Kartografické listy / Cartographic letters, 2020, 28 (2), 53-68

IDENTIFICATION OF EVOLUTION, CHANGES AND TRENDS IN WORLD CARTOGRAPHY USING METHODS OF COMPUTER LINGUISTICS AND VISUAL ANALYTICS

Jakub KONÍČEK

Identification of evolution, changes and trends in world cartography using methods of computer linguistics and visual analytics

Abstract: Cartography is one of the world's oldest sciences. Over time, it has experienced many changes and developmental stages. Current world social, political and artistic trends and issues are affecting this development. Each of these influences has shaped cartography and cartographic creation into the form as we currently know. It is clear that these developments will be ongoing and that various significant and important changes will occur. However, the evolution of cartography must be thoughtfully managed so that it pursues current trends and remains timeless, otherwise the science may well be very easily reduced to an ancient discipline. The paper presents the developments, changes and trends in world cartography. The main aim was to identify the trends and critical milestones which resulted from these developments in cartography over time. The interests and activities of professional and scientific groups of the International Cartographic Association, the thematic focus of papers presented at International Cartographic Conferences and the critical ideas of cartographic articles in peer-reviewed journals are ideal materials for similar research. Methods of computational linguistics and visual analysis allows the study, identification and presentation of the key milestones which directly affect these materials. This type of knowledge can provide effective assistance in the management of future developments and directions in cartography.

Keywords: International Cartographic Association, International Cartographic Conferences, online academic services, data mining, text analysis, data visualization

Introduction

Society, culture and the world have evolved over time. Cartography and cartographic creations are no exception. For many generations, people have depicted various places on Earth and beyond in maps. The first primitive maps were created 25,000 years ago. The Pavlov map, as the oldest surviving map, was engraved into a mammoth tusk. Rock walls (Bedouin map), clay plates (Babylonian map) and textiles (Chinese maps on silk) are examples of materials which have been used to chart the surrounding world of the population, delineate territories and create games. From illustrating ideas of the world to recording the exact location of expeditions, the specific handwriting and motifs observable in the current artistic style have long been conspicuous on maps.

Now that we can move closer to any place in the world with satellite images in a few seconds or find the ideal route to a desired destination using a smartphone, the shift in cartography is enormous. It is extraordinary to compare this with map products even just a few years old. Nevertheless, the cartographic creation process is still guided by trends and evolves with and respects certain rules.

The article identifies the developments, changes and trends in modern world cartography. Specific analytical methods of text mining, quantitative content analysis and visual analytics can

Mgr. Jakub KONÍČEK, Department of Geoinformatics, Palacký University Olomouc, 17. listopadu 50, 779 00 Olomouc, e-mail: jakub.konicek@upol.cz

identify the main trends and factors which have affected, are still affecting, and are likely to affect the direction of modern cartography in the future.

1. Evaluation Process

The process of identifying major trends provides added value to trivial theoretical research. Most information sources are textual. If we wanted to study all these materials physically, it would be an enormous task requiring a significant investment of time.

An effective alternative method for exploring large text-oriented datasets is targeted analysis whose objective is direct identification of key concepts and thematic clusters, specifically, these are methods of text mining and Quantitative Content Analysis (QCA) and can be classed as methods of computational linguistics (Uszkoreit, 2000).

Computational linguistics is a field of applied linguistics dedicated to the study of natural language through computers and information technology. It is a cognitive science which applies the field of artificial intelligence and computational models which draw on human knowledge (Petkevič, 2017). According to the Association for Computational Linguistics, the field's main pursuit is in the provision of computational and analytical explanations of certain linguistic phenomena. Motivations in this scientific field can be exclusively technological in the form of explicit evidence for the existence of certain textual phenomena (Sproat, 2005). The main motivational constituents of computational linguistics fit well with identifying key concepts in the datasets of information sources. They seek to indicate strong and often recurring themes. This feature is very useful in identifying trends present in different cartographic sources, forms and designs.

2. Design and Research Method

Before commencing practical research, it is necessary to plan the entire process and define the individual carefully. An essential component is defining the research questions. A discussion of the answers to these questions should be the results of the entire investigation. Based on this procedure, suitable software solutions can be subsequently allocated. These solutions can be used for data analysis and the evaluation process. Equally important are detailed definitions of the input data and the allocation of specific information sources.

The research seeks to identify significant trends, milestones and areas of cartographic application which have fundamentally affected the formation of cartography. Through the practical implementation of partial steps and answers to the research questions specified in the following sections of the work, explicit identification of these features is possible.

To obtain valuable outputs while maintaining high work efficiency, the breakdown of the entire research process was divided into the six sub-steps described in the following chapters.

2.1 Research Organisation and Data Sources

Good research must have clearly defined rules and progress in work. Through predefined key points in the research process, we can proceed logically without unnecessary errors that may hinder the process or lead in the wrong direction. An equally important component is definition of the source data, their form and appearance, and a description of the processing methods and software solutions used.

The research in the paper explored the information from textual data obtained from a thematically oriented survey of the most used online academic services which concentrate on scientific articles found in publications such as *Web of Science*, *Scopus* and *Google Scholar*. The universal platform www.dimension.ai, which aggregates studies and articles from numerous publishing platforms, provides a comprehensive overview of cartographic research. The database has been expanded with reports from 2013 to 2020 published in the news archive of the International Cartographic Association (ICA). All information were exported to the *.txt* file format, which is the default format for conducting subsequent analysis.

Theoretical information were supplemented with reports, thematic research areas and the scientific specializations of ICA expert groups. Instructive sources such as research areas and presentations from the International Cartographic Conferences (ICC), available online since 2013 as pro-

ceedings, were also considered in the research. Insights for future work are provided by thematic areas planned for conferences in 2021 and 2023, which were already scheduled at the time of research.

2.2 Formulation of the Hypotheses

Formulation of the research questions is a research priority. An explicit definition of these questions allows orientation of the entire research towards precise answers and the achievement of valuable results. The following research questions were postulated:

– *What are the most significant areas of cartographic research today and what were they in the past?*
– *How have research trends changed?*
– *Which trend is currently the most pronounced?*
– *What changes and trends can be expected in the following years?*

The answers to individual hypotheses are contained in Chapter 3. Their descriptions are logically structured and substantiated with specific results from analyses.

2.3 Data Acquisition

The data acquisition process was one of the most time-consuming components of the research. The present study specifically involved searching through the above-mentioned online academic services, archives and thematically focused information from cartographic portals. Often, the information were also found on out-of-date and non-functional websites containing useful data. The form of data was different in individual cases depending on the source, and many times, proceeding was a challenge. A process of harmonizing the information into a uniform appearance was required for further analysis. We used the *PSPad Editor 5.0.1* software to edit information into a uniform text format. The software allows more advanced filtering and editing of text files and guarantees their independence from the original formatting style. Verified functionality and free licensing are some of the advantages of this software.

