

# Big Data in Changes: Is big data bigger then sustainable development and research design?

MILENA JANAKOVA

Department of Informatics and Mathematics

Silesian University in Opava

Univerzitni nam. 1934, Karvina

CZECH REPUBLIC

mija@opf.slu.cz, <https://sites.google.com/site/maninformatics/z-mija>

*Abstract:* - This paper is oriented on research design with links on big data in the global information society. Big data creates a phenomenon in modern society, where firms, organizations, and individuals must work with data, information, and knowledge. At first glance, it is standard matter, but the deluge of data brings many difficulties to instability. Big data is oriented around the derivation of needed and correct information with knowledge from stored data. The solution is to apply optimal research design as a multidimensional approach on big data based on defined preferences. Key is a complete perception of reality based on an optimal research design. Such approach works with many objects and different contexts using available analytical disciplines (layers). Needed spectrum of suitable analytical disciplines is wide. Presented design is focused on browsing selected layers via disciplines such as Artificial Intelligence, Business Intelligence, Customer Intelligence, Competitive Intelligence and Swarm Intelligence. For active work with data via various layers, a good helper is simulation and multidimensional approach. This approach is useful for data warehouses as well as bringing high value for support of sustainable development. The benefit is creating unexpected relations between specified layers with benefits from information technology, and controversies of the global information society. A natural request is intuitive movement from one layer to another in the form of a zoom to needed data.

*Key-Words:* - Big data, design, information, information technology, research, sustainable development.

## 1 Introduction

The general perspective of this paper is oriented on information technology (IT) that influences all fields of human activities. The IT development proceeds very rapidly, and to have actual information about a suitable solution based on information technology is often difficult. Social networks and the Internet bring some help. [1] The Internet, by default, offers access to many information sources and references around the world with support of suitable Internet architecture. [2] This reality is the reason “why information technology lives with people”, and “why information technology involves people lives”. The positive benefit is variability and open access.

Data and information is around us all the time. Organizations, firms and people literally swim in a sea of information. A correct and suitable orientation is difficult for everyone. Visualization brings partial help [3] based on videos, pictures or graphs. This approach is important because standard rows of reports are confusing. The question is easy: How to derive a conclusion from the thousands and thousands of rows of records? In this reality, the visualization helps in every possible form.

Visualization is one condition for better work with information. [4] The second request is to apply the appropriate applications. In today’s global society, it is easy to select a needed application by preferences and actual conditions. The complexity is often created by confusing requirements of IT users, and these requirements must reflect research design with links on software construction perspective. [5] The IT users need information immediately in an appropriate format. Intuitive navigation is preferred. These trends are well-known via various surveys and questionnaires. [6]

A large volume of accessed data naturally leads to big data. The “big data” concept [7] is often used by governments, managers, leaders, organizers, analysts, developers, and of course, standard IT users. In many cases, they work only with needed data based on a traditional approach with support operating and database systems, BI and CRM products. The standard is oriented on the familiar www interface, MS Office products like MS Excel or MS Word.

For successful decision-making, various data and information is needed in various formats. An active IT user approach not only needs standard

information about income and expenses, or profits and losses, but also other data and information related to customer preferences, security society, and sustainable development based on stability. Such needs are visible in many crises and difficulties that involve people's lives around the world. In this actuality, when information technology lives with people and information technology has an influence on people's lives, information technology must help and inspire active solutions to existing difficulties. Big data is a good start point to do it; therefore, the objective of the paper is oriented on big data with needed research design. Optimal research design for work with big data is challenge for everyday life in the global information society. For responsible work on selected topic, the structure of this paper forms chapters focused on perception of context, information technology for large data, problems with unexpected connections between data, solution based on simulation and combination, new challenges for research and global society with necessary conclusion.

## 2 A Dynamic and Compete Perception of Context

In business, companies, organizations and individuals are searching for a gap in the market for successful implementation of offered goods and services. The key is innovation and an upgrade with a regard to customer demand. The reason is clear – the more attractive a product, the higher the demand from customers and the more customers are willing to pay. The global market is overstocked, and also consumers are looking for optimal goods or services with the support of the Internet.

