Veslava Osinska
Nicolaus Copernicus University in Toruń, Poland
wieo@umk.pl

Visualizing the scientific information nowadays: the problems and challenges

Abstract. In recent years, a comparably fresh research field — information visualization has become commonly available for the researchers of all specialties. Information or knowledge maps play a role of interface for the analysis and intensive study of scientific community and knowledge domains development. The popularity of visualization techniques and interdisciplinary framework has resulted in many problems that have not been solved since the field had emerged. The article introduces the instrumental problems and challenges in this field. Exposing the functions information visualization allows to understand the difficulties and barriers within the whole visualizing process. A particular example of insight into the Polish science map is considered in the context of a new knowledge.

Keywords: information visualization, visualization problems, science maps, Infovis.

Introduction

What is typical of the present times is that visualization has been adapted not only in statistical research but also in many fields not necessarily related to science. The visual language has become essential means of communication in the advertisement and any media applications. Its potential was first noted by Edward Tufte, whose book (1990) is being included in the basic tutorials of data visualization. Modern science is based on open access to the knowledge resources which are extremely fast growing now. These dynamics replicates the common trend of big data phenomena as well as technologies development and accessibility to the wide audience. To subjugate information overload in science, the use of various data visualization techniques and methods is applied.

Visualization application based on scholars related data has particularly expanded in the recent two decades (Chen, 2001; Börner 2010). It is a proof of developing various tools and methods needed to represent the scientific knowledge (Osinska, 2016). On the other hand, such features as intensive social communication, open software standard and online resources like open scientific data and platforms foster the dissemination of information (*Infovis*) visualization in science. As Katy Börner calls her book — album illustrated by sophisticated maps of knowledge "Everyone can map" (Börner, 2015). She included there some exquisite examples that were designed thanks to the working groups which are mainly represented interdisciplinary. Openness (big choice of open source tools or online services, open data), availability (user friendly software, Web data), as well as socialization (tools and experience sharing, user communities, published galleries) describe visualization nowadays.

In classic terms, the information visualization was described as a new discipline aimed to interactively represent the abstract data (Card, Mackinlay and Shneiderman, 1999). According to papers, the *Infovis* researchers position this field between science and art (Kosara, 2007; Osinska 2010). Current user requirements rely on no only aesthetics, but also functionality. Thus, visualisation is also referring to the technology because of being measured for utility, for instance, struggling with the problem of navigation (van Wijk, 2005) and it is processing the goods of contemporary world — information (Osinska, 2016). Infovis practitioners, besides the statistical knowledge, must be up-to-date with the graphical applications and design, hence they should possess a broad set of skills that is a desirable feature in the contemporary science as well as other domains of life.

Multifunctionality

Better understanding of the role of visualization in human activity leads to considering all possible functions that might be performed. Furthermore, the practical context of *Infovis* allows to find and elaborate the problems connected with the processes constructing the graphical representations of knowledge.

It is well known that the visualization provides reduction of large amounts of data by presenting them in a condensed, visual way on maps, charts or graphs (Osinska, 2016). Information reduction is the most frequently mentioned advantage and for evidence the old phrase "*a picture is worth a thousand words*" is quoted in many publications (*The meaning and origin …*).

Another role noticed by experts is analytical in nature. The information maps should provide insight, i.e. deliver multifaceted analytical material and provide to discover a new, previously hidden knowledge about particular facts or events. In terms of visualization, the analysis is semantically relative to the management tasks. A good example constitutes the science policy makers who make decisions about the financial streaming and as a support use the statistical data visualization.

It has not been known since today that the visualization is helpful in learning as it provides to create associations between facts, parallel threads, ask yourself questions and easier remembering information presented in a visual form. Visualisation as augmenting textual content, affects both teacher and student and therefore plays intellectual as well as educational roles.

