the quality of data streams in e-business, e-commerce, web.*
A change in methodology is needed in order to obtain industrial results with guaranteed characteristics of quality. Within this context an essential role is played by data, which needs to be complete, correct, comparable and real-time obtainable.

Keywords: *certification, data testing, data quality, information quality*.

1. Introduction

As on the stock market or financial market the error in a rumor or allegation arguing that a certain bank or company are unstable will show indirectly through the bank's capability to face all withdrawal requests or the stock market reaction regarding share prices. Only judging after effects it can be proven that the information was erroneous, a rumor to render unstable a financial system.
To certify an application input data means to evaluate the main characteristics of data, such as: completeness, correctitude, credibility, accuracy, security of access and relevance.
Certifying the results of a process involves verifying supplementary correlations compared to the characteristics of initial data, the existence of a defined structure, and the correctitude of processing in all used cases.
Data certification is even more important as the effects of incorrect data are frequently of an indirect nature. As an example, if the dimensions for making a certain component are incorrect the direct effect is that the component doesn't fit in the product, and then it is discovered that the design dimensions are incorrect. The error in data is discovered due to impossibility to use the product, hence indirect.

2. Data Certification

Certifying data is a complex process that includes any kind of data, while keeping in mind their usefulness and importance.
An important side of data certification is database certification. Certifying databases includes: establishing the concordance between reported and real content, verifying the existent quantity of data, and verifying the concordance between initial and current documents.
For example the certification of a juridical database takes into consideration the time frame, document types, data sources, staff in charge of transferring data into computer memory and

verifying it, methods of searching etc. Certifying such a database can be accomplished step by step or by sampling.
The process of certifying data is strongly connected to the process of testing data. Testing data means undergoing a set of stages in order to discover quality characteristics for a given data collection. Testing data shows if they are representative, complete, correct and accurate.
Testing data becomes important when software users have already obtained the hardware and software needed to process data and therefore input data is the only factor with influence over the final results.
The results of data testing help restructuring documents, redimensioning the effort of writing data in these documents and creating special methods to input data in order to render operational classical or Internet-centered applications.

3. Web page certification

The necessity to certify web pages is given by the large Internet audience. In many situations web pages have presented false or incomplete information with negative effects. In this case certifications means to establish if between the web page and reality is a perfect correspondence. Certifying web pages involves certifying the web page producer or the web page content validity.
For example a train schedule, holiday pricing in a travel agency web site, official information should be guaranteed and certified by the institutions handling the respective information. Data certification is vital for e-commerce. It is also vital in the banking system.
Certified web developers are those who have certified qualities and abilities to produce web pages conforming to all characteristics of data quality.
In this context we introduce the notion of *electronic notary*. This is an authority whose main responsibility is to certify information published in web pages.
The quality of data stored in on-line databases must be ensured since their users browse this information as a prelude to certain activities (purchasing goods, services, education enrollment, getting information from digital libraries, extracting data from official statistics). This is why a convention to protect Internet applications on a large scale is needed, and the elaboration of standards that define the obligation to ensure completeness, correctitude, reliability, accuracy, credibility, exactitude and relevance of data at levels that render it operational. Even if at the beginning such regulations are not perfect they are perfectible.

4. Conclusions

Certifying software companies, software, hardware, data, and web pages needs to have the same level of exigency. When the software is of a high quality and the hardware ensures the necessary processing resources the results of the application is strictly dependent on the input data. Input data varies as quantity on every activation of a software product. Loading a database (large data volume) is different from querying a database. Multi-user environments make data certification even more valuable since errors have a high transfer speed and multi-directional effects.
The credibility of web applications is given by the capacity of its producers to ensure certification of data included by these applications.
A new domain is born, the management of data certification to ensure an efficient and coherent environment for the processes related to increasing data quality.

This page was left blank intentionally.