It is not enough to have an attractive product. There is a need to listen to customer requirements, and offer expanded programs of cooperation with customers, contractors, partners, competitions, science and also all other interested persons. The mutual cooperation must be dynamic with a higher level of quality. This request is standard because new approaches, products and services, and of course, new technologies are being introduced.

A number of companies, organizations and individuals have to face at present new conditions for realized activities. This environment has its mirror in science and its design. And, conversely, science affects our reality. It is a truly indispensable connection for further vision and sustainable development support. The question of sustainable development [8] is a basic theme for human life in

the future. There are connections with expected topics such as:

- Biodiversity.
- The life environment.  
(water and atmosphere)
- Non-renewable resources.  
(metals, coal and petroleum)
- Renewable resources.  
(solar, wind, geothermal and biomass energy sources)

Such approach is extended to other topics [9] which are also important for human life:

- The optimal utility of goods and services to society.
- The promotion of employment.
- The support of acceptable limits for a growing economy.
- The wise utilization of all resources with support innovation and an active approach.

Acceptable limits were, are, and will be a question for further development and science in all fields of human activities. The difficulty is that an inviolable limit for one is an acceptable limit for another. The actual decision depends on the point of view which is based on an accepted abstraction level, the adopted boundaries of the researched system, defined objects and their properties, and the created connections between objects and the ability to change the arrangement of the adopted design.

For innovation and research, design is important. [10] Interest is focused on suitable architecture and approach involving active implementation of new or innovative products and services. The arrangement of the adopted design is the first step to innovation for sustainable development. Such a design has to be verified and realized in practice via an optimal method. A good helper is simulation with IT support and models of data, information, knowledge, and intelligence [11].

Information technology has a major influence on all disciplines and realized activities. The reason is that people work with great volumes of data and information. This volume exceeds our senses, and the new concept is created in the form of “a Digital Universe”. Based on a research study, a digital universe such as a stack of tablets creates two-thirds of the distance to the Moon in 2013. This figure should reach approximately 6.6 stacks from the Earth to the Moon by 2020. [12] Data availability is the primary power to work with new approaches for increasing the value of our activities based on large data.

### 3 Information Technology for Large Data

Large data and increasing value, the Internet of Thinking with benefits and opportunities, needs and imperatives, weaknesses and negatives are at the centre of numerous analysis and surveys. [13] The primary topic of interest is focused on the connection between the digital universe and:

- The expanding.
- The influence on business.
- Opportunities.
- Power.
- The connection around the world.
- Transformation.
- Visualization.

A number of people ask how large the digital universe is and how can I work with this universe. The size of this digital universe is easy to calculate and estimate. With a verified analysis, the digital universe is doubling in size every two years and between 2013 and 2020 will grow from 4.4 trillion gigabytes to 44 trillion gigabytes. [14]

This space is large enough for intensive searching of a new connection between stored data from various sources. A broad view on big data brings diverse and heterogeneous information and methods for solving existing tasks in the global information society. For example, in Twitter [15], Big Data Science tweets are oriented on many and many interests in relation to big data. The discussed topics are:

- 40 Excel Tricks.
- Advanced Analytics at Scale with IBM.
- Being a data scientist in a small country: challenges and solutions.
- Big Data and Analytics: Learn the Key to Business Survival (IBM Whitepaper).
- Big Data Applications to Financial Institutions.
- Big Data Challenges in the Enterprise.
- Challenge of the week.
- Data Integrity - A Sequence of Words Lost in the World of Big Data.
- Data Scientist versus Business Analyst.
- Deriving Value with Data Visualization Tools.
- Discover How to Balance Increasing Data Volumes.
- Doing Data Science.
- How NoSQL Fundamentally Changed Machine Learning.
- List of NoSQL Databases.
- Outliers in Logistic Regression.