To obtain the theoretical knowledge which includes a specific concentration of themed articles from online academic services, a specific search environment enabling advanced filtering was applied. Specific records can be filtered by keywords which define the thematic areas and according to the required period. Based on the results obtained in this manner, a comparison of all outputs was obtained and answers to the hypotheses were formulated. The specific process is detailed in the Results chapter.

2.4 Identification

Computational linguistics can be applied through many processes, procedures and work systems. These can be speech recognition systems, text synthesizers, web browsers, text editors and specialized analytical software (Sproat, 2005). For the research requirements, analytical software was selected according to the knowledge which specifically focuses on detailed data.

Text mining can be understood from three different perspectives: text mining as information extraction, text mining as text data mining, and text mining as so-called Knowledge Discovery in Databases (Hotho et al., 2005). By combining these perspectives, text mining can be defined as the search for and discovery of previously unknown information by a computer using automated extraction of information from various text sources (pages, books, articles, etc.). Typical tasks involve the categorization of texts, creation of similarity clusters, and summarization or identification of relationships between entities (Hearst, 2003). The application of specific software allows statistical description and visualizations of the outputs to be formed. From this, a deeper understanding of the text can be acquired.

Quantitative Content Analysis (QCA) is denoted as a research tool, procedure, technique, method, approach or conceptual framework. Despite the wide range of approaches, the essence of QCA can be captured in its technical complexity and accessibility at the conceptual level. Simply, in the context of content analysis, texts or images are examined according to several predefined characters. Characters are words, phrases or typical images monitored for their occurrence and frequency. The process integrates a qualitative and quantitative approach (Dvořáková, 2010).

In combination with text mining techniques, such as relational analysis to identify relationships, a comprehensive and information-rich view of the researched data set can be obtained.

2.5 Technology

To streamline the work and support the present study, many thematic materials and an automated evaluation of the collected data were included. Free (open source) licensing, the solution's application and usage in similarly focused research, and a user friendly interface were emphasized to select suitable software. After taking into account these criteria, the choice was narrowed to *RStudio* and *KH Coder*. The availability of detailed documentation, high analytical functionality and a wide range of visualization outputs in these solutions allow the simple and effective addition of information value to theoretical knowledge.

RStudio is an open source and freely available development environment for effective work with the *R* programming language (RStudio Team, 2011). The software provides a means of analysis using commands written in the *R* programming language. The software is characterized as an environment for advanced statistical analysis of data and graphical display of these data. It supports modelling capabilities, classical statistical tests, time-series analysis, clustering and much more. A main feature is the ability to extend *R* with numerous functions and add-ons developed by a community of developers who provide regular updates and descriptions of functionality with clear tutorials (The Comprehensive R Archive Network, 2020).

R offers several useful functions for text mining. The functions indicate and highlight the most commonly used keywords in paragraphs of text. A popular method of effective data presentation is Word Cloud, which *R* can visualize. Unlike other online generators, *R* generates much more accurate visuals according to user preferences supplemented by statistical descriptions. Special libraries and packages have been developed to create these. A suitable method for the present study was clustering to indicate the existence of relationships between selected words.

For the research requirements, several packages and libraries intended for *RStudio v.1.2.5033* were used with version *R 3.6.3*. A distinct advantage of the software is direct synchronization and versioning on GitHub, which allows the project to be reproduced and ensures that the author, readers and potential reviewers can re-create and duplicate the research from the same workflow, computational process and software settings (Kmoch et al., 2019).

KH Coder is open source software for QCA or text mining which operates under a General Public Licence (GNU) and allows free use of the software and obtained results. Versions for Windows, macOS and Linux are available along with detailed manuals (Higuchi, 2020a).

The software can be applied in many scientific disciplines, such as neuroscience, sociology, psychology, education and informatics. Over 500 English professional articles in Google Scholar use the keyword "KH Coder", indicating its high relevance (Google Scholar, 2020). Even 3,687 English articles on the software's official website have actively applied it (Higuchi, 2020b). Several studies have rated its qualities as comparable to the analytical software *WordStat*, which is one of the best-rated text mining solutions; however, it is not available for free.

The software allows advanced search and the creation of specific statistics from a text file. The main advantage of this solution is text file filtering without unwanted components, such as punctuation, duplicate expressions, conjunctions and other words that would unnecessarily affect the research results. Files modified in this way can be exported directly and analysed manually or automatically. The option of direct application of the included analytical tools is also available.

The most significant functions used in the present study were word extraction and counting, word occurrence and connection, document clustering and classification, advanced search, and export as intuitively compiled and editable visualizations. The software supports 13 languages. For the research requirements, analysis in English was used.

The user environment of version *KHcoder3.Beta.01* allows work in either a graphical WYSIWYG interface or with a command line, which permits execution of SQL statements. More advanced features include installation of definable plugins, connection to a MySQL Database and custom data uploading. All outputs can be exported in various formats. Partial steps with settings can be saved in a local database.

2.6 Evaluation

The final research stage was the evaluation and interpretation of outputs and subsequent presentation as a comprehensive and descriptive text. As the final part of the research, this stage produced answers to the research hypotheses and a clear explanation in the context of the specific analytical outputs and references to the existing literature.

The research results were supplemented with data visualizations and infographics which represented the results in a simple and engaging form. These were outputs of visual analysis processes and assisted in providing a better understanding of the information. In addition to the above-mentioned software, individual visualizations using the online tool *Flourish* and desktop software *Adobe Photoshop CS6* were created.

Generally, people understand visual representations of information much more quickly than the same equivalent expressed in textual or numerical form. In addition to visualizations, the use of visual analytics also involves human factors: thinking, cognitive aspects and data mining. Visual analytics combines these factors into a single large unit and reveals the hidden connections in the examined data. The entire process was aptly described by Keim et al. (2004): "*Analyse first. Show importance, zoom in, filter and analyse again*."

The described procedure was applied practically to evaluate the obtained research results. A manual and specific analysis of all the theoretical starting points was first conducted. Only relevant data describing the cartographic trends of the given period or components dedicated to their prediction were selected from the data set. The information specified in this manner could be entered into automated, software-processed analyses, which also revealed its hidden connections. Based on these narrow concretizations of outputs, final interpretation of the results, which are clearly described in Chapter 3, were attained.

