Some of text analytics applications in higher education institutions

Predrag Stolić[1], Snežana Stolić and Aleksandra Milosavljević[2]

[1] Technical faculty in Bor, Bor, Serbia
[2] Mining and metallurgy institute Bor, Bor, Serbia
e-mail pstolic@tfbor.bg.ac.rs

Abstract: *Previous knowledge of the institution environment were based on the use of structured data. In recent years, unstructured data are more used by predictive analysis. The impact of unstructured data is a dominant in different types of texts, so the one of the leading approaches to analysis and obtaining the necessary information is based on the use of text analytics. A real use-value text analytics is shown in this paper with particular emphasis on some of the possible aspects of text analytics using in the area of higher education institutions.*

Keywords: *competence; teachers; modern media*

1. INTRODUCTION

The common characteristic of all higher education institutions in the world is the realization of study programs in a very dynamic environment. Mentioned dynamism is a result of the various parameters combination, starting from sociological, via technical and technological, all the way to economic. Unlike higher education in a traditional sense, where all the improvements carried out periodically and with a certain delay due to the tendencies that expressed the society then, modern higher education implies continuous improvement in all parts of its processes in order to keep competitiveness.

Modern Higher Education set the goal of modernization of study programs as a continuous process that ensures excellence in achieving the expected learning outcomes and necessary competences for future graduates at all levels of academic and professional studies [1]. As the education of students performed in order to qualify students for future tasks that will be done within the framework of qualification, the modern study programs must be designed to respond to the demands issued by the current labor market in the economic, technological, social and scientific sense. In order to achieve the set goals in terms of compliance between study program and current market, in processes of creation, design, reform, verification and validation of study programs must be included hierarchically all relevant national authorities, scientific and research community, employees within the higher education institutions, but also students and, as a very important component, those who will absorb the final product of higher education, employers as end users.

Data sources that can be used in a previous analysis in order to achieve competitive study programs at all levels of study are found everywhere within the eco-system in which higher

education institutions achieve their work. The question is how to extract relevant and useful data from large volumes of data which are generated on a daily basis at high speed and how that data processed in the shortest possible period of time so that forehand analysis could carried out in order to provide instruments for defining new directions of development and improvement of study programs. Traditional analysis and appliances that were used until recently gave certain results but still some delay is observed in adequate reactions to the more volatile labor market.

In recent years in the context of the development of information and communication technology a set of techniques known as Big Data are appeared which allows the processing of large amounts of information in a reasonable period of time with adequate resources and relevant data recognition, forehand analysis and the corresponding presentation of the results of completed analytical actions. In the following chapters will be presented one of those techniques known under the name text analytics and the possibilities of its application to developing institutions within higher education.

2. UNSTRUCTURED DATA

For years, the world of information systems and information technology relied on a strictly structured data stored in databases. Almost all of the data which the computer systems encountered, from raw data that were the result of some type of data acquisition till the results of the data processing, regardless of the type, quantity and usability, were subordinated to a some kind of database management system. The most common type databases were relational database, based on well-known mathematical postulates of relations and collections. By information technology improvement, an approach based on relational databases proved to be very limited and in certain moments inadequate for modern challenges which the world of new information technologies brings to customers in the near and distant future [2].

In accordance with the mentioned above, opposed to strictly structured data, the process of using unstructured and semi-structured data is initiated. It can be said that for the expansion of unstructured data using, instead of traditional structured, responsible three highly coupled, important factors: the mass computer communications, the explosion of social networks using and the tendency of man to express himself in a world of two previous using a way close to the everyday expression and natural speech. Data are generated by enormous speed with appereance of extreme number of data sources (blogs, comments on sites, social networks, various documents in the cloud, etc.), but due to unstructured mode of presentation, the most of the data seems irrelevant or inappropriate for further processing over them.

By further analysis can be determined that the corresponding sets of unstructured data, which are considered to carry irrelevant data, can lead to significant, relevant findings depending on the ways in which data is treated. Therefore, it is necessary to change approache to data analysis. Traditional techniques and technologies used in the analysis and processing of data must be abandoned in whole or modified to allow actions to be performed on the domain of unstructured data.

