Application of Big Data in Historical Science

Nataliia Khymytsia[1], Taras Ustyianovych[2]

1. Social Communication and Information Science Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12, E-mail: nhymytsa@gmail.com

2. Social Communication and Information Science Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12, E-mail: ustyk5@gmail.com.

Abstract – The importance of the research is that it presents the usage of Big Data and history. The prospects and the main directions of the usage of the Big Data in historical science were analysed.

Keywords – history, historiography, techniques, Big Data, computing, the Internet.

I. Introduction

The exponential increase of electronic resources that is observed at the beginning of the 21st century opens new prospects for historical science and historiography researches. The Internet global computer network offers a lot of sources for historians: electronic copies of archival documents, scanned historical sources, materials of archaeological and ethnographic expeditions, collections of historical photos. Electronic journals, various scientific articles and books, and virtual conferences are accessible in "cyberspace".

However, the modern Ukrainian historiography lags far behind European and American trends. For example, the usage of the Big Data west methodological toolkit in domestic historical science is an elitist phenomenon.

II. Analysis of research and publications

There are several definitions of "Big Data" term. One of them describes Big Data as a volume of data impossible to process in a traditional way because of its amount. Another one says that it is a phenomenal acceleration of data accumulation and its complication. The third definition claims that it is a set of tools that allows working with data, regardless of its type and amount.

The Cambridge English Dictionary defines it as: «very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods» [1].

Some methods and techniques of analysis of Big Data historians have already been successfully used in research activities. Practical methods for modeling historical processes are applicated by Russian scientist L. Borodkin. [3]. Approaches of mathematical modeling of long-term social and historical processes, of theoretical history, historical macrosociology, of creation and analysis of historical databases, of studies of social evolution, historical demography are combined by cliodynamics that develops mathematical models "of age" socio-demographic cycles [4]. As research and analytical tools in cliometric studies monitoring information systems are also used. Combining methods of economic theory, the use of quantitative modeling techniques lets describe and explain historical processes and phenomena in the economic development of the society in general and its communities living in certain administrative areas [5,6].

III. The prospects of the usage of Big Data in historical science

In the professional activity of a historian, a special definition of Big Data is used, namely various tools and methods of processing structured and unstructured arrays of information. Big arrays of information collected by a historian are presented in the form of unstructured data, such as text, images, photos.

Big Data provides a possibility of preservation of big arrays of information, allows preserving "all data", without worrying about what part of data is actual for the next analytical activity. To get useful information, you need to carry out processing of big arrays of information.

On the basis of application of Big Data tools and methods in processing structured and unstructured arrays of information, wide possibilities of optimization historical knowledge are being created.

Today, Big Data has the potential to become an important sector of the IT industry. Big Data has been developing for twenty years, and experts claim that it will be difficult to imagine our life without its application. We will see a new era of Big Data soon. History can be efficiently used in this process. The combination of historical science and this modern technology can help us find a new, prospect field of study.

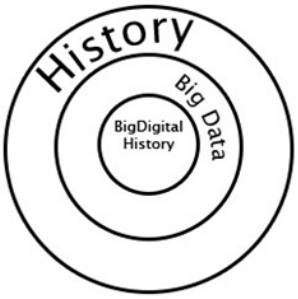

Fig.1Creation of Big Digital History.

Circle 1: "History" – the main object of the research, it contains a lot of information that should be processed quickly. The main component of the scheme.

Circle 2: "Big Data" – the data processing system that is connected with the first circle. It can process a lot of data for a short period of time. It is also a very important component. It includes large communities, network and modern technologies.

Circle 3: "Big Digital History" – the result of the combination of the first and second circles. It creates a new field of science that needs detailed studying and can be used for processing historical data. Its development will give a good result for the improving of historical science in general and will help preserve a lot of valuable materials.

Human need to take a direct part in this process. Their task is to select information according to the following characteristics: topicality, adequacy and usefulness (the

most important one). Without a human's activity, though insignificant, Big Data in historical science won't be able to develop, and "stagnation" can occur. This approach of Big Data and history will make them more productive. Historian will extract complex knowledge from the smallest crumbs of evidence that history has left behind. Big Data offers a complementary path to knowledge. Hiscorian will start with complicated work but then it would be simplier to work with the material. Big Data will show us more historiograhphy and preserve a lot of value information. That can also help to mend a "split" between them which may get much bigger. But Big Data has some disadvantages. Historians that are not able to use Digital History may be in the danger of getting mired in data.

