

Enhancing Encyclopedic Characteristics Using Geotagging Why It Matters?

Jasmina Tolj

The Miroslav Krleža Institute of Lexicography, Zagreb, Croatia
jasmina.tolj@lzmk.hr

Ivan Smolčić

The Miroslav Krleža Institute of Lexicography, Zagreb, Croatia
ivan.smolcic@lzmk.hr

Zdenko Jecić

The Miroslav Krleža Institute of Lexicography, Zagreb, Croatia
zdenko.jecic@lzmk.hr

Summary

Through the last couple of decades, encyclopaedias have transformed significantly, becoming digital and born-digital, full of multimedia, and today among other characteristics being well connected via hypertext and metadata. At the same time, in digital humanities an increasing emphasis is being put on mapping or geotagging archival data, with special emphasis on networking projects and international cooperation. Digital platforms are being designed for publishing, describing, presenting, searching or browsing through historical sources. Some projects take the temporal aspect further, enabling not just retrieving historical data but also offering navigation through space and time via interactive maps. Encyclopaedias joining such initiatives would allow for new ways to explore its content and connecting encyclopaedic knowledge with specific artefacts and locations so users could visit them. Some encyclopaedic projects include a similar approach to their content, but further development is needed. In this paper, authors analyse existing conditions among web-based encyclopaedic projects (such as the Brockhaus Encyclopaedia and the Slovenian biographical lexicon), the data types appropriate for geotagging, opportunities and perspectives. This paper will also explain how using geotagging enhances encyclopaedias' characteristics. This would contribute to further encyclopaedia's interactivity and allow for new ways for users to explore content and learn, in turn contributing to its greater usage.

Key words: web-based encyclopaedia, geotagging, geospatial anchoring

Introduction

Starting from their very beginnings, encyclopaedias have introduced a new method of organizing and structuring knowledge and were thusly designed as a high-quality, simple, and quickly available source of information. As such, encyclopaedic works are characterized by accuracy, relevance, objectivity, comprehensiveness, credibility, timeliness, consolidation, complexity and structure (Jecić, 2013). With technological advancements, encyclopaedias first became digital, mostly meaning that previous works were digitized and supplemented by multimedia, and later have developed further to become comprehensive web-based projects. Wikipedia, initiated in 2001 as the global multilingual web-based edition based on mass collaboration, was among the first to explore a new way of thinking about encyclopaedic works. It was the first to offer open access to its knowledge, making it available for anyone with an internet connection. Encyclopaedic characteristics have since developed as well and now hold an even greater epistemological value regarding its update possibilities (timeliness), collaboration potential, being unlimited in scope, having far more possibilities for information retrieval and options for content interconnectivity (Smolčić et al., 2017).

Further advancement, especially in digital humanities and archival communities, allow for continued development. Increasing emphasis is being put on mapping or geotagging data, with special emphasis

on networking projects and international cooperation, such as the Topotheque¹ and the Time Machine² project. Such digital platforms are designed for publishing historical sources, such as photographs, documents or audio-visual records and then using interactive IT tools to describe, present, search or browse through them. As a data base of relevant information and a source of condensed knowledge, encyclopaedia must also keep up. Regarding encyclopaedias' future, Prelog (2010) writes about placing the developmental emphasis on more complex linking of articles, where expanding the original context in which some concept is introduced is enabling a dynamic approach. This would allow new creative ways of searching (or browsing) and using relevant information. Since much of the information contains a spatial dimension, it would be worth exploring what information can and should be linked to a geographical location, thereby enhancing search options, interactivity and comprehensiveness, thusly furthering the encyclopaedic concept.

The idea of enriching information on the internet with a spatial component was first introduced in the 1990s (Herring, 1994) which later developed into the term geoweb, relating to virtual maps on the internet and the ability to link abstract information (textual, multimedial or entire web pages) with geographic locations, allowing for search options based on geographic locations (Voloder, 2010). Encyclopaedic content, with its comprehensiveness and knowledge synthesis, surely offers plenty opportunity to be tagged to geographic locations (geotagged). The authors believe that encyclopaedias could play an even greater role in the so-called *knowledge ecosystem* and should therefore be further advanced by geotagging its content. This paper will present a brief overview of a few encyclopaedic and non-encyclopaedic projects that have in a way geotagged their content, and propose guidelines for further development in encyclopaedics. For encyclopaedias to reach their full potential in geotagging their content, however, it will be necessary to determine typology of data appropriate for geotagging and develop a more complete methodology for geotagging encyclopaedic content.