3. Results

Monitoring and identifying trends are a relatively lengthy and systematic activity. Computer linguistics and visual analytics methods allow us to reach valuable conclusions. The results of the present study were divided into four sub-units according to the main source of data. Their informative value corresponded to the hypotheses of cartographic trends over time. The answers to the hypotheses are included as a summary of the study's research topic.

3.1 eCARTO News

The International Cartographic Association (ICA) regularly publishes the *eCARTO News* newsletter on its website. It is a monthly magazine which describes the latest cartographic news, trends and innovations from around the world. News can be considered one of the most relevant sources for identifying current cartographic trends. The ICA is the highest decision-making body in the cartographic hierarchy, therefore its relevance as a source is indisputable.

From the archive of published reports on the website https://icaci.org/ecarto-news-archive/, which contains publications back to 2013, identify older news was also identified. Since the content of individual reports is relatively extensive, a content analysis was applied. By identifying keywords and terms in individual messages, the content of specific records at given times could be approximately determined. Fig. 1 charts the occurrence of the most represented keywords and uniformly visualizes them according to categories for specific years (note: available interactively https://public.flourish.studio/visualisation/3014286/).

The most numerous terms identified were "map", "mapping" and "cartography". Logically, given the topics under discussion in the field of cartography, these terms were the most represented. It is interesting to note the development of smaller representations of keywords, especially their variability and quantity. It can be deduced that the years 2015 and 2016 were the most comprehensive in terms of diversity of topics covered. These years also saw a noticeable increase in online and web-oriented concepts, which logically reflects the rise of online solutions and technologies, not only in cartography. Compared to 2013 and 2014, the difference is significant. In the last three years, we can observe a noticeable decrease in the use of the terms "mapping" and "map" (compared to 2013, by almost 100 %). We can also detect a slightly visible increase in the occurrence of terms such as "education", "history" and "cartography". In addition to changing pro-

fessional direction, the ICA and its affiliates are attempting to pursue an educational direction which aims to promote an awareness of cartography. This statement is supported by Virrantaus, Fairbain and Kraak (2009) in a study on the research agenda in the field of Cartography and Geoinformatics, where the authors explicitly issue ICA recommendations for increased educational activities, especially among the public.

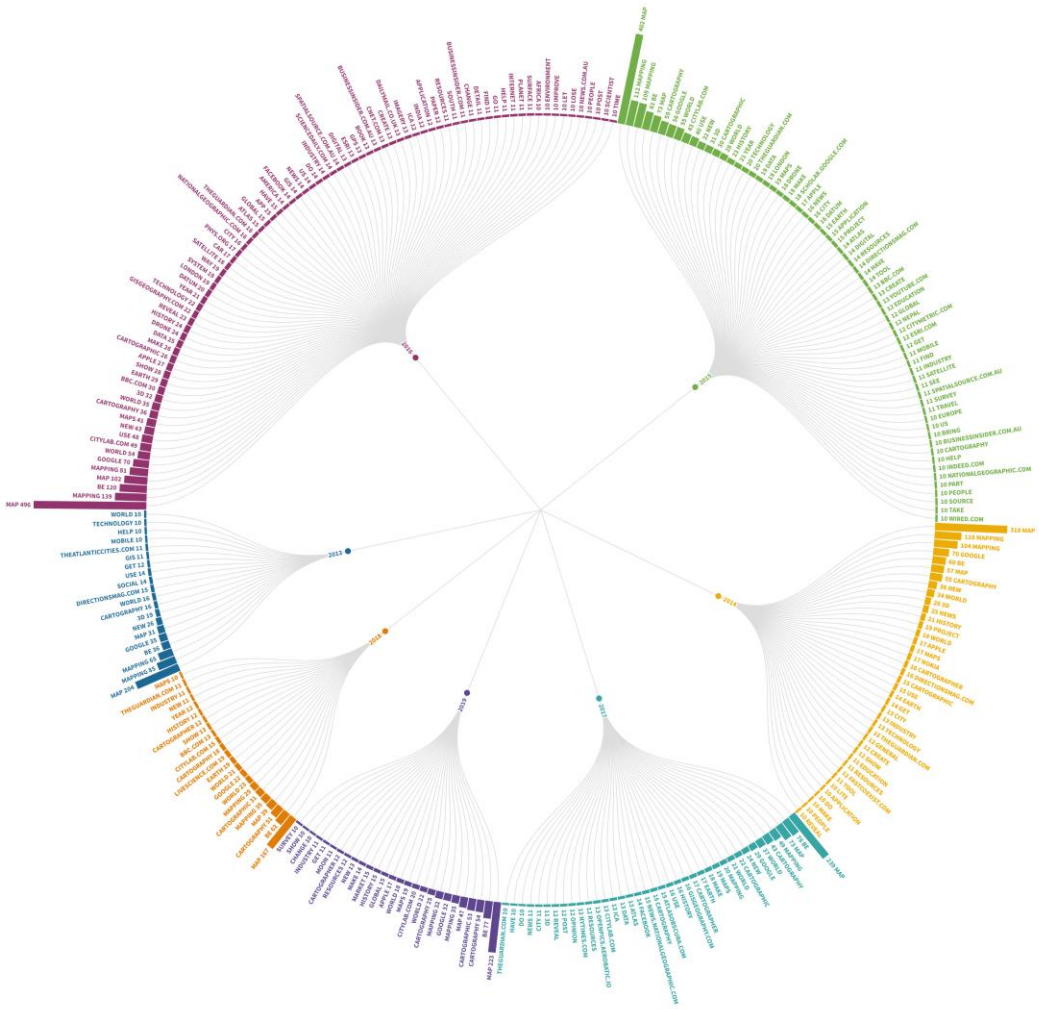

Fig. 1 Frequency of keywords in *eCarto News*

The Sankey diagram in Fig. 2 visualizes the representation of keywords for individual years of the analysed ICA *eCarto News*. Strong expressions dominant in individual years can be observed in the static image. A more detailed preview is provided by the interactive version of the diagram available at https://public.flourish.studio/visualisation/3014233/. The timeline of key terms can also be viewed in the animation https://bit.ly/news_years.