3. TEXT DATA MINING AND TEXT ANALYTICS

Expansion of the use of information and communication technologies in carrying out tasks aimed to the modern business, texts of various types, shapes and lengths are beginning to be more importance as carriers of information in the form of various reports, letters, memos, or the whole brochures, instructions. Therefore, the essential value of the possibility to obtain

data from various types of texts is recognized and searching for implementation methods of mentioned is started.

At the beginning, data extraction from text consisted of a simple counting words, or the appearance of given words in a text. Over time, tasks become much more complicated and the complex analytical operations on the text itself are perfomed, which led to the creation of new computing discipline marked as text data mining (often in the use is the term text mining). Text data mining means the data relevation from the text or processing of unstructured text information [3] in order to be able to use information from the text in further flows of data processing such as importing relevant data to the database, processing in the statistical and machine learning algorithms and similar [4]. Typical tasks that are accomplished by text data mining are various types of marketing analysis through processing questionnaires and forms, automatic classification of text, such as automatic detection of unwanted messages (spam messages) [5], the analysis of user behavior on social networks via sentiments [6], the analysis of information published on web sites, blogs, pages on the social networks and similar [7].

Text analytics means a consolidated set of linguistic, statistical and machine learning methods for extracting information from text sources purposes. Obtained information using text analytics, are intended for use in the context of Big Data systems, business intelligence, various statistical analysis and processing, prediction of behavior and future events [8]. Today in the use prevails the term text analytics as a synonym for text data mining, although there are some differences in the use of these two terms. However these differences are becoming lesser and common characteristics are much more emphasized so text analytics, in the world of modern business, are more notable as the dominant term.

4. TEXT ANALYTICS FOR THE PURPOSES OF HIGHER EDUCATION INSTITUTIONS

4.1. Detection of potential forgeries

The use of text analytics in terms of higher education institutions in the world and in our country is based primarily on an analysis of various types of documents (scientific papers, theses, doctoral theses, etc.) in order to identify potential forgeries and their authors. If we consider that, a few years ago, the average number of published scientific papers per year is several million worldwide, and that most of these papers now available at some of the ways in text format, the use of text analytics in this segment is necessary link in the whole process [9,10]. In this way, the integrity of the institution strengthens through raising the quality of scientific papers and reducing incidence of possible forgeries.

Fig. 1 shows a schematic view of the system that uses text analytics to analyze the scientific content in order to detect possible forgeries. This system could be fully applied in the analysis of the content of textbooks and other supporting materials which are used in the context of study programs implementation. However, due to laws in the field of copyright and related rights, system would be limited in work because it could analyze only those issues which are not protected by copyright, or could analyze only parts of the editions that the publisher made publicly available. Functionality of the system will be questionable so these systems are for now still rarely used in the field of analysis of textbooks and other similar literature.
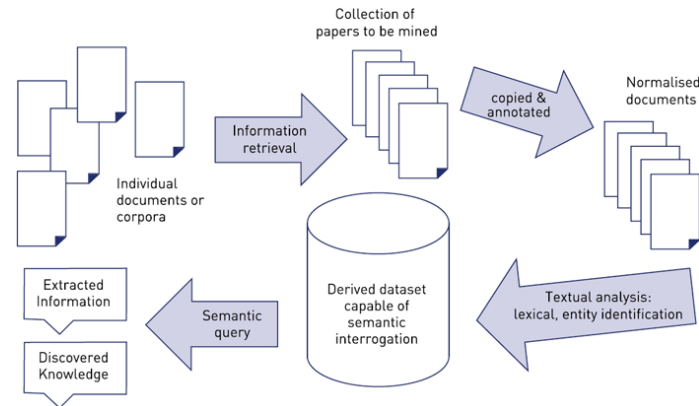
Figure 1. *Schematic representation of the processes involved in text analytics for scientific content [9]*

4.2. Analysis of the market environment

Today's higher education institutions achieve their goals in terms of a very variable market environment. Mentioned changes are very dynamic and it is necessary to develop mechanisms to monitor and identify them in order to students, after completing their education on these institutions, could fit into the corresponding market trends. Traditional techniques for market monitoring and segmentation for projection purposes and harmonization of higher education programs give certain results, but there is a large degree of inertia that has manifestation after the application of the results to higher education programs in terms of their modernization and harmonization. Again, we have a certain amount of obsolescence and non-compliance with current situation in the market.

