Analysis on the Metrics used in Optimizing Electronic Business based on Learning Techniques

Irina-Steliana STAN, Ion-Sorin STROE
University of Economic Studies, Bucharest, Romania
irina.stan@outlook.com, soso@soso.ro

The present paper proposes a methodology of analyzing the metrics related to electronic business. The drafts of the optimizing models include KPIs that can highlight the business specific, if only they are integrated by using learning-based techniques. Having set the most important and high-impact elements of the business, the models should get in the end the link between them, by automating business flows. The human resource will be found in the situation of collaborating more and more with the optimizing models which will translate into high quality decisions followed by profitability increase.
Keywords: *electronic business, learning systems, KPIs, optimization, performance, models, decision.*

1 Introduction

IT has been a component with major implications in the structure of all businesses even a century ago. Those companies that have treated this factor with high importance, by investing strategically, had a significantly higher rate of profitability in the long term under all internal departments.

Predictions for 2013 [1] show that IT spending will increase by 5.7%, exceeding the nominal amount of 2.1 billion dollars, due to the evolution of cloud technology, Big Data - from search to discovery and prediction (a classification, analysis and prediction of the data), specific platforms and social systems.

Having a developed market in financial terms, it is imperative that companies seek IT technologies in order to lead to the fulfillment of a plurality of objectives such as:

- Early detection and removal of socio-economic crisis, of system issues, of processes systematically affecting the company, penetrating to the highest level;
- Design an efficient and sustainable system, having a larger yield;
- Remove harmful processes and system gaps;
- Adapt business to changes independent from socio-economic environment.

In other words, in terms of investment in IT innovation, we distinguish three types of business:

- *Classical*. This category includes those businesses for which information technologies don't have a major influence, which gives the company a static kind by taking fundamental decisions in a slow pace;
- *Hybrid*. This includes all those companies that have discovered the advantages of using IT side, which try to automate and optimize by integrating informatics solutions. Their objective is that the technology should have a growing influence on decision-making systems;
- *Online*. Within these businesses, the main communications channel is online, integrating all technical innovations in the field.

Analyzing the evolution of the number of users of Internet services, there is a spectacular increase of 600% compared with 2000, meaning that in a population of about 7 billion, 2.5 billion frequently

accessed the Internet, having the penetration rate of 35% [2].
The statistics presented above reveal a new direction in terms of channel of communication between business and customers, so that companies that are open to new technologies should base their foundation in online environment.
It is worth mentioning that the online environment is a relatively new branch of information technology, the state of knowledge in an interdisciplinary approach, is at an average level, which is why research is needed in this area so as to develop integrated models to substantiate it.
Electronic business affairs do not eliminate the essential elements of classical businesses, but translate them so that models previously used in economics, finance, marketing and management, remain valid and used in the business in one way or another, the focus being on the information technology-based model.
Orientation on this topic leads to an increase in the average yield on long term, automatically on profitability too, the best examples coming from the analysis among the most profitable companies.
In the Top 100 Most Profitable Companies in 2012, the majority are those which adopted technologies or which exclusively perform activities in online environment, such as Apple (No. 3), Microsoft (No. 4), Intel (No. 12), Google (No. 18), Oracle (No. 23), Apache (No. 39) [3].
"One of the most important factors in the company is the technology" [4], according to a study made by IBM from over 1,700 company board of directors of the top companies, as presented in **Fig. 1**.