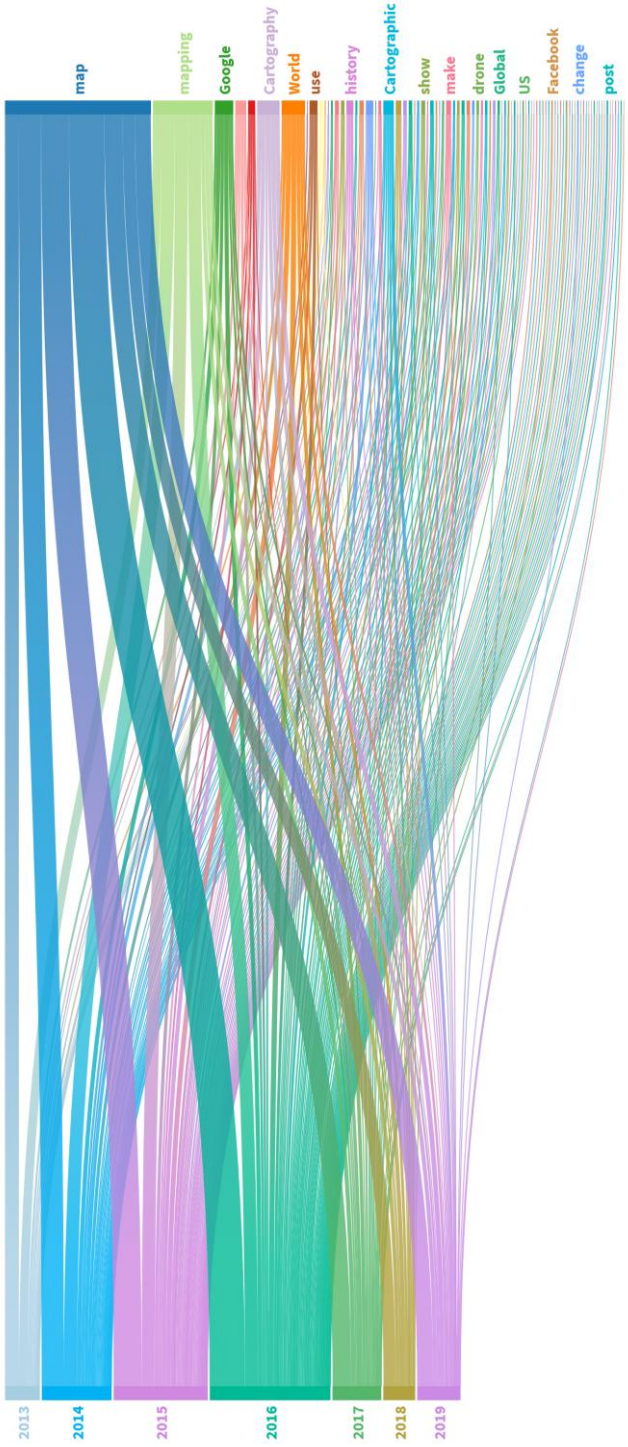

Fig. 2 Distribution of key terms in 2013–2020 in *eCarto News*

Significant clusters which again omit the terms “map”, “mapping” and “cartography” were “Google”, “New”, “World” and “Apple”. We expected these terms to occur since the news mainly publishes world news searched on Google and includes a reference. However, from a thematic point of view, the occurrence of the words “3D”, “technology”, “drone”, and the already mentioned term “history”, can be considered definitive. These terms reflect the important orientation of cartographic research in these areas. But it is also interesting to see such a low occurrence of the terms “GIS” and “geoinformatics”, which are inseparable to contemporary cartographic work.

3.2 Research papers

Peer-reviewed scientific articles are the main output of any valuable research. Cartographic research is no exception. Academic services such as Web of Science, Scopus, Google Scholar and dimensions.io provide access to the most important papers published in prestigious scientific journals. These include books, conference proceedings, reviews, bibliographies and other reviewed publications. Suitable sources are therefore available for conducting exploratory analyses to determine the scientific direction and value of cartographic research.

By using advanced search tools and filters for individual online academic services, we can define searches in the field of cartography much more effectively. Specifically, we can base an advanced search on the term “cartography” in the categories “topic”, “abstract”, “keywords”, “annotation” and “research area”. The time range for the observed period can be set to 2013–2020. Because the date for the year 2020 was only valid as of July 10, 2020, when the data were downloaded, the data could not be compared to other periods in a full range. However, the trend could be estimated approximately.

The number of cartographically oriented articles for each monitored year was precisely visualized in Fig. 3. The data were collected from the online tool *dimensions.ai*, which brings together the world’s largest number of scientific articles. Collection is performed immediately after the publication of any article. Based on the metadata, the data are categorized, and thus subsequent analysis is more objective. The division between the three most used academic services is depicted in Fig. 4.

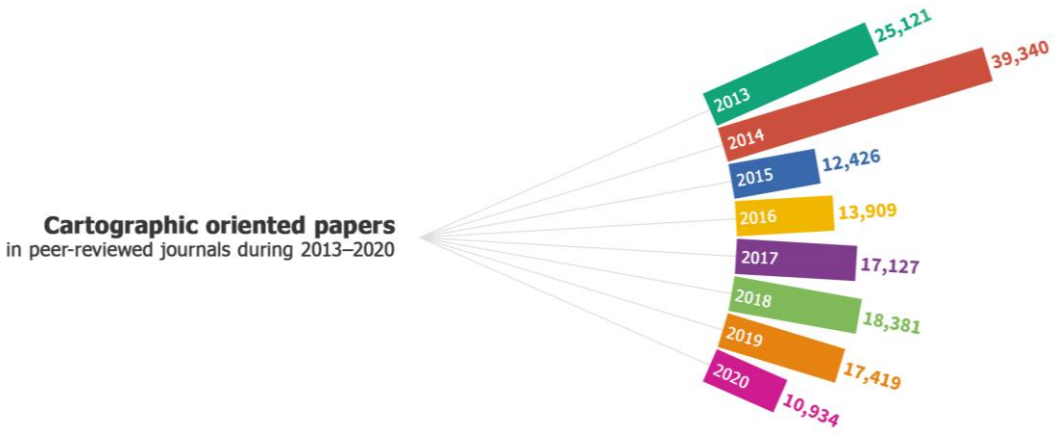

Fig. 3 Number of research papers which included the keyword “cartography”

A significantly greater number of articles was noticeable for 2013 and 2014, followed by a decline. An explanation for this may be that the direction of cartography, even at the direction of the ICA, began to expand into several scientific areas not devoted purely to cartographic research after 2015. These areas include *remote sensing*, *geology*, *urban studies* and *mobile and web applica-*

tions. The chord diagram visualized in Fig. 5 (interactively available at https://public.flourish.studio/visualisation/3163370/) depicts the cartographically focused studies in individual areas of research. We can also find supporting arguments in the sub-chapters devoted to the ICA conference thematic areas and its commissions, whose thematic focus become broader every year.