The visual language is omnipresent today in almost all domains of life and visualization is to be considered as one of the communication means, also in scientific fields. There is a tendency to use courageously *Infovis* instead of text, as its equivalent. For example, posters, illustrated abstracts, infographics — the instances of elements well adopted in scientific communication, evoked discussions between the specialists regardless of their specialization. We can see the communicative and social roles mutually interlace on the ground of information visualization. Sometimes visual information distorts real facts, whether or not in line with the authors' intension. Manipulation examples can be often met in media.

Current data or knowledge visualisations convey the aesthetical charge that arouse the observers' interest which would not have been otherwise triggered. Why ones are more attractive for the majority of people than other — this issue lies within a new subdiscipline — *neuroesthetics* (Osinska, Osinski and Kwiatkowska, 2016). It covers the scientific approach to the study of the aesthetic perceptions of art in the context of complex visualisations. Therefore *Infovis,* similar to art, influences a human imagination and emotions.

Visualisation process

Visualisation process consists of several phases which is illustrated on Figure 1. The first step is to collect the research data and it does not constitute a major effort. Next, data need to be cleaned, processed and prepared for the visual representation (steps 2 and 3). Data science competences are required from the researchers at that phase. Input dataset is represented by vector of attributes, which are difficult to visualize in 2D or 3D configurations. There are two ways to resolve this problem: to reduce dimensions of data or extend output exploration space. In the first approach, different algorithms originated from statistics, machine learning, graphs drawing or artificial neural networks algorithms (*ANN*) have been applied (Chen, 2004; Börner, 2010). The second method, connected with topological transformations, is used in early *Infovis* stage (2000 – 2008) by adopting such techniques as "fish eye" view, zooming glass, that is particularly often met in the semantic browsers (Osinska, 2010). The final phase of the mapping process focuses on improving the usability of visual layout through matching glyphs and colours, constructing legend and adding data manipulation functions like sorting, filtering, zooming and another facility.

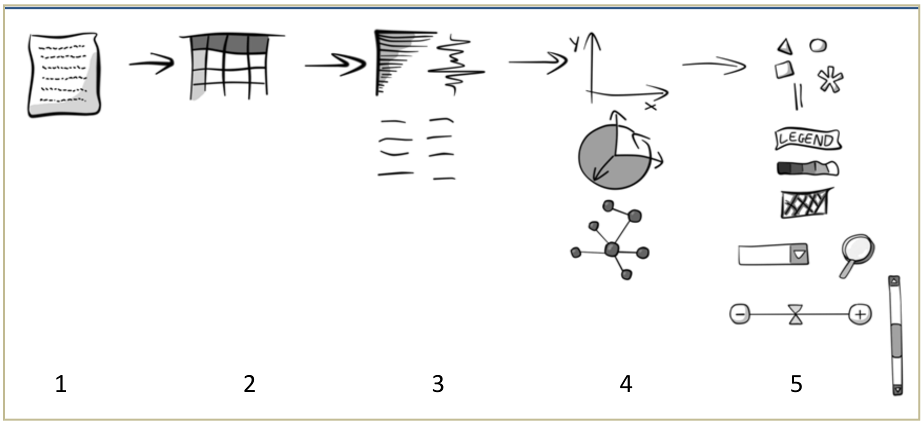

Figure 1. The five phases of visualization process: data gathering, processing, preparation, reduction and visual layout design.

Pre-visualisation: problem with data

The process of visualization, as can we see, is multifaceted, consisting in sequential steps, required to apply methods used in different domains of knowledge. In the subsequent phases, researchers meet difficulties of which nature cannot be unambiguously described. For instance, technological obstacles may appear at each step of research, and require involvement of professionals with specific skills. Therefore, it is impossible to select, group and order visualization problems according to the sequence of particular stages. A comprehensive approach implicates that in order to provide further development of *Infovis* we need to identify, characterize, rank and systematize all difficulties the researchers and practitioners encounter while experiencing visualizing and all labours resulting from those.