The main task of the usage of Big Data is a simplification of the work in historical science in order to be able to process large volumes of information. Three main steps enabling a close contact between Big Data and history must be highlighted:

Step 1 – Digitization. Books have to be turned into images.

Step 2 – Transcription. Transformation into text.

Step 3 – Preservation of data. That allows making data more accessible and carrying out more thorough research. This is needed to be made carefully because of a value of the works.

Humans will get access to a great amount of information and will be able to use it for the development of historical science. American historian Roy Rosenzweig made a huge contribution to the development of Digital History in USA. His work in digital history was recognized in 2003 with the Richard W. Lyman Award for "outstanding achievement in the use of information technology to advance teaching in the humanities.". The historian said: "The injunction of traditional historians to look at "everything" cannot survive in a digital era in which "everything" has survived. There is no sense to conceal something that cannot be concealed. Big Data made data more accessible for learning history and analyzing the past. It will help solve an important number of questions". And that is really so. With the development of Big Data, everyone will be able to get necessary information in just a few minutes. Free and open access to the data will be provided. We'll be able to learn about all the most important events in human history and not only about the events belonging to only some short period of it.

We have to realize the necessity of this process which will safe our history and a great amount of historical researches. The aim is to highlight immediately what often requires careful thought. Humans without big data cannot do that, so this modern technology is needed. History has many kinds of materials but now digital material is much better for her.

Also, big data opens up new perspectives for the development and optimization of historiographical research today. For example, The Seeley Historical library, Great Britain, Cambridge, contains 95,000 historiographic works. They create Big Data. Computers can read this amount of information quickly, whereas a human would spend all his life on that. Even though that the sets of data can be very big, the information that consists in them has much less size. Processing of such amount of information requires special methods. One of these is Data Mining (intelligent data analysis). Data Mining is gaining knowledge throug data research, processing of data samples, and clearing and collecting data. It will open us a new way of studying history and searching for information if we use it right. Data mining is an interdisciplinary subfield of computer science, involving the methods at the intersection of artificial intelligence, machine learning.

Conclusion

Modern fundamental changes in the number and quality of historical information particularly actualize a necessity in application of concepts and ideas of Western European and American historiography in domestic researches, extension of scientific tools of Ukrainian historians. The key benefit of Big Data technology usage is that it allows comparing the historical information that was earlier incomparable. That can change our view of historical events and make data accessible. And that will develop the history science and make researches full of earlier unknown information.

References

[1] Cambridge Learner's Dictionary. Fourth edition. // Cambridge University Press. 2013.

[2] E.B. Belova, L.I. Borodkin, I. M. Garskova, T.F.Izmestyeva, V.V. Lazarev, A.I. Tikhonov Computerized statistical analysis for historians / Ed. by L.I. Borodkin and I.M. Garskova. – M.: 1999. – 187 p.

[3] P.V. Turchin Historical Dynamics: Towards the Theoretical History. – M.: «URSS», 2007.

[4] S. Golyb, N. Khymytsia. The use of multi-level modeling in the cliometric studies process / S. Golyb, N.Khymytsia // Proceedings XIII-th International Conference "Modern Problems of Radio Engineering, Telecommunications and Computer Science" (TCSET'2016) : Lviv, February 23-26, 2016. – Lviv-Slavske, Ukraine. – p. 733-735.

[5] S. Golub, N. Khymytsya. Assessment of utility models in the process of cliometric researches / S. Golub, N. Khymytsya // Proceedings of the 5th International Scientific Conference "Information, communication and society": May 19-21, 2016. – Lviv: Publishing house at Lviv Polytechnic, 2016. – p.238-240.

[6] N.Khymytsia, S.Lisina, O.Morushko, P.Zhezhnych. Analysis of Computer-based Methods for Processing Historical Information / N. Khymytsia and other // Advances in Intelligent Systems and Computing: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017, September 5-9 Lviv, Ukraine, Shakhovska N. (Ed.). – Springer International Publishing: 2017.– Volume 1.– p. 365-367.