Scope and methodology of research

Three encyclopaedic and three non-encyclopaedic projects will be analysed in order to present the potential for use of spatial data and to what extent it is implemented in encyclopaedics. The non-encyclopaedic projects included are the Topotheque project (platform) which is an online archive operated in local entities, the eCultureMap³ which is an online geographical knowledge map connecting and visualizing digital cultural objects, and the Time Machine project, a platform to build a map of European history that spans thousands of years. These projects all relate to archival science, part of social sciences just as encyclopaedics, and were chosen to give example of what can be done using, among other, geospatial data.

The encyclopaedic projects presented include the Slovenian Biography portal⁴, which is an ongoing web-based project, the German Brockhaus encyclopaedia⁵, which is the most significant German encyclopaedia and one of the largest in general among encyclopaedic works, and the Encyclopedia Virginia⁶, which focuses on the history and culture of Virginia, USA. These three, biographical, general and national encyclopaedic projects, have in different ways put the geospatial component in use and therefore represent the practice of geotagging in encyclopaedic projects. Analysis of these six projects will allow insight into how to further improve web-based encyclopaedic projects using the spatial component of information and how this in turn enhances encyclopaedic characteristics.

The use of spatial features in digital humanities and archival science

Different archival and cultural communities have recognised a need to become more approachable to the general public and some to seek cooperation with gathering more materials to catalogue and save. One project that started as a private toolkit to manage (index, date and localize) a private footage has evolved into a collaborative online archiving platform for the public to save local and historically

¹ Topotheque <https://www.topothek.at/en/> (26.7. 2019)

² Time Machine <https://www.timemachine.eu/> (26.7. 2019)

³ eCultureMap <http://eculturemap.eculturelab.eu/eCulture14m/Map.html> (30.7. 2019)

⁴ Slovenska biografija <https://www.slovenska-biografija.si/> (30.7. 2019)

⁵ Brockhaus <https://brockhaus.de> (30.7. 2019)

⁶ Encyclopedia Virginia <https://www.encyclopediavirginia.org/> (30.7. 2019)

relevant material or knowledge with a European perspective – the Topotheque⁷. The platform allows users to create archives where image and file content is sorted, and therefore can be searched, by key words, date (timeline slider) and perspective⁸ on the map⁹. There is usually additional information, as a description or information on the owner of materials (often time museums and other institutions, not just private persons). The Topotheque was designed to not only act as means to preserve culture and history, but to also serve as a research tool. V. Lemić (2019) points out that all local collections are aggregated on the Europeana, European Union's digital platform for cultural heritage, facilitating swift exchange, gathering and presenting of information and allowing for new connections to be made in presenting cultural heritage, and to be used in other cultural, scientific or educational projects.

The eCultureMap is also related to Europeana¹⁰. It is an interactive online map connecting and visualizing digital cultural objects¹¹. It allows geographical overview of Europeana's content, the use of spatial tools for searching and browsing, and it consists of four components, the mapping (the interactive map, with automatic translation from English to native languages of metadata), route planning (for browsing objects along a selected path, with distances, images and links to detailed information), search (of Europeana content free text), and mobile component (using locational services with previous three components). The quality of added geographical metadata is crucial for efficacy. This map provides a valuable resource for creative industries, promotion of culture, education, cultural tourism and adds even more value to Europeana (Zakrajsek, Vodeb, 2014).

The most ambitious project among its peers is the Time Machine project that would map millennia of European historical and geographical evolution. It proposes to build a large-scale digitisation and computing infrastructure (through simulation, multi-scale modelling and artificial intelligence technologies) that would digitise great volumes of information from Europe's historical archives, museum, libraries and geo-historical datasets. The project is one of six that won The European Commission's pan-European competition for researchers addressing grand scientific and technological challenges of strategic importance for Europe and with potential to change the future (Aigner, 2019). One of many proposed project's outcomes is to identify larger patterns, correlations and connections that open new perspectives for in-depth analysis of the culture and history¹². It is evident that mapping information, providing a new dynamic approach to research, and connecting institutions that safeguard history and culture has been put high on Europe's priority list.