The tasks associated with the data and database are included in the first step (Fig. 1, 1). This is a time-consuming phase, making up sometimes even 80 percent of the whole scientists' effort. Working with data describes strongly directed competences and experience and therefore a new profession/specialization "data scientist" have appeared in response to the needs of contemporary science and business. First of all, researchers have to collect data reflecting research questions that are connected with sources that should be identified. The latter are generally global indexes, public statistics, digital libraries but also different websites or own observations. In the case of dedicated portals, the data are structured which allows to download and easily shape them into tables. If data are unstructured, i.e. cluttered by HTML tags, then the harvesting consists in the use of specialized scripts adapted to the Website information architecture. The whole process is called *Web scraping*. Scraped data must be cleaned, ordered and brought to some organized structure like tabular form related to a flat database. These tasks describe data processing stage of which success depends on several factors such as: complexity, uniformity of data (how many outliers are in a pattern), data scientists programming skills (scripts coding, formats interchange), computing power and volume size sometimes. All mentioned issues characterize not only the preparing phases of the visualizing process but first of all are typical for big data framework and data storage, particular noSQL mechanism (Aalton, 2016). Therefore, they can be considered as a distinct dimension of *Infovis* problems.

The problems and challenges in visualizing information

In the early stage of Infovis, Chaomei Chen, the leading specialist in this field, detailed ten unsolved visualisation problems and grouped them into technical, user-centred and disciplinary (Chen, 2005). His study involved information visualization as an unripe scientific discipline, which still needed many theoretical elaborations that would determine the presented in paper conclusions which are prevailing up to now. Besides the problems, the criticism of Infovis has been discussed in the subject literature from 2000-2010. That is a specific feature for that period —separating this research area from the scientific visualization, focused on data having representation in nature (i.e. physics, astronomy, biology, genetics and so on). Thus, data communication scientist in Tableau, Robert Kosara, underlining sublime of visualization postulates that the artistic approaches are more acceptable in *Infovis* than in scientific visualization (Kosara, 2007). The value of visualization based on the effectiveness as well as its limitations from the technological viewpoint are described by Jarke J. van Wijk (2005). Sabrina Bresciani and Martin J. Eppler - the co-author of a great *Periodic Table of Visualization Methods Table*[1] — project where visualization plays a double role (subject of study and the means of presentation), undertook to describe the disadvantages of *Infovis* (2008). They attempt to classify them into three categories: cognitive, emotional and social from the two perspectives: designer induced and user induced problems.

Nowadays, when differences between the scientific and “non-scientific” kinds of visualisation became no essential (the scientists often use sophisticated graphs and charts typical for the big data), while the scientific charts often include blocks of infographics and the frontiers between the usability and artistic scale of the visual presentation has become distinguished (Cairo, 2014), and when we observe the democratization of *Infovis* technology, it is worth to systematize our knowledge about its disadvantages.

The information visualisation problems and challenges assisted to the whole transformation process taken among others from own experience and observations of colleagues conducting akin research can be divided into the following categories: theoretical, topological, technical, cognitive, aesthetical.

1. Theoretical (scientific)

As mentioned above, *Infovis* is considered a scientific discipline by the inventors (McCard, 1999; Hall, 2008), but this argument can be easily disproved. The discipline is recognized on the basis of two criteria: epistemology-methodological and empirical (Cisek, 2002). The first one — the principle of providing theoretical framework in order to explain given phenomena and predict the trends, despite visualization draws from many research methods (maths, statistics, computer science, graphics), does not work (Osinska, 2016, pp. 46-49). The reason is

[1] http://www.visual-literacy.org/periodic_table/periodic_table.html

"visualization is intrinsically complex, has many aspects, and can be approached from many perspectives" (Wijk, 2004). On the other hand, many practical implications of *Infovis* pleads in favour of its scholar ground.