- Predictive Analytics in Campaign Management.
- The 7 Most Unusual Applications of Big Data You've Ever Seen.
- The Data Science Toolkit - taking your first steps towards becoming a Data Scientist.
- Tutorial: How to determine the quality and correctness of classification models?
- Word Clouds of Big Data, Data Science and Other Buzz Words.

Upon closer examination, the number of retweets (answers to a tweet) is from 0 to 12 tweets, with an average of 4.32 retweets to one tweet. Is this a lot tweets with retweets, and data with information? What is the motivation and meaning behind this information sharing? The above-mentioned list shows only tweets that were sent over 24 hours through twitter.com. This example of a big volume of data is not unique. There are other examples of information sharing from organizations and individuals with a much larger volume of data as CERN, the European Organization for Nuclear Research, Geneva [16]

Above mentioned examples show only marginally the volume of data and information in society. The global information society works, and will work, with many and many data and information in various relations and formats. There are models that analyzes billion web pages with billion hyperlinks, and there models bring good orientation within the volume of available data. [17, 18] In the future, the description of such complexity will be harder and harder based on aggregation levels, data format, and indexing with network format. Firms, organizations, and individuals must process a large volume data and information as soon as possible by defined preferences and priorities.

The positive fact is that IT users select from many applications and tools focused on big data. [19] There is open-source as well as commercial software. Existing discussion groups and blogs support correct implementation. Such information sharing brings needed advice and orientation to a topic with all benefits and challenges, but from the complete view, the searching out of unexpected connections between data is not all that easy.

### 4 Problems with Unexpected Connections between Data

Connection searching between stored data is a standard activity realized with the support of information technology. This task has links to

monitoring, alert processes and security needs. For standard IT users, there are natural requests based on implemented IT support. The well-known tasks are:

- Analysis of priorities and the overall time required to solve user requests.
- Detailed overview of activities.
- Key Performance Indicators and their development.
- List of transactions.
- Performance by price development.
- Relationship sales and margins.
- Sales development by prices, customers, location, or product and time.

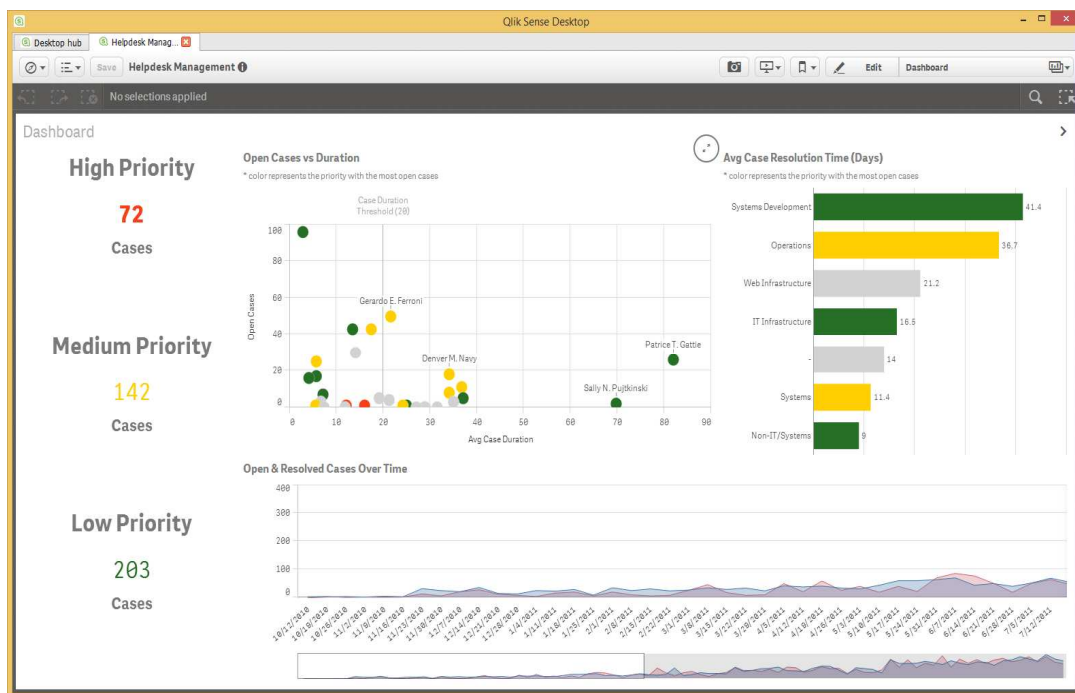
Needed analyses are shown via dashboards with various graphs, lists, or summaries and filters. The Qlik Sense Desktop [20] is a good example which provides a complete approach.