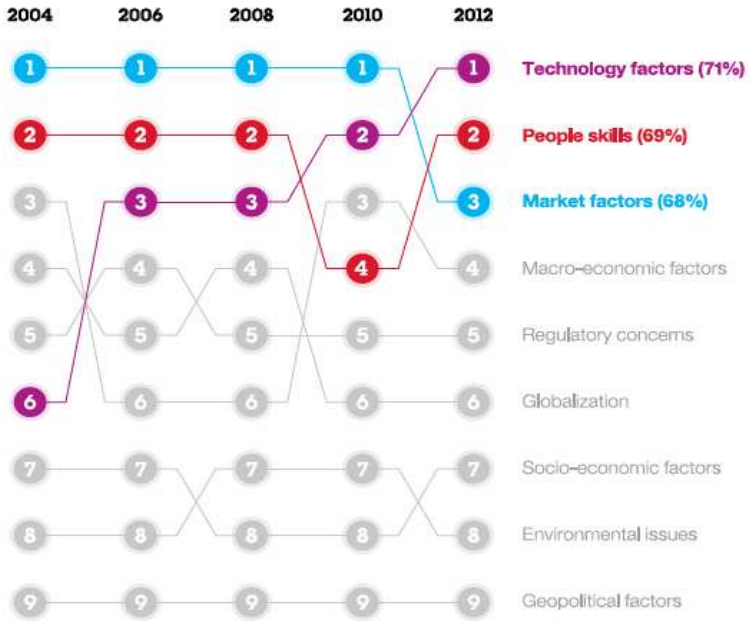

Fig. 1. "Technology pushes to the top of all the external forces that could impact their organization over the next 3 to 5 years, CEOs now see technology change as most critical." (Source: [4])

Electronic business can be divided into several categories, depending on the actors involved, on the technologies used, services offered, etc. But these topics will be discussed later, as the paper is focusing on the general area of this type of business.

Once specific technologies are integrated in the electronic business, their advantages should be used in a more efficient manner.
An important aspect is the one regarding business automation, in order to eliminate possible errors in a very short time or even predict them and get to take high-level decisions in the idea of building a decision

plan and a long term strategy that will ensure consistent services by positioning prior over competition. All must be able to perform in real time, to be flexible, intuitive and reversible, fault tolerant.

Starting point to build such a project is the data, whether these initially do not exist or are in low numbers.

One way to implement the above specifications can be by engaging a learning based system.

The first step is to train the learning system with a set of initial information in order to give a starting point, while, in parallel, the data providing process continues.

"Jerusalem Declaration" [5] shows the importance of current society in collecting data: "We are entering the era of a high rate of production of information of physical, biological, environmental, social and economic systems. The recording, accessing, data mining and dissemination of this information affect in a crucial way the progress of knowledge of mankind in the next years. Scientists should design, explore and validate protocols for the access and use of this information able to maximize the access and freedom of research and meanwhile protect and respect the private nature of part of it."

If in the past the collection of information supposed a major contribution of human capital by using expensive technology resources, now it can be obtained in a relatively short time, at a cost that would justify the investment.

Currently, obtaining important information that has major contributions can be made in real time due to interaction between several systems, by using their capacity of analysis and processing. As an example, more and more predictive models are used [6] with regards to economic future, election results and other socio-economic developments [7]. Another example is the "Google Dengue Trends" which uses information received from users based on their search on the search engine in order to determine the risk of disease by region due to mosquitos' stings [8].

Considering that the system is intended to be as general as possible, the need to cover vast areas of information makes the collected data diverse and heterogeneous. This leads to the need to build a set of minimum requirements in terms of data analysis and efficiency in processing because we aim to have a system which will obtain realistic results.

In the era of "Big Data" [9] there is no longer a question of how to obtain data but its analysis and processing. The International Data Corporation (IDC) shows that the data provided by users are exponentially growing, current information exceeding the storage capacity of 1.8 billion gigabytes, with a growth rate of 9 in less than five years [9].

The learning system must therefore include complex processing algorithm, which:

- Finds relevant information from many data sources;
- Makes a logical integration of information from different sources;
- Removes useless information;
- Focuses on continuous data homogenization, even on predefined steps;
- Assigns data and results to individual, giving a real meaning depending on the context;
- Provides an interface easily to integrate into other services which addresses to a wide variety of users so that their knowledge level does not represent an impediment in using the system.

The data collected at the moment can be classified according to many criteria but a special category is the human factor in terms of the contribution of this component in decision making and thus in making unwanted events.

The financial crisis of the past years was due to result of collective human behavior in the social environment on a large scale, which is why understanding individual typology based on historical information can facilitate similar items expectation of future crises disasters or other events.

Current technology provides an understanding of the human factor-based systems with very high accuracy, so as to design a complete system that would suggest strategies for future actions.
Stored data collection and analysis should:

- Determine pattern of collective actions in the social environment relying on a set of highly heterogeneous data;
- Define models of human behavior;
- Share responsibility for decisions, depending on cases;
- Determine initial points that favored specific actions;
- Show relationships between different models and other subjective but important factors such as human intuition, simulations of specific cases or test cases [10].

A very important thing to note is that we can automate many of the business processes using learning techniques on data but cannot remove the human factor, as it has the ability to determine alarm signals that the system might not be able to detect, which is why the objective is not to eliminate human intake but total optimization of the tasks that can be tech, leading to lower costs through dynamic adaptation to different situations and by replacing certain responsibilities that would be done on a longer period of time by individuals.