The lack of theoretical descriptions follows the problems with evaluation of visualization. Generated complex patterns on information or knowledge maps are difficult to put under benchmarking or any general assessment — there are still no explicitly defined quantifiable measures of quality (Chen, 2005). How to evaluate the visually presented knowledge became the acute problem today, whilst the researchers apply different methods and techniques to the same data. As Ch. Chen noted, the uncertainty — the cost of representing the scientific knowledge, simultaneously gives a better understanding of its status and evolution during research (Chen, 2018, p.32). The uncertainty implies the next *Infovis* problem: research re-productivity, i.e. providing clearly procedures to repeat the whole analysis process (Monya, 2016). This requires from authors to include full descriptions of attached maps in publications. This manner may help scholars to reproduce all stages of presented research, that is needed to verify results. As it seems, open data movement, when the scientists share own data, will more highlight the problem of visualisation evaluation. Sharing research data, to be subsequently visualized by different scholars, is featuring the era of social *Infovis* (see above).

2. Topological (scientific empirical)

Permanent problem of visualization which the scientists grapple with is the limited output space for the final presentation involving reference system and points, data glyphs, relationships, metaphors, legend and so on. In this context, *Infovis* researchers have a small choice: flat or 3D representation displayed on a computer screen. Input dataset is usually put into a vector of multidimensional features. Thus, the main difficulty of visual mapping consists in dimensions mismatch between the data representation and the predefined information space, which is supposed to serve exploration, insight and sometimes retrieval. As mentioned above, there are applied algorithms used in machine learning or ANN. There are also used topology deformation methods such as: exponential distortion, zooming glass or 3D graphs (Klavans and Boyack, 2014).

The problems of the representation space are covered by the topological modelling and appropriate algorithms of searching, which is addressed to the research interest of maths. This does not mean that these issues are limited to the theory. The empirical studies of large scale data visualization, particularly graphs exemplification, very clearly signalize the needs of novel effective methods of representation. The practitioners, without waiting for solutions, are experimenting with different ways of big data rendering. They sequentially zoom the relevant fragments of visual layout (for instance: Börner, 2015; Kaminska, 2017), publish interactive results (for instance; Chen's Homepage[2]), create animations providing scaling of the analysed data. Chen classified this issue as a scalability problem of information visualization which should be studied from the two perspectives: high-performance computing and individual user (Chen, 2004)

3. Technical

Technical side of visualization relates to the choice of appropriate algorithms resolving distinct tasks of visualization usability. It can be connected with practical implementations of

[2] http://www.pages.drexel.edu/~cc345/

topological modelling, dimension reduction, feedback within interactive layouts. Generally, information visualization has not required any advanced supercomputing techniques, they have been primarily reserved for the scientific visualization, such as genome analysis or topographical surfaces modelling (*Ibid*.). But the challenges of designing interactive, effective visualization applications covering the above listed technical issues needed a suitable technological support.

The basic *Infovis* principle “overview first, zoom and filter, then details-on-demand”, called *focus plus context* was originated by Ben Scheiderman (2001). If visual interface is interactive, it usually delivers such functions as zooming, filtering or manipulating by data. This means, focus plus context is not a challenge but prior standard in visualization applications and technical problems mainly revolve around ensuring interactivity and feedback with user, that is consistent with the HCI tradition. Visualization interface is often applied for information retrieval tasks which generates the current challenge in *Infovis* engineering (Osinska, Bala and Gawarkiewicz, 2012).

4. Cognitive

The way how *Infovis* designers code information can be distinct from the user’s comprehension. It can be caused by visual pattern overcomplexity, users unfamiliarity with the visual language, different cognitive schemes, semiologic ambiguity. Particularly, the use of specific in given domain metaphors can complicate the understanding of knowledge included in the maps. This all results in multi interpretations of information visualization — the basic cognitive problem in the scientific research. This difficulty evokes many scepticisms of *Infovis* implementation.

According to Chen, the high level of perceptual-cognitive tasks “include the recognition of a cluster of dots based on their proximity, the identification of a trend based on a time series of values, or the discovery of a previously unknown connection” (2004).