For example, the average time to solve a request (in days for system development, standard surgery, web infrastructure, IT operations, web infrastructure, or other requirements outside of IT) use dimension (Case Owner Group), facts (Avg Case Duration), and filter (by average time and case owner group).

IT users use priorities for better resolution in detail with an advantage. There are also:

- A number of cases with high, medium and low priority.
- Open and resolved cases by the time required to solve
- Open cases and their term solutions.

For these analyses, the dimensions create a Case Owner or Date, and the facts create Avg Case Duration, Open Cases, Overall, Total New Cases, Total Closed Cases. Please see Fig. 1.



**Fig. 1: A window for active work with data in QlikSense Desktop.**

It is not a problem to design an optimal dashboard for default and well-known connections with stored data. The question is going further and designing new unexpected connections. A suitable way is to attempt to combine existing analyses based on dimensions and facts. This approach is haphazard and requires cooperation between business managers and IT specialists on large data and data science.

Many IT users (and also business managers) prefer individual work with data; therefore, Excel is very popular for work with data. Everyone is familiar with contingency tablets. For suitable work with large data, it is insufficient. Initial activities have to be focused on clarification of specified requirements and needs:

- To identify business limits.
- To remove these limits.

- To look for suitable replacements.
- To look for improved results.

Business limits constitute a list that includes the unfulfilled processes and the limiting business associated with the negative activities. For active identification of business limits, good helpers are the following questions:

- What do you think you are not so successful at? In what area would you like to improve?
- What processes of your business would you specify as unqualified?
- What new processes are you afraid of realizing?
- What are your favourite processes?
- When you wish to create an optimal business, what do you realize?

Removal of existing limits has close links with difficult conditions in a dynamic and competitive market economy. The interest is often focused on the question of the price of the offered products, and searching for suitable substitutes, innovations via processes, products, or organization of work. The default part of this question is the correct evaluation of specified limits and their influence on competition and customers. Looking for suitable replacements is focused on realization of selected activities in a different way in order to produce optimal results using available sources, processes and skills. The aim is to define processes for change that will be able to produce positive results. Looking for improved results represents knowledge about existing processes which are suitable for change based on knowledge management. [21] The interconnection of all the specified requirements and needs creates a good starting-point for accepting business into a new internal map of reality.

Based on this internal map of reality, needed objects are easy to create. The benefit object approach [22] is the ability to define optimal properties and methods for active work and searching for unexpected connections between data. Good results provide a combination object approach with multidimensional methodology [23]. This approach uses various levels as adopted abstraction, IT support, needed time, but also interest which is focused on data, information, processes, functions, organization, finance, management, software, and hardware. Based on such information, simulation provides enough space for a combination of defined objects with needed connections in appropriate data models. [24]

## 5 A Solution Based on Simulation and Combination

The main interest is oriented on optimal research design in relation to searching out of unexpected information for support of sustainable development. From the perspective of the global information society, the focus is concentrated on big data in dynamic changes of existing conditions. Big data influences realized processes based on IT support; realized processes in society also influence big data. The continuum of IT development needs stability with regards to analysis, design, and implementation of selected software and applications.