By application of text analytics in domain of market environment analysis, minimization of the traditional approach characterized problems is performed. Neutralizing these factors is realized by converting the text data into the data of outstanding organizational and business value that can be automatically accessed and then perform analysis and interpretation, understanding them and act in accordance with the information detected from the electronic texts that are the subject of accession and analysis [11]. This approach does now, during various daily analyzes, that advantages could be taken from the resources of social networks, electronic questionnaires, communications made through e-mail and also from hidden resources in the form of stored information in the portfolios of companies, job ads, labor markets and similar domains which, in the traditional sense, did not reveal much, or reveal only in some kind of indirect contact with them [12]. Now, these resources are directly accessed by text analytics and seemingly hidden information becomes a valuable tool in the identification of all the advantages and potential problems which could have influence in sphere of higher education by market changes.

4.3. Evaluation of students satisfaction

During the last years, there is a expressed focus of higher education in the Republic of Serbia to monitor student satisfaction, especially after the forming of the Commission for Accreditation and Quality Assurance. Various types of surveys, external evaluation and self-

evaluation have become a key segment and tool for improvement of higher education institutions. However, in practice, real problems have occurred in the form of obtaining data whose reliability and usefulness are debatable. For example, in the evaluation of subject teachers appeared expressed preference that the students get a lower grade to teacher who focuses on discipline, and opposed to that, high score will receive a teacher who tolerates extremely frequent absences from lectures, delays and so on. Also, students continue to express doubts about the anonymity of each survey conducted by higher education institutions.

On the contrary, students expressed a preference to objectively assess the condition and work of higher education institutions in the context of less formal communication which take place, for example, in various forums and social networks segments. It is believed that this kind of data and information can lead to a more realistic and comprehensive look to all potential problems that students encounter in the context of higher education institutions and thus, after identification, appropriate mechanisms could be made to improve the work of institutions themselves [13-15]. In this case, the text analytics is a way to process these information for the purpose of allocation larger set of data and knowledge about the satisfaction of students and obtaining a more authentic and better picture about state in institutions of higher education, and in higher education in general [16].

Fig. 2 provides a simplified (generalized) schematic representation of text analytics phases which are processed when analyzing student satisfaction on the basis of a separated text regardless of the source of the text (social networks, forums, email, etc.).

Figure 2. *Analysis of students opinions using text analytics*

Certain text are analyzed and, after the separation of sentences, further analysis of each sentence partially is processed in order to identify all the parts of each obtained sentence [17]. Clearly defined analytical procedures are processed over obtained parts in order to characterize and evaluate each of the parts [18]. The objective of analytical processing represents an isolated opinion from the initial text. This procedure is known as sentiment analysis [12] and it presents one complex process that beside text analytics, as a core of the procedure, includes natural language processing (NLP) and computational linguistics.

The advantage of this approach will be presented by a short example based on evaluation of subject. Within some evaluation survey, student can evaluate the subject with the grade B (very good). This kind of assessment can be used for the purpose of further statistical analysis and evaluation. However, a far better conclusion about the subject matter can be obtained for example from the text about the subject written by a student at a forum. For example, a student explained to another student on the forum fundamentals of the subject, what is learned, what is tested, how the matter is innovative, how the subject is modern and similar and at the end of posted item wrote the text using slang "In a word this subject is awesome!!!". Using text analytics from this post we can obtain information, for example, that a student who wrote the text is feeling enthusiastic about the subject, that it is encouraged him, that recommends the subjet to other colleagues and a whole other set of relevant information. As

was mentioned above, we get the whole set of information which can be used for measuring the success of the subject through the student satisfaction, as opposed to an initial one grade from the survey (B - very good).

5. CONCLUSION

Higher education institutions achieve their activity in conditions of highly variable and dynamic environment. Traditional methods of analysis can not fully follow the latest trends which are currently presented in the field of higher education and the immediate environment that surrounds it. To ensure further competitiveness it is necessary to introduce modern methods in all aspects of higher education institutions work.

Text analytics is a method that can give tremendous contribution to the analysis of higher education institutions work. This paper presents a few aspects of the use of text analytics in the field of higher education, while the real use is practically unlimited. Essentially, as a man almost all processes in a particular language and in the form of a text, the text analytics can be shaped into an essential tool in various segments of work.