The use of spatial features in encyclopaedic projects

Slovenian biography

The Slovenian biography portal gathers biographies of notable figures from Slovenia's history and was created by the Slovenian Academy of Sciences and Arts, its Scientific Research Centre, the Jožef Stefan Institute and Seven Past Nine Ltd. It is comprised of biographies from three biographical lexicons, two of which were first published in print: Slovenian Biographical Lexicon that was published 1925–91 (around 5,000 entries entirely digitized and available on the portal, edited by Cankar et al.), Primorska Slovenian Biographical Lexicon that was published 1974–94 (containing 4,400 entries, out of which about 40% digitized and published on the portal so far, edited by Jevnikar et al.) and New Slovenian Biographical Lexicon that started as a web-based lexicon in 2013 (currently containing around 500 entries, edited by Svetina et al.). The portal has over 9,000 entries and is an important reference work for research in Slovenian humanities, social sciences and the history of the natural sciences.

Slovenian Biography uses open software and standards (Erjavec, Dokler, Vide Ogrin, 2017). To encode and structure information XML is used as a markup language, as it is compliant with the Text

⁷ It was initiated by Alexander Schatek, an entrepreneur and industrial designer, in cooperation with the International Centre for Archival Research (ICARUS) and Europ's co:op project – Community as Opportunity: the Creative Users' and Archives' Network.

⁸ Meaning that a photograph's position is not only shown as a point on a map, but the perspective in which the tagged image or video was taken can also be presented.

⁹ The portal uses an embedded Google Map.

¹⁰ In fact, it re-uses its content.

¹¹ It was developed in the EU project Carare (Connecting Archaeology and Architecture in Europeana).

¹² The Time Machine factsheet (updated June 2019). https://ec.europa.eu/newsroom/dae/docum ent.cfm?doc_id=60088 (6.8.2019)

Encoding Initiative guidelines for electronic text encoding and interchange (TEI, an extensive and flexible schema used to digitally represent texts)¹³. Among other (name, sex, occupation, etc.), data is encoded using the TEI module for places, which allows for detailed annotation. Information that is detailed includes settlement name, region type and name, country and geographical coordinates. Gazetteers are used to ensure the standard form of settlement names and additional information is encoded for settlements that no longer exist or have changed name.

This kind of detailed markup is then used to present information and allow for various browsing and searching options. Slovenian biography portal offers a search bar for simple and complex search queries, and many browsing options: alphabetical index, timeline (to select the period in which persons lived), by profession and occupation, status (nobility, academy member, etc.), born or died on this day, by family, and using the map¹⁴. The map presents the places of birth and death of all persons included in the lexicon, allowing users to browse by place of birth and death. When a user chooses a place on the map to explore, initial information given are the names of person (in *Surname*, *Name* form, that act as hyperlinks), divided into categories of *born here* and *died here*, along with years of birth and death. The hyperlinks link to the person's biography article in the lexicon. Other than that, the complex search option allows to search for persons by place of birth or death by name of settlement, region name or country. This allows for a new perspective on browsing through or researching Slovenia's remarkable persons.

Brockhaus Encyclopedia

In a paper discussing geospatial anchoring (geotagging) of encyclopaedia articles, Kienreich, Granitzer and Lux (2006) present how this was done for the German encyclopaedia Brockhaus. The work references the online version of the encyclopaedia at that time, which today does not offer the discussed features, but the work gives valuable insight into problems and possibilities of geotagging encyclopaedic content. This general encyclopaedia was published by Bibliographisches Institut and F. A. Brockhaus AG as the leading work in European German-speaking countries, with about 240,000 articles and the world atlas with 2,000,000 named geospatial references as well as detailed surface maps¹⁵.

Encyclopaedia's articles were geotagged¹⁶ to named geographical entities in the atlas, and Kienreich, Granitzer and Lux (2006) divided geospatial references chosen for an article in four categories: direct, indirect and symbolic geographical references, and temporal spatial references. A direct reference is formed when an article's topic is explicitly linked to a geographical entity, e.g. an article on a specific town is referenced to the same town on the atlas. An indirect geographical reference is formed when a geographical entity to be tagged (anchored) does not correlate to an article in the encyclopaedia's knowledge space. In a biographical article, for an example, a place of birth will be a geographical reference, but it might not have its own article (a non-German town, for an example). Such geographical references create relations between encyclopaedia articles that would otherwise not be connected because they don't overlap in the topical but do in the geographical domain. In cases where an article does not name a specific geographical entity, a symbolic reference is created, for an exemplary instance of the object class in question, for the largest of such objects or for the most important such object (as decided by editorial staff ranking). Temporal spatial references are created for articles on historical events, which require making references in a time-dependent manner (e.g. reference to a country, but only in a specific time period).