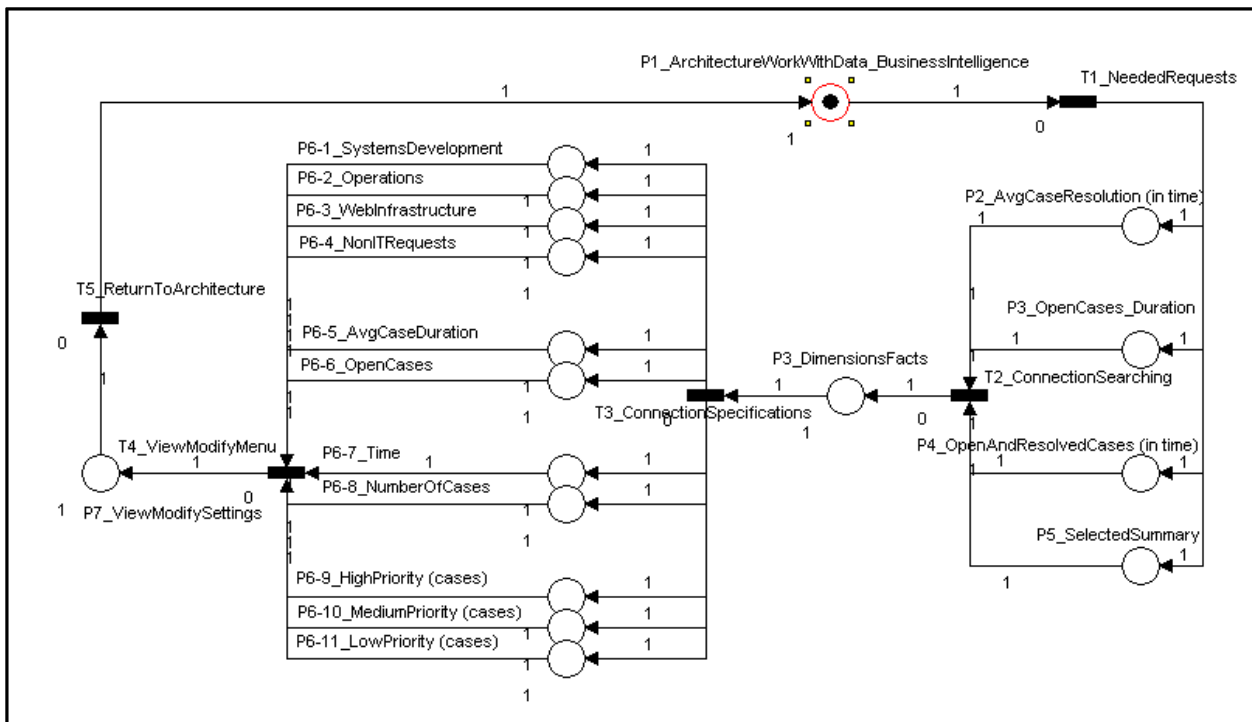
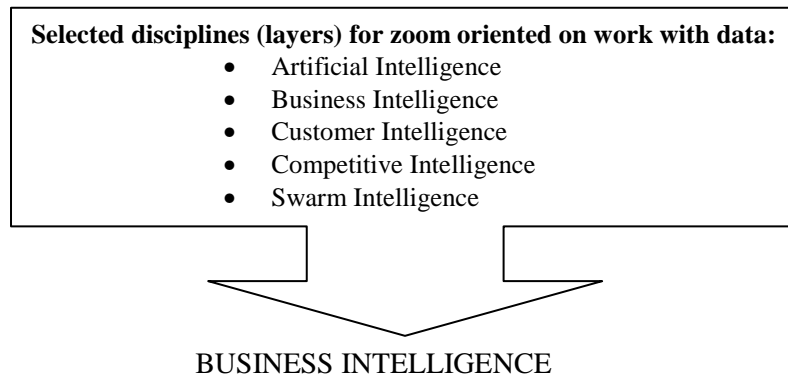
Optimal work with data and information is complex. It is not a pile of data, information and knowledge in a vacuum. Immediate results bring well-known an analysis of selected topics by experience, advices of suppliers implemented IT products, but also other users' experiences with support the Internet. The optimal solution has two ways:

- To create a team of experts with good orientation on all disciplines for suitable impact on work with data.
- To create a team of a few experts with good orientation on selected disciplines for suitable impact on work with data.