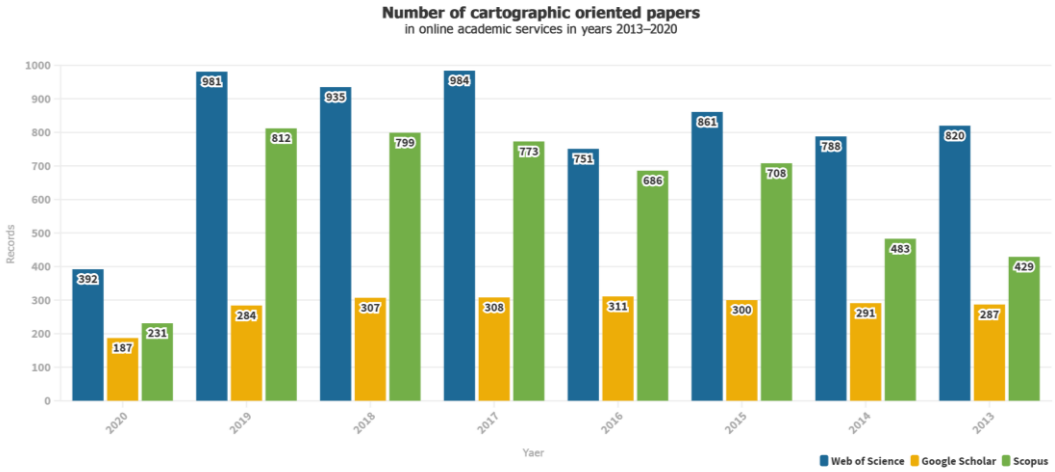

Fig. 4 Representation of cartographic articles mediated through academic services

In addition to geography and history, cartographic research has been significantly represented in the fields of environmental studies and educational fields for several years. In the last three years, remote sensing, physical geography and multidisciplinary geosciences have come to the fore. However, the surprising fact is a significant departure from the scientific field of computer science. This could be explained by the rather broad multidisciplinary application of cartography and the emergence of new scientific fields which have fallen into larger research areas in recent years.

3.3 International Cartographic Conferences

Conference papers may be considered one of the best tools for monitoring current and future trends in science. Contributors explore currently addressed topics and research, or present planned innovations. The ICA regularly organizes International Cartographic Conferences (ICC) every two years. From this, we can search for proceedings with individual contributions.

The analysis included the individual names of concrete thematic areas presented at conferences (2013–2019), and also from 2021, which are as yet only planned. Fig. 6 (interactively available at https://bit.ly/ICA_proceedings) charts important topics most represented at conference presentations according to individual years. The most stable topics discussed were web cartography, map projections, design, atlases, modelling and cartography and children.

Specific contributions indicate certain thematic clusters, i.e. significantly similar groups of topics. Their significance is visualized in Fig. 7. We can clearly identify the major topics as spatial data and analyses, design, history, open data and the often neglected cartography of mountains dedicated to tourist maps.

A more detailed analysis of the last two conference proceedings indicates an increase of up to 12 thematic areas. The areas which examine the link between cartography and current trends were included robotics, autonomous vehicles, artificial intelligence, open data and software, big data, games and crowd-sourced information. It is interesting to see a return to the field of theoretical

cartography. This fact is associated with educationally motivated ICA activities defined in the ICA strategic plan (International Cartographic Association, 2010).

A comparison of the thematic areas of the Tokyo Conference in 2019 and the planned cartographic conference in Florence in 2021 shows a surprising decline in some areas. A total of seven thematic areas will be eliminated, and only one new area will be added. The conference will not focus on toponyms, visual analysis, environmental mapping or topographic mapping. A new area dedicated to web cartography, map services and cloud computing will be added. This area precisely captures the primary direction of technology and research, not just in the field of cartography and geoinformatics.