The map representing this geotagged content¹⁷ can be browsed and searched, as there is a search bar to search for geographical entities that then displays search results in a relevance-ranked list. Further options include focusing of a selected rectangle in the view, taking measurements, and overlapping satellite imagery, topological and geopolitical information. The authors point to challenges in encyclopaedic and spatial ambiguities, as well as name changes in geographical entities over time. They found that users benefit from geospatial anchoring by having one more option to navigate

¹³ TEI: Text Encoding Initiative <http://www.tei-c.org/> (30.7. 2019)

¹⁴ The portal uses an embedded Google Map.

¹⁵ Apart from detailed surface maps, the atlas also contains high-precision altitude information, not just on Earth, but on Moon and Mars as well.

¹⁶ The authors refer to it as geospatial anchoring.

¹⁷ Which utilises a web-based atlas client application (not Google Maps).

between encyclopaedia articles (through geospatial references), making the content better connected and the encyclopaedia more dynamic.

Encyclopedia Virginia

Encyclopedia Virginia is a project of Virginia Humanities and the Library of Virginia¹⁸ focusing on the history and culture of Virginia. It is an ongoing project, first published in 2008. Encyclopaedia's entries are topical and biographical, accompanied by primary documents and media objects, including images, audio and visual clips, links to Google Street View tours of historic sites and museum objects (in partnerships with museums and cultural institutions in Virginia)¹⁹. Each entry contains rich metadata, allowing users to search and browse content in various ways, via map, static or interactive time lines and with *This Day in Virginia* feature. Future plans are to expand options with, among other, an option for mobile users to read encyclopaedia's entries appropriate to the geographical spot on which they stand.

The map²⁰ allows searching by search term, time span and over sixty categories (such as Slavery, Geography, Architecture etc.), or it can be visually browsed. Each marker gives a brief description and a link to a correlating entry, and sometimes a date, event type (death, birth etc.) or place (e.g. highway marker, a park, a stream). The encyclopaedia offers over sixty virtual tours of museums, historic mansions, plantations, former slave dwellings and other. The tours offer 360° views of interior, exterior and sometimes the surrounding area. This provides a more vivid approach to researching Virginia's past, especially the reminder of unjust treatment of people of colour, to general public.

Discussion

There are some similarities in how the Brockhaus Encyclopedia and the Slovenian biography use the geotagged content. Both display the information on a map and allow for new ways to search and browse content, along with making new connections where there were previously none, and allowing for a more dynamic browsing session. The Slovenia's biographical portal holds one type of encyclopaedic article, the biographical article, so naturally there are fewer opportunities to advance the content by mapping than there are in a general encyclopaedia such as the Brockhaus. This is where it becomes clear that anchoring information to geographical locations can provide new paths (connections) for users to explore. Where, for an example, a search for a certain writer can lead to a town of his birth, then the town is connected to some historical event, which in turn leads to a new historical figure and so on.

What this type of encyclopaedia could further include as a feature is to point to not only its own articles, but to geographical locations, such as artefacts in museums or on display in public (statues and such), or perhaps provide images of such objects and locations. The Encyclopedia Virginia offers a similar feature, if not more advanced, providing virtual tours. It, however, lacks other features, such as the map that would visually represent locations of all virtual tours. These described encyclopaedic projects use geographical component of information to further their users' experience and also to enhance encyclopaedia's characteristics.

Geotagging advances encyclopaedia's connectivity, search options, interactivity and adaptability. Connectivity is made better throughout encyclopaedia's own content, but it can also be enhanced by connecting to outside content, such as objects in open space or museums. Spatial metadata used in geotagging allows search options by place, and when presented on a map it also allows better interactivity and adaptability as it facilitates new ways of finding (or browsing) content. While geotagging enhances encyclopaedic characteristics, it also presents new research options in other branches of social science, such as history, political science or sociology, just as it does in non-encyclopaedic projects.

Regarding the described encyclopaedic projects, it can be seen that there is no significant attempt in making many connections to other institutions, while described non-encyclopaedic projects try hard to do so. Perhaps encyclopaedias should also try to be included in Europeana or similar aggregating

¹⁸ The Library of Virginia became a partner of the project in 2012.

¹⁹ Encyclopedia Virginia. <https://www.encyclopediavirginia.org/about> (30.7.2019)

²⁰ The website uses an embedded Google Map.

projects, such as the Topotheque and the eCultureMap are, because they too are a sort of *storage* of history and knowledge.