2. Overview on performance indicators

Term optimization involves the adoption of a mechanism that can be applied to electronic business and contains various internal processes that support adaptability to changes in a competitive context, fundamental decision-making dynamics, so, increasing profitability.
The main element of the optimization process must have as its starting point the possibility of measurement / analysis of all business flows.
The heterogeneity of different types of electronic business requires offering for adopting a generalized optimization solution which will homogenize the various input and output in order to get an easy to understand structure and adaptable to change, similar to the approach of Model Driven Architecture (MDA) [11]. According to this view, the end result is to get the definition, analysis, generalization and use of assessment tools sets for various categories of activities in the company.
MDA proposes an incremental framework that includes different theories such as:

- *The Goal - Question - Metric* (GQM) is a concept that assumes that measuring various indicators can be optimal only after the list of objectives is completed. Also, according to this principle, all information related to the objectives should be analyzed and used for recording the progress;
- *Balanced Scorecard* (BSC) is a strategic planning and management system used in various types of business that provides uniform internal activities to the vision and strategy, improving various impact factors;
- *Business Process Model and Notation* (BPMN) is a graphical representation of various business processes;
- *Semantics of Business Vocabulary and Business Rules* (SBVR) is a standard that defines the vocabulary and rules related to business terms, rules and procedures;
- Other theories.

Whatever core framework was used, proposed or already existing, the primary objective is not only to remain at the theoretical level but to materialize the optimization models by applying them on as many as possible electronic businesses.
The first step is to identify the key elements of the firm which suffer over time minimal or major changes. Thus, we use a system of metrics and indicators that will summarize the facts.

By definition, a metric can identify the deviation or potential deviation from achieving a goal set.

The main categories of indicators will be:

- *Key performance indicators (KPIs).* These will measure the changes in terms of returns on different time intervals. Performance indicators will express the quality of decisions implemented. Having the evolutionary history of the entity and of the performance indicators, they will identify any deviation that was not performance. Means of identification, correlation and automation of these indicators will be presented later on in this article;
- *Key Risk Indicators (KRIs).* Are those statistical indicators that can synthesize the positioning of the company in terms of risk, which are dynamically revised and aim to provide consistent signals on changes with high influence [12]. This type of indicators are the safety features to prevent incorrect application of decision strategies. Risk indicators will be able to determine the times when certain items will change and lead or can lead to behavior that could divert an aggravating behavior and could act in unfavorable directions.

Proposed sequence for defining the optimization platform in terms of metrics.

Step 1. Will review all elements defining the flow of business and will create a weighted matrix to express their importance. Also in this step will classify the main categories that can influence company's performance.

Step 2. Based on previously defined categories and on decisional factors, will express a number of key indicators for each stream of business.

Step 3. Each indicator will have detailed important information needed in the framework:

- The need for adoption, as in **Fig. 2.**;
- Whether or not an indicator has high impact, it will be classified in risk category indicators;
- The influence level between indicators;
- Reporting interval - hourly, daily, weekly, monthly, yearly, etc.;
- The period of storage and / or use of historical data;
- Range of monitoring the evolution of the index.

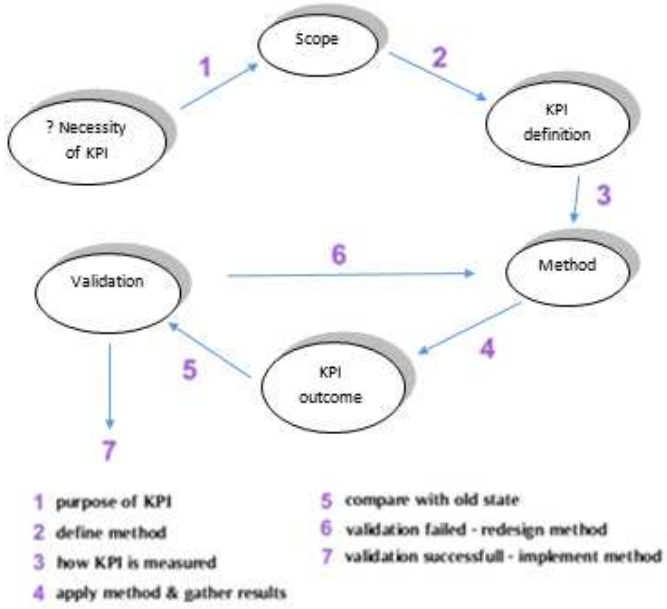

Fig. 2. Indicator analysis

Depending on the specific of category to be optimized, we will define optimization models that include previously defined indicators.