It is not easy to represent parallel relationships, for example clarify connections between the cited and citing authors by means of a single graph. Researchers make such mistakes as they point both kinds of ties to a single node and code them by arc direction that is difficult to insight (for instance: Kaminska, 2018). Intuitively, it would be proper to use several layers or several concentric circles, like a Sun Burst[3]. As noted by Wijk, the problems of visual insight may be induced by both designers and users (2008) — thus different channels of thinking, also within different sciences (Osiński, 2018) should be taken into account during the *Infovis* design process.

Elimination or minimizing of the cognitive problems will allow to extend usability of visualization and direct it to the young audience. This property can be used in learning — many researchers express the role of visualization in education (Ursyn, 2015; Osinski and Osinska, 2018). There are also useful online projects aimed to support high school students learning: Gap Minder, Visual Theasurus. Collecting of information and knowledge maps for educational purposes is important challenge today.

[3] https://www.cc.gatech.edu/gvu/ii/sunburst/

5. Aesthetical

Visualisation must be beautiful – the famous book of David McCandless (2009) reflects this sentence by title "Information is Beautiful" with attached examples of ingenious infographics. The same name dedicated to website[4] makes available high artistic visual layouts to be admired and insight. Any interdisciplinary research field as *Infovis*, needs a choice, readily balanced between the artistic and pragmatic values within visualisations (Kosara, 2007). Pragmatic content in the case of science maps means that the scientists can successfully analyse information structures and discover new knowledge about data. One hundred examples of such visualization currying deep informative dimension are at dedicated web service *Places & Spaces*[5]. Another, native science visualisation is illustrated below.

If a single graphical representation is not enough to cover all study results or express research paradigm, multi layouts are used on overall image (dashboard). Composition of the dashboard must follow the contemporary rules of graphical design, among others Gestalt's approach, however the historical knowledge visualizations rely basically on the natural geometry involving circles, trees and hybrids, known today as fractals. (Osinska and Osinski, 2018). These issues relate to both cognitive and art problems.

Polish science map

On Figure 2, the Polish science map, based on disciplinary classifications of all Polish scientific journals (N=4 500) is presented. Insight into complex patterns consisted of distinctive more or less clustered coloured dots allow to capture the selected aspects of Polish science:

- Contemporary medicine draws on cooperation with science and engineering which reflects in literature;
- Computer science issues are studied by the use of both exact sciences and engineering;
- Problems of natural sciences is closer to medicine on one hand, and engineering and technology on the other;
- Natural sciences are divided into two groups: biological-based close to medicine and social sciences and geological ones applying mostly engineering and technology.
- Single cluster assemblies forestry, agricultural and veterinary, which is not marginal but centred within engineering, natural and exact sciences,
- Art is characterized by the same isolation, but within humanity and social problems that means there is no up to now Polish writing about modern art using ICT,
- Social sciences are the majority of scientific journals,
- Despite the LIS associated with humanity, it arranges near to social issues; there is also lack of ties with computer and information sciences.

[4] https://informationisbeautiful.net/ . The similar blog about pretty pictures of visualisation is avalable at: http://www.visualcomplexity.com, written by Manuel Lima, who is looking for the connections between current visual layouts and historical sources