While searching for information on the internet a user is flooded with enough to never read through it all, what is available is unstructured and often unverifiable. Unlike an internet search, encyclopaedias provide a compendium of trusted and concise information and knowledge, where articles are prepared by experts in a given domain and are guaranteed to describe a topic accurately and exhaustively. Due to their basic characteristics (accuracy, objectivity, relevance, etc.), encyclopaedias have the potential to play an important role in informing and educating the public, which is of great importance in this so-called *post-fact era*, where evidence is often overpowered by emotion and confidence in institutions, expertise and the media is on the decline. The European Union has recognized the importance of encyclopaedias and had conducted an analysis of European countries' online encyclopaedias (Bentzen, 2018). It found disproportion between the sheer amount and verifiability of information and how this impacts the European democracy system and values. The analysis emphasized the importance of information availability regarding at least basic history and culture in the so-called *knowledge ecosystem* and the importance of encyclopaedias as crucial factors in avoiding manipulation as they hold great potential for informing users seeking information and knowledge. For these reasons it is crucial to keep advancing encyclopaedic science to maximize encyclopaedias' usability.

Geotagging in encyclopaedic projects is underutilized and many encyclopaedias do not use spatial tags at all²¹, giving a lot of room for further research on this topic. For these reasons it is crucial to advance encyclopaedic science further and to create a more complete methodology for geotagging encyclopaedic content. To do that, it is necessary to investigate solutions to many challenges that arise, such as to determine which information is relevant for creating geospatial tags, developing automatic or semi-automatic tools to geotag existing projects, and most importantly determine typology of data appropriate for geotagging.

References

- Aigner, T. (2019). Europe builds a Time Machine. Big Data of the Past is becoming a reality. Vienna: Insights, ICARUS – International Centre for Archival Research
- Bentzen, N. (2018). Europe's online encyclopaedias. Equal access to knowledge of general interest? Bruxelles: European Parliamentary Research Service
- Erjavec, T., Dokler, J., Vide Ogrin, P. (2017). Slovenian Biography. // Proceedings of the Second Conference on Biographical Data in a Digital World, (Vol-2119) / Fokkens, A., ter Braake, S., Sluijter, R., Arthur, P., Wandl-Vogt, E. (eds). Linz: CEUR-Workshop Proceedings, 16-21
- Herring, C. (1994). An Architecture for Cyberspace: Spatialization of the Internet. Champaign: The United States Department of Defense, U.S. Army Construction Engineering Research Laboratory
- Jecić, Z. (2013). Enciklopedički koncept u mrežnom okruženju. // Studia lexicographica 7, 2(13), 99-115
- Kienreich, W., Granitzer, M., Lux, M. (2006). Geospatial anchoring of encyclopedia articles. // Proceedings of the International Conference on Information Visualisation. / Bilof, R. S. (ed). London: Institute of Electrical and Electronics Engineers, 211-215
- Lemić, V. (2019). Topoteka – naša povijest, naš arhiv. <http://www.skole.hr/upload/portalezaskole/newsattach/16617/Topoteka.pdf> (6.8.2019)
- Prelog, N. (2010). Od tko i što do kako i zašto – budućnost online enciklopedija. // Studia lexicographica 4, 2 (7), 164-176
- Smolčić, I.; Tolj, J., Jecić, Z. (2017). Epistemological Value of Contemporary Encyclopedic Projects. // INFUTURE 2017: Integrating ICT in Society / Atanassova, I., Zaghoulani, W., Kragić, B., Aas, K., Stančić, H., Seljan, S. (eds.). Zagreb: Faculty of Humanities and Social Sciences, University of Zagreb, 141-149
- The Time Machine factsheet. (2019). https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60088 (6.8.2019)
- Voloder, I. (2010). Strategija razvoja geowebe s informacijskog, tehnološkog, kulturološkog i poslovnog stanovišta. Doctoral dissertation. Zagreb: Filozofski fakultet
- Zakrajsek, J. F., Vodeb, V. (2014). eCultureMap – Link to Europeana Knowledge. // Communications in Computer and Information Science (Vol. 416 CCIS) / Bolikowski, L., Casarosa, V., Manghi, P., Goodale, P., Houssos, N., Schirrwagen, J. (eds.). Springer International Publishing Switzerland, 184-189

²¹ Encyclopaedia Britannica, for instance, does not geotag its content.