Fig. 3. illustrates the operation and application of two indicators.

Monitoring the risk state. Note that these performance indicators are at the same time risk indicators, which is why there are two levels - $MR(i_1)$ and $MR(i_2)$ - representing the values for indicators for which the risk status is triggered (Maximum Risk value). At the onset of risk status we will apply certain models of prevention and depending on how it will impact, will revert to the previous state. For example, if the indicator i_2 at time t_1 - $V(t_1, i_2)$ - triggers risk status, the core framework will manage time by an

eventual recovery of i_2 to the value recorded at time t_0.
Also, the complexity of the optimization algorithm will be determined in proportion with the influence of other risk indicators in creating the state.

- For i_1 and i_2 indicators, reporting intervals are different, weekly and monthly respectively, possibly due to their importance;
- It is worth mentioning that we are interested if the state of risk triggers only when the indicator is reporting. We can opt for a shorter reporting period but should take into consideration the nature of the indicator because changing the timeframe and fitting into alert level may be short-lived. Consequences can lead to misinterpretation of the indicator and beyond even taking wrong decisions. An example is highlighted in the chart below: the value V (t_3, i_2) is not included in the risk class, even if the index exceeds the maximum allowed by the next reporting period t_4.

Fig. 3. KPI evolution & evaluation

As a general conclusion on performance indicators, in order for them to bring a significant contribution to the optimization process they should be included in metrics whose utility will increase the yield rate and automatically improve profitability.

3. Optimization models for specific cases

The complexity of electronic business is similar to classical business, involving various fields, which combined create a flow system often difficult to track, control and predict. So, as part of the business, are elements related to the financial, economic, human, marketing, politics, etc.

The objective of this article is to take into account the business specific elements that develop their activity wholly or partly in the online environment, the reference categories being classified as follows:

- Software;
- Hardware;
- Human;
- Finance;
- Miscellaneous.

Optimization models to be proposed will be composed of metrics and indicators designed to partially or completely automate decision making system, having a dynamic opening, meaning fault-tolerant and adaptable. The goal is not to eliminate

the human factor but to provide assisted decision opportunities so that they act in time and understand the impact that decision might have.
Interaction between technology and the human factor will be an environment based on continuous improvement to ensure success. Only this approach will ensure efficient and early limiting of the human error by the monitoring performed on the computer system. Also, the individual will notice and correct possible deviations of the system before it produces results considered optimal while they are not.
In the following pages we propose models for the above mentioned categories.

3.1 Hardware

From the hardware point of view we can distinguish two subcategories, namely infrastructure, including communications channels, and the second one, of processing and storage of the resources.
As primary infrastructure can change based on the decisions taken by some important actors (large companies, governments, etc. - for example, all cars produced in the European Union have the obligation to incorporate the technology Anti-lock Braking System (ABS) [13]). Average users have low influence, which is why we will not discuss optimization techniques in this regard.
In terms of processing and storage of the resources, there is an adaptability of the market solutions according to the specific of the company. It assumes the use of Cloud Computing paradigm which allows scalability on infrastructure in a very short time (minutes). Increased storage and processing capacity correlated with lower price per unit allowed providing these general solutions but also personalized for each individual case [14], giving up the traditional solutions such as Virtual Private Server (VPS), Dedicated Servers, etc. [15].
Optimization will cover how hardware-solutions will be selected to meet the needs of online business platform and get a larger yield.
In terms of optimization, performance is perceived by users as high when the level of response from the platform is bigger, the maximum load being the instant one, in milliseconds. It is true that maintaining the performance level - level of satisfaction – involves other factors, which we will specify in other optimization categories, the main being software.
Website performance is a key factor in achieving user satisfaction and automatically converts them into active customers. At high level, Amazon shows [16] that a simple increase in the response time of just 100 milliseconds on the web platform results in decreased sales by 1%.
Hardware optimization goal is to obtain a model that:

- Minimizes the cost of hardware resources, by holding a level of satisfaction as high as possible;
- Justify the performance - response time per customer profitability.