[5] scimaps.org

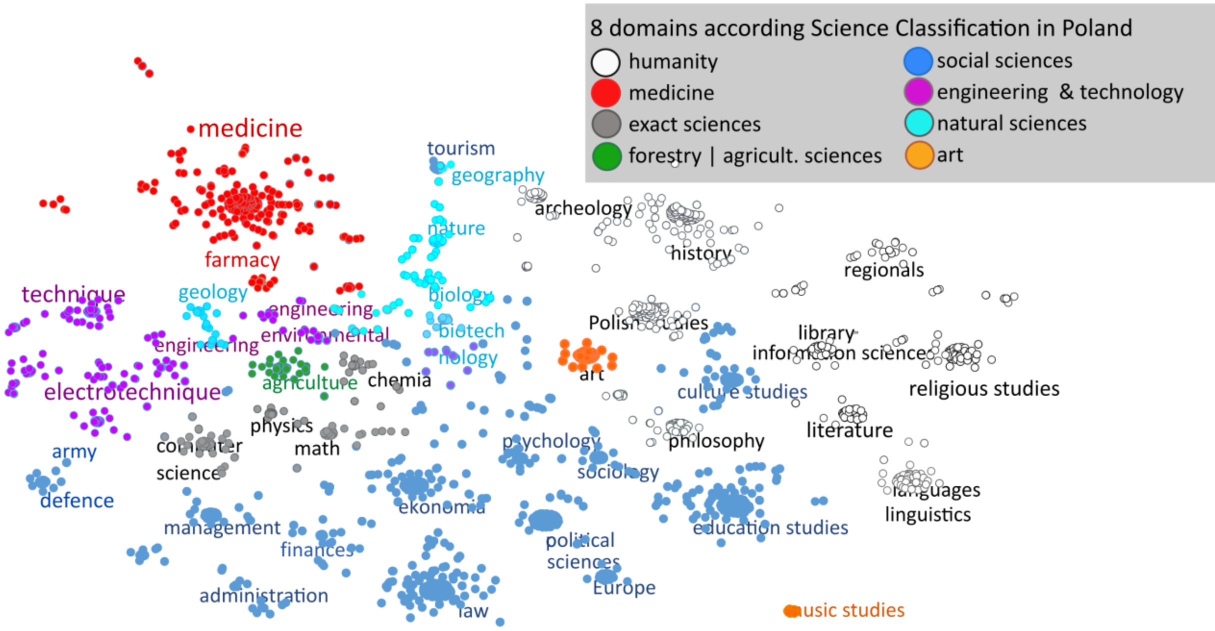

Figure 2. Polish science map based on 4.5 thousand Polish journals categorisation.

Summary

In recent years, it can be observed that *Infovis* technologies have become commonly available and therefore popular among all major scientists. Dissemination of visualization techniques and interdisciplinary framework causes many problems that in majority have not been solved since the field have been emerged. The author discusses the problems and challenges of *Infovis* starting from exposing the usability context of visualization. If we list all possible functions visualization included, then it is possible to focus on difficulties and barriers and classify them. The main problems scientists face today can be grouped into scientific theoretical, scientific experimental (topological), technical, cognitive and aesthetical. Before visualization process begins we have to collect and process research data, that is the crucial task in initial stage.

Among the current challenges affecting the scientific environment we can mention the need of theoretical elaborations, i.e. fundamental theory, strong establishment of *Infovis* solutions on education area as well as more explicit aesthetics input under the neuroesthetics study. There are no problems concerning the democratization of visualisation methods and therefore some experts write about the inevitability of visualization development (Alton, 2016), most of all in the scientific world.

Acknowledgements

This research is sponsored by Polish National Science Center (NCN) under grant 2013/11/B/HS2/03048/ Information Visualization methods in digital knowledge structure and dynamics study. I would like also to thank the author of unique professional database Arianta (http://www.arianta.pl/) – dr Aneta Drabek.

References

1. Alton, L. 2016. *4 Potential Problems with Data Visualization*. [online]. Scientific Blog. Available at: <https://www.datasciencecentral.com/profiles/blogs/4-potential-problems-with-data-visualization> [Accessed 8 March 2018].

2. Baker, M. 2016. Is the reproducibility is crisis. [online]. *Nature*, Vol. 533. Available at: <https://www.nature.com/polopoly_fs/1.19970!/menu/main/topColumns/topLeftColumn/pdf/533452a.pdf> [Accessed 8 March 2018].

3. Börner, K., 2010. *Atlas of science*. MA, Cambridge: MIT Press.
4. Börner, K., 2015. *Atlas of knowledge*. MA, Cambridge: MIT Press.