The first way needs many resources and support rather large teams in the opulent and global companies. Cooperation is the key for their success. The second way needs also resources a work with team, but range is variable according to existing options. This approach has good links on agile methods and agile development tools. There is also risk that given team does not use all the opportunities of current knowledge.

Such risk is limited by a continuous shift between disciplines supported by the zoom on the present question. The benefit is work with only a few selected objects, and obtained results are compared via further analysis with orientation on other disciplines by preferences. This approach reduces errors caused by the omission of important objects from various disciplines. If important object is deprived then the substantive context is naturally addressed via other zoom.

Please see Fig. 2.



**Fig. 2: An architecture for active work with data based on selected dimensions and facts (Business Intelligence).**

The multidimensional view [25] on big data is aimed at a wide area of interests and active work with controversies and difficulties. The benefit is in the active connection between existing controversies and changes in the global society.

Suitable view on big data is important for the global information society. People wish to know the near and distant future. The aim is an adequate response to changes, and preventing unexpected fluctuations and crises with the support of intelligence. Searching unexpected connections between data uses individual zoom as layer for work with data. Further work continues via other zoom, for example, based on Customer Intelligence.

## 6 New Challenges for Research and a Global Society

Current existing needs show that the future trend will be oriented on closer connections with the controversies and difficulties of the global information society. This is a great challenge for all societies and research design for big data. Such a trend brings other positive factors like active development of new approaches for optimal implementation of information technology in all realized human activities. The uncertainly matter will play a more important role in the variable layers of big data. This layers use needed background for processes with data, information, and knowledge.

The special challenge is automation and self-service in societies, firms, organizations, and individuals by preferences. There are differences in size and orientation, time spent with big data, continuing education, and monitoring current trends. This high instability needs regular actualization for a background of created simulation. The changed simulation is run for searching for a hidden relationship.

## 7 Conclusion

Big data is all around us. All societies, firms, organizations, and individuals must be resistant to this data flood. The aim is to resist existing controversies and crises with the preferred way of sustainable development. The big data area is extensive with a focus on analyses, statistics, challenges, transforming processes, visualization, learning, databases, and predictions. Such a topic is hard to keep in good condition as a unified entity.

In this situation, it is difficult to search for unexpected connections between data. Existing complex reality is reflected in many forms of intelligence such as artificial, business, customers, competition and also swarm. The suitable recommendation is to implement a multi-dimensional view on big data with partial views based on simulation with a combination of individual intelligences as additional layers. The transition has to be easy such as with Petri Nets based on defined objects and their properties.