The use of text analytics in the field of higher education will constantly growing. That projections presented primarily the use of text analytics in forehand analysis of the factors that affecting the work of higher education institutions, and using that analysis will be possible to predict future trends and to form a correct answer to this trends. Therefore text analytics as a tool of the future will enable better segmentation, better results, better positioning of higher education institutions.

It is expected that the introduction of text analytics in domain of higher education lead to the significantly reducing of time required for performing analysis. Economically analysis will become more acceptable and accessible to a variety of higher education institutions. Organizationally, the analysis will be applicable at all levels of the organizational structure, from departments within the faculty to the university senate under almost identical conditions.

There is no doubt that future higher education will apply text analytics, through the different scope of action, in almost all segments of its work.

REFERENCES

[1] Ministarstvo prosvete i nauke Vlade Republike Srbije. (2012). *Strategija razvoja obrazovanja u Srbiji do 2020. godine.* Službeni glasnik Republike Srbije broj 107/2012. Belgrade, Serbia: Službeni glasnik.

[2] Zaki, A. K. (2014). *NoSQL databases: new millennium database for big data, big users, cloud computing and its security challenges.* International Journal of Research in Engineering and Technology, Vol. 3(3), pp 403-409. doi: 10.15623/ijret.2014.0315080

[3] He, W. (2013). *Examining students' online interaction in a live video streaming environment using data mining and text mining.* Computers in Human Behavior, Vol. 29(1), pp 90-102. doi: 10.1016/j.chb.2012.07.020

[4] StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, Oklahoma, USA: StatSoft.

[5] Manning, C.D., Raghavan, P. & Schütze, H. (2009). An Introduction to Information Retrieval. Cambridge, England, UK: Cambridge University Press.

[6] Irfan, R. (2004). A Survey on Text Mining in Social Networks. The Knowledge Engineering Review, Volume 30, pp 157-170. doi: 10.1017/S0269888914000277

[7]   Bail, C. (2014). *The cultural environment: measuring culture with big data*. Theory and Society, Vol. 43(3-4), pp 465-482. doi: 10.1007/s11186-014-9216-5

[8]   Ellaway, R. H., Pusic, M. V., Galbraith, R. M. & Cameron, T. (2014). *Developing the role of big data and analytics in health professional education*. Medical Teacher, Vol. 36(3), pp 216-222. doi: 10.3109/0142159X.2014.874553

[9]   JISC. (2012). The Value and Benefit of Text Mining to UK Further and Higher Education. Bristol, England, UK: JISC.

[10] Chen, N.-S., Kinshuk, Wei C.-W. & Chen H.-J. (2008). *Mining e-Learning domain concept map from academic articles*. Computers & Education, Vol. 50(3), pp 1009-1021. doi: 10.1016/j.compedu.2006.10.001

[11] SAS. (2012). SAS Text Analytics - Derive more value from unstructured content. Cary, North Carolina, USA: SAS Institute Inc.

[12] Chen, H., Chiang, R. H. L. & Storey V. C. (2012). *Business intelligence and a nalytics: from big data to big impact*. MIS Quarterly, Vol. 36(4), pp 1165-1188.

[13] Picciano, A. G. (2012). *The Evolution of Big Data and Learning Analytics in American Higher Education*. Journal of Asynchronous Learning Networks, Vol. 16(3), pp 9-20.

[14] Eynon, R. (2013). *The rise of Big Data: what does it mean for education, technology, and media research?*. Learning, Media and Technology, Vol. 38(3), pp 237-240, doi: 10.1080/17439884.2013.771783

[15] Romero, C. & Ventura, S. (2013). *Data Mining in Education*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 3(1), pp 12-27. doi: 10.1002/widm.1075

[16] Daniel, B. (2015). *Big Datas and analytics in higher education: Opportunities and challenges*. British Journal of Educational Technology, Vol. 46(5), pp 904-920. doi: 10.1111/bjet.12230

[17] Abdous, M. & He, W. (2011). *Using text mining to uncover students' technology-related problems in live video streaming*. British Journal of Educational Technology, Vol. 42(1), pp 40-49. doi:10.1111/j.1467-8535.2009.00980.x

[18] Xu, Y. & Reynolds, N. (2012). *Using Text Mining Techniques to Analyze Students' Written Responses to a Teacher Leadership Dilemma*. International Journal of Computer Theory and Engineering, Vol. 4(4), pp 575-578. doi: 10.7763/IJCTE.2012.V4.535