Finally we get a relationship that will help us estimate the response time depending on profitability at a certain moment, taking into account historical data regarding various useful indicators such as:

Indicator	Description
ProcessingUnits	number of processing units used in Cloud for day X, month Y, year Z
StorageUnits	number of storage units used in Cloud for day X, month Y, year Z
TrafficUnits	number of traffic units used in Cloud for day X, month Y, year Z
VisitorNumber	number of platform visitors
CR	conversion rate = number of platform visitors who became clients/ total number of visitors
ClientIncome	income for each client
ResponseTime	response time (seconds) for each loaded page

Fig. 4. Miscellaneous hardware performance indicators

The model will serve to human part from the decision-making process in order to find answers based on existing information on questions such as "*What is the primary objective in terms of offered performance?*", "*What is the maximum cost that the company can support on the hardware infrastructure, when performances fall?*", "*What is the effect on limiting resources, meaning performance degradation, on the cost and the profitability per user?*".

The first step is to determine the causal relationship between *response time* (in seconds), meaning the *platform performance* and *profit per customer*.

Response time is based on indicators related to the number of sessions and requests in a given day, which involves a number of processing units, storage, traffic, etc.

Historical data modeling requires the use of an econometric model, the pattern resulting from the analysis of the number of requests for a certain period:

- *Requests per Day*: demand is not constant and is due to sociological factors. You can create oscillating trends driven by human factor availability to access platform, restrictions being imposed by the employer;
- *Requests per month*: in this case you can see a variable set of data, its nature being given by specific periods of each month, such as holidays, national events, etc.;
- *Requests per Year*: their number can vary due to political and economic actions, global or local, as a result of business performance.

Based on this analysis we conclude that the specific model is nonlinear. It requires using a sinusoidal model, starting from the general form:

$$Y_i = C + \alpha \sin(\omega T_i + \phi) + E_i$$

where:
C - is the constant defining the mean level;
α - is the wave amplitude for the model;
ω - is the frequency;
T_i- is time variable;
φ - is the phase;
E_i – is the error sequence in the sequence Y_i by approximating the model.

Translating into an econometric model for use, in this case, of historical data set, it will result in the following formula:

$$P = \sum_{i=m}^{n} (a_i * sin(b_i * RT + c_i) + E_i)$$

where:
P - is the estimation of customer profit, correlated with response time (RT);
m – is the starting year for the dataset involved in the model;
n – is the end year for the data set involved in the model;
RT – is the response time estimation, based on profit per customer (P);
a_i, b_i, c_i - model parameters used as training set;
E_i - is the error of approximation in the model.

Discovering the correlation between response time and profitability, human factor will have decision options in a range of project expenditures by adjusting the hardware infrastructure, namely the maximum number of processing units, storage and bandwidth used for a certain period.
Model results can be interpreted as a solution for a learning method. Basically, having the information available, the data is trained together with the links between them from time to time, the system being able to learn from previous periods and human decisions being taken so that ultimately the financial return is the maximum.

3.2 Software
The main element of this category is the online platform, the website, as an intermediary between business and customer flows.
Thus, performance indicators will be related to how online platform works and to most of the features coming from users. Indicators related to how to make and implement the site to make it functional, such as optimization by improving code written in a particular programming language or the use of specific utility programs, are connected to the infrastructure and can be included in the model from 3.1. We also exclude those indicators that can be included in the aforementioned category, as the demographic ones.
It may be created a variety of metrics [17] that can be used in an optimization model, for example:

- *Return of Investment* (ROI), conversions, subscribers;
- *Search Engine Optimization* (SEO): number of backlinks, quality of backlinks, pages, indexes, different ranks;
- *Usability*: time on site, bounce rate, links clicked, eyetracking, page visit, returning visitors;
- *Society*: on Twitter messages (tweets), Facebook likes.

In **Fig. 5.** we present three types of indicators that can justify the profitability per user in terms of software, the latter having a linear response correlated with the level of satisfaction. It is worth mentioning that a direct correlation between a variety of indicators and profitability can be found.

Indicator	Description
FunctionalityAccess	number of users that accessed the functionality in day X, month Y, year Z
FunctionalityIncome	number of users that become clients to that functionality in day X, month Y, year Z
UsedTechnology	a certain technology used by the visitor

Fig. 5. Miscellaneous software performance indicators

Choosing a linear model is motivated by the fact that the use of a given functionality is given by the customer satisfaction level for that utility.
The optimization model should assist the human in decision making process by answering questions such as *"What is the correlation between a specific user functionality and profit per user?", "How to reduce the cost of certain functionality by maintaining optimal levels of profitability?" "What consequences has removing a functionality within the financial structure?"*
The econometric linear model to correlate the functionality within the business and its profits will have the following form:

$$P = \sum_{i=m}^{n} (a_i * F + b_i + E_i)$$

where:

m – is the starting year for the dataset involved in the model;
n – is the end year for the data set involved in the model;
P - is the estimation of customer profit, correlated with accessing functionality (F);
F – is the estimation of accessing functionality, based on profit per customer (P);
a_i, b_i - model parameters used as training set;
E_i - is the error of approximation in the model.
Obtaining the correlation between functionality and profitability, human factor will get decision options in a range of project expenditures, by allocating funds for developing or disposing the service provided. The model determines the level of satisfaction of users.