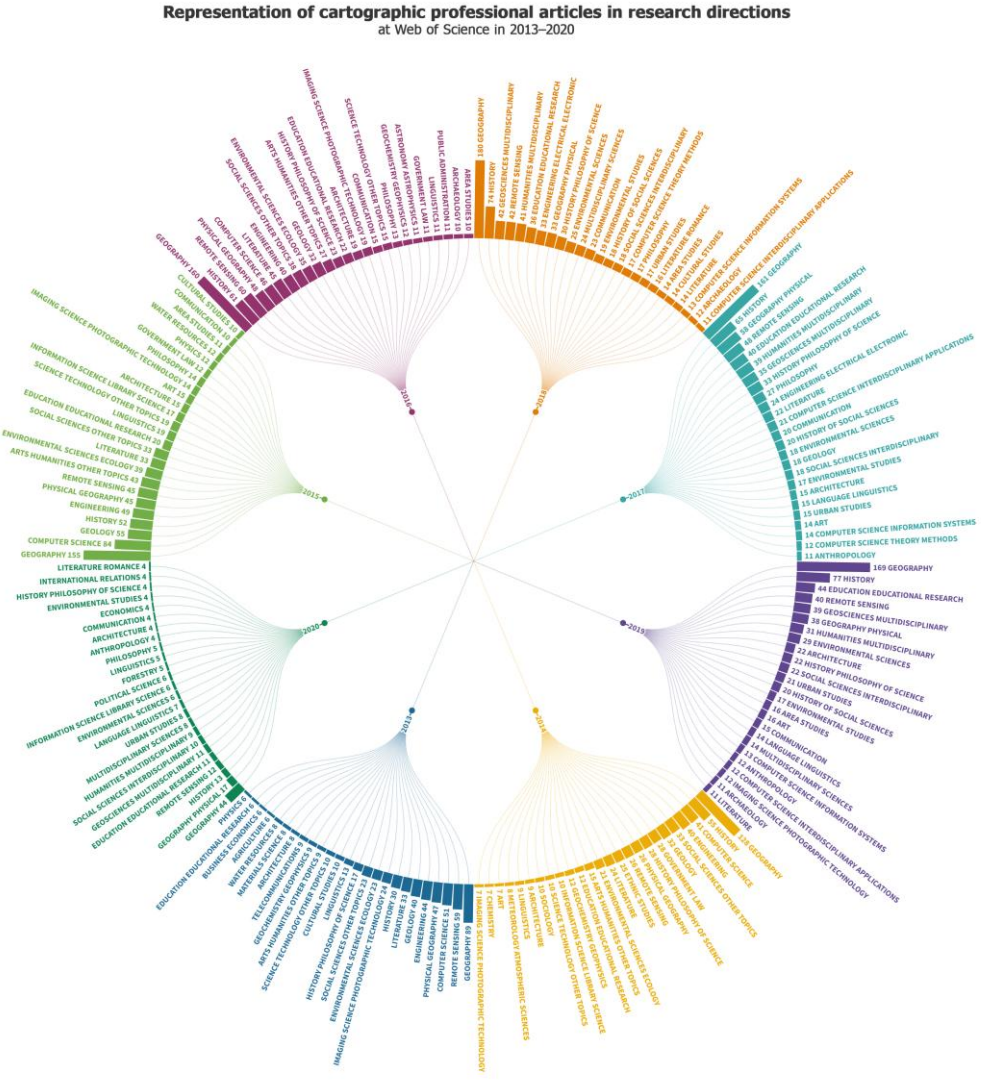

Fig. 5 Representation of cartographic research in scientific fields

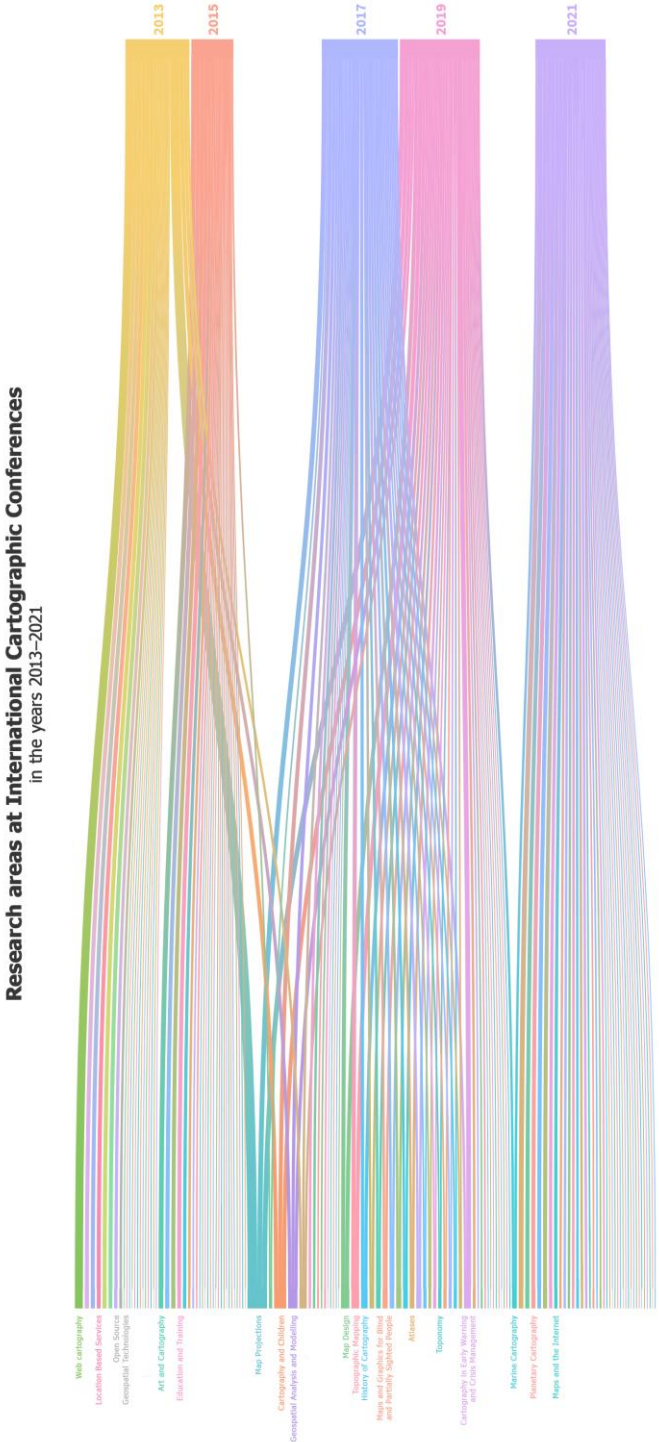

Fig. 6 Research areas of the ICC

Fig. 7 Number of topics in the generated clusters according to their thematic focus

3.4 ICA Commisions

The ICA commissions cover current issues, topics and trends in the field of cartographic research. Based on their composition, we can follow individual changes in the research orientation of world cartography. Commissions are set for four-year periods.

An overview of thematically focused commissions since 2007 is visualized in Fig. 8 (interactively available https://bit.ly/ICA_komise). In 2007, 22 commissions were appointed. Their continuation or removal in other functional periods can be monitored in the visualization according to individual lines. A total of 11 commissions are still active, while “Use and User Issues” has been renamed twice, although its principle remains unchanged.