### References:

- [1] *Top 15 Most Popular Social Networking Sites*, 2015. From <http://www.ebizmba.com/articles/social-networking-websites>.
- [2] Battilotti, S. Priscoli, F. D., Giorgi, C. G., Pietrabissa, A., Monaco, S., Panfili, M., Canale, S., Suraci, V., Approaches for Future Internet architecture design and Quality of Experience (QoE) Control, *WSEAS Transactions on Communications*, Vol. 14, Art. #9. E-ISSN: 2224-2864, pp. 62-73, 2015.
- [3] Clark, D. *Data Visualization Is The Future - Here's Why*, 2014. From <http://www.forbes.com/sites/dorieclark/2014/03/10/data-visualization-is-the-future-heres-why/>.
- [4] Simon, P. *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*. NJ: John Wiley Sons Inc, ISBN 978-1-118-79438-8, 2014.
- [5] Yong, L. Communication Complexities: A Software Construction Perspective. *WSEAS Transactions on Computers*, Volume 11. E-ISSN: 2224-2872, pp. 227-240, 2012.
- [6] *Information Technology Market Reports*, 2015. From [http://www.reportlinker.com/report/best-keywords/Information%20Technology?utm\\_source=adwords1&utm\\_medium=cpc&utm\\_campaign=High\\_Tech\\_And\\_Media&utm\\_adgroup=Information\\_Technology\\_All&gclid=CNne3s6NksYCFSLmwgodACYAXg](http://www.reportlinker.com/report/best-keywords/Information%20Technology?utm_source=adwords1&utm_medium=cpc&utm_campaign=High_Tech_And_Media&utm_adgroup=Information_Technology_All&gclid=CNne3s6NksYCFSLmwgodACYAXg).
- [7] Mayer-Schönberger, V., Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. UK: Eamon Dolan/Mariner Books, ISBN-10 0544227751, 2014.
- [8] *Sustainable Development Information*, 2015. From: <http://www.sustainabledevelopment.info.com/>.
- [9] *Sustainable Development - Additional Topics*, 2015. From: <http://science.jrank.org/pages/6642/Sustainable-Development.html>.
- [10] *Design for Innovation - European Commission*, 2015. From: [http://ec.europa.eu/enterprise/policies/innovation/policy/design-creativity/index\\_en.htm](http://ec.europa.eu/enterprise/policies/innovation/policy/design-creativity/index_en.htm).
- [11] Wang, Y. Formal Cognitive Models of Data, Information, Knowledge, and Intelligence, *WSEAS Transactions on Computers*, Vol. 14, Art. #75. E-ISSN: 2224-2872, pp. 770-781, 2015.
- [12] *Data Growth, Business Opportunities, and the IT Imperatives - Executive Summary*, 2014. From: <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>.
- [13] *EMC Digital Universe Study with Research and analysis by ICD*, 2014. From: <http://www.emc.com/leadership/digital-universe/index.htm>.
- [14] Price, G. *How Large is the Digital Universe?*, 2014. From: <http://www.infodocket.com/2014/04/16/how-large-is-the-digital-universe-how-fast-is-it-growing-2014-emc-digital-universe-study-now-available/>.
- [15] *Big Data Science on twitter*, 2015. From twitter.com.
- [16] *CERN on twitter*, 2015. From twitter.com.
- [17] Granville, V. *Big data set - 3.5 billion web pages - made available for all of us*, 2013. From <http://www.bigdatanews.com/profiles/blogs/big-data-set-3-5-billion-web-pages-made-available-for-all-of-us>.
- [18] Santhanakumar, M., Christopher Columbus, C. Web Usage Based Analysis of Web Pages Using RapidMiner. *WSEAS Transactions on Computers*, Volume 14, Art. #46. E-ISSN: 2224-2872, pp. 455-464, 2015.

- [19] *Top 27 Free Data Analysis Software*, 2015. From <http://www.predictiveanalyticstoday.com/top-data-analysis-software/>.
- [20] *Qlik Sense Desktop*, 2015. From: <http://www.qlik.com/us/explore/products/sense>.
- [21] Concas, G., Pani, F. E., Lunesu, M. I. A New Approach for Knowledge Management and Optimization using an Open Source Repository, *WSEAS Transactions on Information Science and Applications*, Issue 5, Vol. 10. E-ISSN: 2224-3402, pp. 139-148, 2013.
- [22] Metz, S. *Practical Object-Oriented Design in Ruby: An Agile Primer*, Addison-Wesley Professional, 2012.
- [23] Cravero, A., Sepúlveda, S. Multidimensional Design Paradigms for Data Warehouses: A Systematic Mapping Study, *Journal of Software Engineering and Applications*, No. 7, pp. 53-61, 2014.
- [24] Jiacheng, Z., Chonglong, W., Xinglin, G. A Dynamic Landslide Simulation Algorithm Based on Multitask Spatio-Temporal Data Model. *WSEAS Transactions on Computers*, Volume 14, 2015, Art. #67. E-ISSN: 2224-2872, pp. 658-667, 2015.
- [25] Dzemyda, G., Kurasova, O., Zilinskas, J. *Multidimensional Data Visualization: Methods and Applications*. Springer, ISBN-13 978-1489990006, 2014.