4. Conclusions

In terms of optimization, it can extend across multiple categories and subcategories, including those ones already presented, financial, human, etc. Providing to the learning system a larger set of training data will give it the ability to almost completely automate decision-making process on matters of business. The quality of results on the learning process given from training the information will also be high as it is evidence-based and comprehensive.
Developing models presented in this article in future studies may lead to a refinement and determination of the specific of users, so that the optimization level will exceed expectations and investments will be made on items which confirmed the degree of profitability.

Acknowledgements

This work was co-financed by the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/134197 „Performance and excellence in doctoral and postdoctoral research in economic science from Romania”.

References

[1] Frank Gens (2013), *Competing on the 3rd Platform Opportunities at the Intersection of Mobile, Cloud, Social, and Big Data*, on-line: http://www.idc.com/research/Predictions13/downloadable/238044.pdf
[2] World Internet Users Statistics Usage and World PopulationStats (2013) , *Internet World Stats - Usage and Population Statistics.*, on-line: http://www.internetworldstats.com/stats.htm
[3] Fortune Magazine (2012), *Top companies: Most profitable*, on-line: http://money.cnn.com/magazines/fortune/fortune500/2012/performers/companies/profits/
[4] IBM Global Chief Executive Officer Study (2012), *Leading Through Connections*, on-line: http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03485usen/GBE03485USEN.PDF
[5] Kahng Byungnam (2009), *The Jerusalem Declaration - Declaration on Data Access, Use and Dissemination for Scientific Research*, on-line: http://portale.unipa.it/ocs/home/jerusalemSubscription.html
[6] Sunstein Cass R. (2006), *Infotopia: how many minds produce knowledge*
[7] Abramowicz Michael (2009), *Predictocracy: market mechanisms for public and private decision making* ;
[8] GOOGLE (2013), *Google Dengue Trends - Dengue trends around the world.*, on-line: http://www.google.org/denguetrends/
[9] John Gantz, David Reinsel (2011), *IDC - Extracting Value from Chaos*, on-line: http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

[10] D. Helbing, S. Balietti (2011), *From social simulation to integrative system design*, The European Physical Journal Special Topics, Volume 195, Issue 1, 69-100
[11] Gawel Bartlomiej, Skalna Iwona (2012), *Model Driven Architecture and classification of business rules modelling languages*, 2012 FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS (FEDCSIS), 949-952
[12] Coleman (2009), Les. *Risk strategies*
[13] Europa (2007), *EC type-approval system for motor vehicles*, on-line: http://europa.eu/legislation_summaries/internal_market/single_market_for_goods/motor_vehicles/motor_vehicles_technical_harmonisation/n26100_en.htm
[14] Arutyunov V.V. (2010), *Cloud computing: Its history of development, modern state, and future considerations.*, Scientific and Technical Information Processing 39, no. 3, 173-178
[15] STROE SORIN (2011), *MySQL databases as part of the Online Business, using a platform based on Linux*, Database Systems Journal, vol. II, no. 3, 3-13
[16] Michele Mazzucco (2010), *Towards autonomic service provisioning systems*, IEEE International Symposium on Cluster Computing and the Grid, 273-282
[17] Sean Power (2010), *Metrics 101: What to Measure on Your Website*, Velocity Web Performance and Operations Conference

Irina-Steliana STAN graduated the faculty and master organized by the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies, with graduation and dissertation topic-themed works in database and management fields. She is currently a PhD candidate in the same area, Economic Informatics. Nowadays the interest domains are related to decision processes, learning systems and improvement on economical business.

Ion-Sorin STROE graduated the faculty and master organized by the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies, with graduation and dissertation topic-themed works in the electronic business field. He is currently a PhD in the same area, Economic Informatics. From the environmental point of view he presents Word Wide Web interest in new technologies and platforms, both conceptually, software and hardware too, developing over time many large and profitable web project.