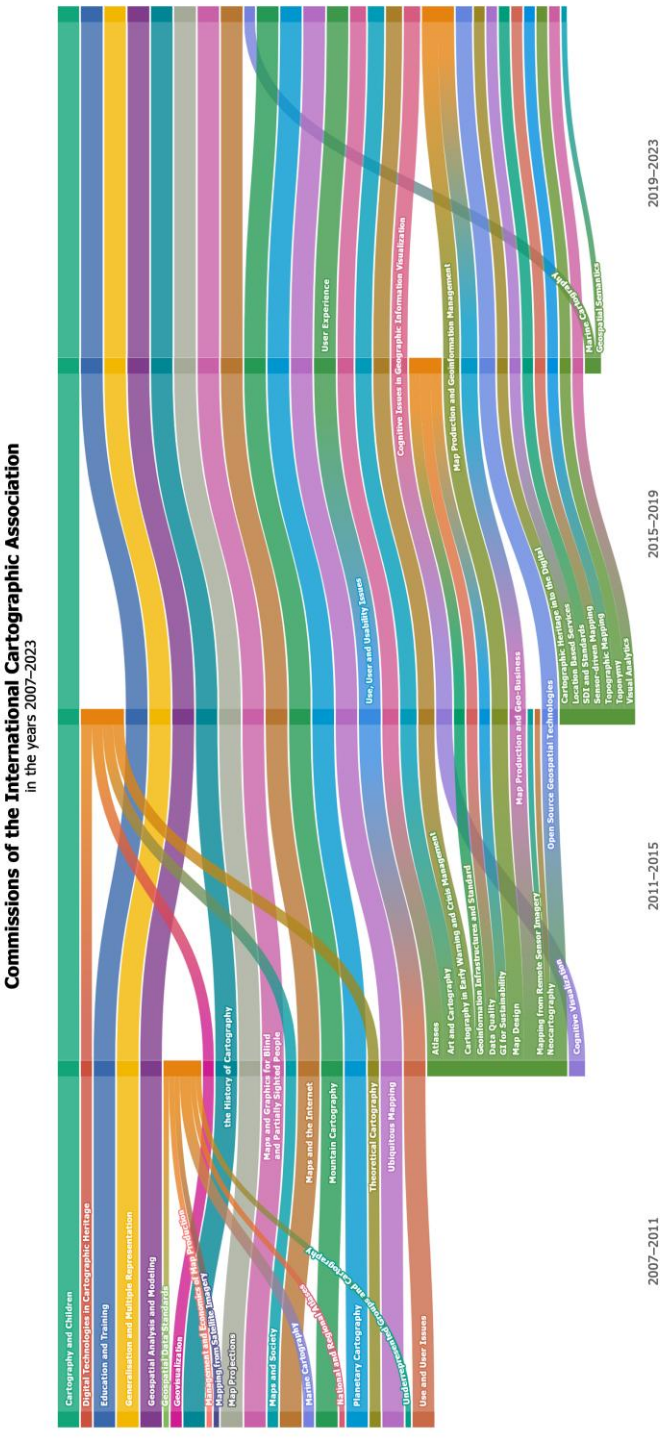

Fig. 8 ICA commissions in time

In 2011, 10 new commissions were established. Unfortunately, 5 of these expired within two terms of office. Thematically, they mainly focused on cartography, design and visualization. The year 2015 again saw an increase in seven new commissions operating up to the present. These were more technically oriented and digitally focused commissions which monitored the issues of visual analysis, sensors, SDI and LBS. By this time, commissions which monitored theoretical cartography, data quality, neo-cartography and management and the economic aspects of mapping ceased to exist. A total of 28 commissions were recently active. These strongly focused on current information technology trends, data issues, legislation and cognitive aspects of cartographic production. An interesting fact is that the commission on marine cartography was renewed, which was active until 2011.

Conclusion

Cartography has changed significantly over the last few years. Like many scientific fields, it is evolving as modern technologies develop. The times when maps were created by hand have long since vanished. The current direction is taking us towards innovations in information technology. Modern information systems, virtual reality, 3D, artificial intelligence, big data, and encouragement of open technologies are forcing cartographers to continually innovate and update their methods.

The connection with geoinformatics and information technology is continuously becoming stronger. Cartography is no longer the same science or even art as it was several years ago. Maps are now often created by non-cartographers who have no knowledge of basic cartographic rules, which significantly reduces the quality of resulting products.

The ICA and the individual cartographic societies of its members are aware of the current phenomenon. Through its activities, it seeks to reverse this scenario and follow the current trends as best as possible and adapt as flexibly as possible, not only at the scientific level but also the practical and commercial levels. This approach is certainly appropriate, but it is questionable how effective and sustainable it will be in the long term.

Cartographic foundations are still based on the same rules that were established before such a significant boom in information technology. We can observe this fact in this study through an analysis of ICA commissions and the thematic areas of conferences. More than half of all areas arise from foundations which are at least 5 to 10 years old. In the context of information technology development, which frequently changes, it is necessary to introduce more significant changes. As this study has shown, several paths are available. It is necessary to select the most current options and continually monitor developments, changes and trends.

Acknowledgment

This paper was created within the project "Advanced application of geospatial technologies for spatial analysis, modelling, and visualization of the phenomena of the real world" (IGA_PrF_2020_027) with the support of Internal Grant Agency of Palacky University Olomouc.

References

DVOŘÁKOVÁ, I. (2010). Obsahová analýza / formální obsahová analýza / kvantitativní obsahová analýza [online] [cit. 2020-05-10]. *Antropowebzin*, 2, 95-99. Available: http://www.antropoweb.cz/media/webzin/webzin_2_2010/Dvorakova__I-2-2010.pdf

GOOGLE SCHOLAR. (2020). *KH Coder Keyword*. [online] [cit. 2020-04-13]. Available: https://scholar.google.com/scholar?lr=lang_en&q=%22KH+Coder%22+%7C+khcoder&hl=en&as_sdt=1,5&as_vis=1

HEARST, M. (2003). *Marti Hearst: What Is Text Mining*? Berkeley (SIMS, UC Berkeley) [online] [cit. 2020-04-13]. Available: http://people.ischool.berkeley.edu/%7Ehearst/text-mining.html

HIGUCHI, K. (2020a). *KH Coder*. [online] [cit. 2020-04-13]. Available: https://khcoder.net/en/index.html

HIGUCHI, K. (2020b). *Research Using KH Coder*. [online] [cit. 2020-04-13]. Available: http://khcoder.net/en/bib.html?year=all&lang=English&key=

HOTHO, A., NÜRNBERGER, A., PAAß, G. (2005). A brief survey of text mining. *Ldv Forum*, 20(1), 19-62.

INTERNATIONAL CARTOGRAPHIC ASSOCIATION (2010). *Strategic Plan For The International Cartographic Association 2011–2019*. [online] [cit. 2020-05-20]. Available: https://icaci.org/files/documents/reference_docs/ICA_Strategic_Plan_2011-2019.pdf

KEIM, D., ANDRIENKO, G., FEKETE, J. D., GÖRG, C., KOHLHAMMER, J., MELANÇON, G. (2008). Visual Analytics: Definition, Process, and Challenges. In Kerren, A., Stasko, J. T., Fekete, J. D., North, C. (eds.) *Information Visualization*, LNCS 4950, 154-175, Berlin Heidelberg (Springer-Verlag).

KMOCH, A., NÜST, D., UUEMAA, E. (2020). *Proceedings of the 5th AGILE (Association of Geographic Information Laboratories for Europe) PhD School 2019*, November 25-28, 2019, Tartu (Zenodo). doi: 10.5281/zenodo.3835766

PETKEVIČ, V. (2017). Komputační lingvistika. In Karlík, P., Nekula, M., Pleskalová, J. (eds.) *CzechEncy - Nový encyklopedický slovník češtiny*. Praha (NLN).

RSTUDIO TEAM. (2011). RStudio, new open-source IDE for R [online]. *RStudio Blog. RStudio, Inc*. [cit. 2020-05-25]. Available: https://blog.rstudio.com/2011/02/28/rstudio-new-open-source-ide-for-r/

SPROAT, R. (2005). *The Association for Computational Linguistics: What is Computational Linguistics*? [online] [cit. 2020-05-10]. Available: http://www.aclweb.org/archive/misc/what.html

THE COMPREHENSIVE R ARCHIVE NETWORK. (2005). *The Comprehensive R Archive Network* [online] [cit. 2020-05-20]. Available: https://cran.rstudio.com/

USZKOREIT, H. (2000). *What is computational linguistics*? [online] [cit. 2020-05-20]. Available: http://www.coli.uni-saarland.de/~hansu/what_is_cl.html

VIRRANTAUS, K., FAIRBAIRN, D., KRAAK, M. J. (2009). ICA research agenda on cartography and GI science. *The Cartographic Journal*, 46(2), 63-75.