5. Bresciani, S. and Eppler, M. 2008. *The Risks of Visualization. A Classification of Disadvantages Associated with Graphic Representations of Information*. [online]. ICA Working Paper. University of Lugano, Switzerland. Available at: <http://www.knowledge-communication.org/pdf/bresciani-eppler-risks-visualization-wpaper-08.pdf> [Accessed 8 March 2018].

6. Card, S.K., Mackinlay, J.D. and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kauffmann.
7. Chen, Ch., 2001. *Information Visualisation*. London: Springer-Verlag.
8. Chen, Ch., 2005. Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications*, Vol. 25 (4). https://doi.org/10.1109/MCG.2005.91
9. Chen, Ch., 2017. *Mapping Scientific Frontiers - The Quest for Knowledge*. London: Springer-Verlag.

10. Cisek, S. 2002. *Filozoficzne aspekty informacji naukowej*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego. [in Polish].

11. Hall, P. 2008. Critical Visualisation. In: P. Antonelli, ed. 2008. *Design and the Ealstic Mind.*, New York: Museum of Modern, pp. 122−131.

12. Kaminska, A.M. 2017. Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN. *Torunskie Studia Bibliologiczne*, 2017 (1). [In Polish].

13. Kaminska, A.M. 2018. Visualizations of the GRUBA Bibliographic Database: From Printed Sources to the Maps of Science. In V. Osinska, and G. Osinski, ed., *Information Visualization Techniques in Social Sciences and Humanities*. IGI Global, Ch. 9.

14. Klavans, R. and Boyack, K. 2014. Mapping altruism. *Journal of Informetrics*, 8(2), pp. 431-447.

15. Kosara, R. 2007. *Visualization Criticism – The Missing Link Between Information Visualization and Art*. In IV (Information Visualization), 11th International Conference. 4-6 July 2007. Zurich, Switzerland

16. McCandless, D. 2014. Information is Beautiful. HarperCollins Publishers.

17. Osinska V. 2010. *Documents Retrieval and Visualisation*. Warszawa: SBP. [In Polish].

18. Osinska V. 2016. *Information Visualisation. Information Science Perspective*. Torun: UMK Publishing. [In Polish].

19. Osinska, V, Bala, P. and Gawarkiewicz, M. 2012. *Information Retrieval across Information Visualization*. In Fedcsis (Federated Conference on Computer Science and Information Systems, Wrocław, 20-26 September 2012.

20. Osinska, V, Osinski, G. and Kwiatkowska, A.B. 2015. Visualization in Learning. In A. Ursyn. ed. Handbook of Research on Maximizing Cognitive Learning through Knowledge Visualization, Ch. 13.

21. Osinska, V. and Osinski, G. ed., 2018. Information Visualization Techniques in Social Sciences and Humanities. IGI Global Publishing.

22. Osiński, Z. 2018. *Infographics in Humanities: Communication of Information or Information Noise? Polish Case*. In V. Osinska, and G. Osinski, ed., *Information Visualization Techniques in Social Sciences and Humanities*. IGI Global, Ch. 4.

23. Shneiderman, B., and Wattenberg, M. 2001. Ordered Treemap Layouts. In INFOVIS'01 (Information Visualisation). Proceedings of the IEEE Symposium on Information Visualization 2001. p. 73.

24. *The meaning and origin of the expression: A picture is worth a thousand words*. [online]. The Phrase Finder. Available at: <http://www.phrases.org.uk/meanings/a-picture-is-worth-a-thousand-words.html> [Accessed 8 March 2018].

25. Ursyn, A. ed. 2015. *Handbook of Research on Maximizing Cognitive Learning through Knowledge Visualization*. IGI Global Publishing.

26. van Wijk, J. J. 2005. *The Value of Visualization*. [online] In VIS IEEE. Available at: <https://www.cs.ubc.ca/~tmm/courses/cpsc533c-05-fall/.../vov.pdf> [Accessed 8 March 2018].