R e s u m é

Identifikácia vývoja, zmien a trendov vo svetovej kartografii s využitím metód počítačovej lingvistiky a vizuálnej analýzy

Hlavným cieľom tohto článku je predstaviť vývoj svetovej kartografie v čase, explicitne identifikovať kľúčové trendy a míľniky s ním spojené. Východiskovými dátovými zdrojmi sa stali konkrétne kartografické články uverejnené v recenzovaných časopisoch dostupných online cez akademické služby Web of Science, Scopus, Google Scholar a portál dimension.ai. Rovnako dôležitými zdrojmi, obsahujúce aktuality svetovej kartografie, sú dokumenty a správy zverejňované Medzinárodnou kartografickou asociáciou (ICA), ktorými sú eCartoNews, zborníky z konferencií, strategické výročné správy a ďalšie zdroje.

Vzhľadom k textovej povahe všetkých dostupných zdrojov a ich enormnému objemu sa na analýzu dát využili metódy počítačovej lingvistiky – text „miningu“ a kvantitatívnej obsahovej analýzy. S využitím softwarových nástrojov R Studio a KH Coder bolo možné dátové podklady dať do jednotného formátu a pomocou definovaných metód uniformne analyzovať. Na zvýšenie miery interpretovateľnosti výsledkov bola využitá vizuálna interpretácia s jednoduchou analýzou.

Výsledky práce prehľadne odpovedajú na vopred stanovené výskumné otázky prostredníctvom okomentovaných grafických výstupov vytvorených pomocou online nástroja Flourish. Informačne bohaté výstupy sú dostupné v statickej i interaktívnej online podobe vhodné k ďalšej interpretácii, diskusii alebo analýze.

Medzi hlavné výstupy patrí vizualizácia kľúčových slov obsiahnutých v eCarto News (obr. 1) zverejňovaných ICA počas obdobia 2013 – 2020, na ktoré sa tento výskum zameriava. Zreteľne je možno identifikovať najpočetnejšie zastúpené výrazy „map“, „mapping“, „cartography“. Zaujímavé je sledovať vývoj menších zastúpení kľúčových slov, predovšetkým ich variabilitu a množstvo. Jednoznačne možno dedukovať, že roky 2015 a 2016 boli najobsiahlejšie z hľadiska rôznorodosti tém. Tiež je znateľný nárast online a webovo zameraných pojmov, čo reflektuje rozmach online riešení a technológií nielen v kartografii. Sankey diagram na obr. 2 vizualizuje zastúpenie kľúčových slov za jednotlivé roky v analyzovaných ICA eCarto News. Za ďalšie výrazné zhluky, okrem pojmov „map“, „mapping“, „cartography“, možno považovať „Google“, „New“, „World“ a „Apple“. Ich výskyt bol očakávaný, nakoľko správy väčšinou uverejňujú svetové novinky dohľadateľné prostredníctvom vyhľadávača Google, kde vždy bola obsiahnutá referencia. Z tematickej oblasti sa však dajú považovať za smerodajné výskyty slov „3D“, „technology“, „drone“ a „history“.

Články v odborných recenzovaných časopisoch sú najrelevantnejším študijným materiálom každého výskumu. Na základe ich publikovania prostredníctvom renomovaných online akademických služieb, kde sa dajú jednoducho dohľadať, sú články ideálnym prostriedkom pre exploračnú analýzu („mining“) textu. Nástrojmi filtrácie sa vyhľadali len kartografické publikácie. Prehľad početnosti kartograficky orientovaných článkov za sledované roky prehľadne vizualizuje obr. 3. Údaje sú prevzaté z online nástroja dimensions.ai, ktorý združuje najväčšie množstvo svetových vedeckých článkov, keďže zber prebieha okamžite po zverejnení akéhokoľvek článku. Na základe metadát je každý článok kategorizovaný, a tým je následná analýza ob-

jektívnejšia. Rozdelenie medzi tri najvýznamnejšie akademické databázy poskytuje obr. 4. Signifikantne vyšší počet článkov je znateľný v rokoch 2013 a 2014, po ktorých nasleduje pokles. Vysvetlením môže byť fakt, že od roku 2015 sa smerovanie kartografie, aj vďaka odporúčaní ICA, začalo rozrastať do viacerých vedných oblastí, ktoré nie sú venované len kartografickému výskumu. Ide o oblasti ako diaľkový prieskum Zeme, geológiu, urbánne štúdie, či mobilné a webové aplikácie. Viditeľne zobrazuje túto skutočnosť i chordov diagram na obr. 5, kde je vyobrazené zastúpenie kartograficky zameraných štúdií v jednotlivých oblastiach výskumu.

Podobne ako odborné publikácie, aj príspevky obsiahnuté v rámci medzinárodných kartografických konferencií (ICC), sú významným ukazovateľom súčasných trendov v danej problematike. Do analýzy vstupovali názvy jednotlivých tematických okruhov na konaných konferenciách (2013 – 2019), a taktiež plánovanej konferencie v roku 2021. Obr. 6 priraďuje k jednotlivým rokom významné okruhy, ktoré sú na konferenciách zastúpené v najväčšej miere. Najstabilnejšími prejednávanými témami sú webová kartografia, mapové projekcie, dizajn, atlasy, modelovanie či okruh kartografia a deti. Pri pohľade na konkrétne príspevky je možné definovať určité tematické klastre, resp. významnosťou podobné zhluky. Ich významnosť je vizualizovaná na obr. 7. Zreteľne najdominantnejšími témami sú geopriestorové dáta a analýzy, dizajn, história, otvorené dáta či často zanedbávaná horská kartografia venujúca sa turistickým mapám.

Záverečnú tematickú orientáciu a vývoj svetovej kartografie z pohľadu ICA vizualizuje obr. 8, ktorý zobrazuje časový vývoj jednotlivých odborných komisií ICA. Na základe sledovania ich vzniku, resp. zániku je možné konštatovať, že aktuálne trendy sú výrazne orientované na informačné technológie, problematiku dát, legislatívu, a taktiež kognitívne aspekty kartografickej produkcie.

Na základe analýzy prípadových štúdií sa dá konštatovať, že kartografický základ je stále postavený na rovnakých pravidlách, ktoré boli stanovené v dobách pred významným rozmachom informačných technológií. Tento trend dokazujú hlavne analýzy zamerania komisií ICA a tematických okruhov ICC, podľa ktorých nadpolovičná väčšina z nich pochádza spred 5 až 10 rokov. Preto treba v kontexte vývoja informačných technológií, ktoré sa menia takmer každý mesiac, očakávať významnejšie zmeny.

Prijaté do redakcie: 15. októbra 2020
Zaradené do tlače: december 2020