

Epistemological Value of Contemporary Encyclopedic Projects

Ivan Smolčić

The Miroslav Krleža Institute of Lexicography
Frankopanska 26, Zagreb, Croatia
ivan.smolic@lzmk.hr

Jasmina Tolj

The Miroslav Krleža Institute of Lexicography
Frankopanska 26, Zagreb, Croatia
jasmina.tolj@lzmk.hr

Zdenko Jecić

The Miroslav Krleža Institute of Lexicography
Frankopanska 26, Zagreb, Croatia
zdenko.jecic@lzmk.hr

Summary

Contemporary encyclopedic projects are a continuation of a multi-century evolution of encyclopedic work. The first generations of digital (electronic) encyclopedias brought a significant shift in the practicality of their use and enriched them with multimedia and hypertext. Contemporary encyclopedic projects are a step further and contribute new epistemological values as opposed to traditional and digital editions. This paper presents this added epistemological value through the example of a few modern projects, and provides a contribution to detection and systematization of the newly created values.

Key words: encyclopedia's epistemological value, web-based encyclopedia, contemporary encyclopedia.

1. Introduction

As comprehensive projects, encyclopedias hold a great epistemological potential as a high-quality, simple, and quickly available source of information. Since the era of Enlightenment, traditional encyclopedias have put forward a new approach to organization and structuring of knowledge, and thus represent the forerunner of modern information systems. The paradigm of encyclopedic work is transforming as encyclopedia enters a new era of development, of constantly being developed by editors, associates, and users, and as it is now based on ad-

vanced IT and communication technology solutions. For these reasons, a new epistemological assessment is needed to evaluate these new characteristics.

2. Scope and methodology of research

To present the encyclopedia's evolution brought on by technological development and new media, epistemological characteristics of encyclopedic works were analyzed, beginning with traditional printed ones, through early digital works on CD-ROMs and DVD-ROMs, up to contemporary dynamic web-based encyclopedic projects. Some aspects of few well-known general encyclopedias will be concisely analyzed, that were foremost created as printed works to later have their content presented on the internet, and thus made accessible to a larger number of users. Firstly, *Encyclopedia Britannica*¹ will be analyzed, the oldest encyclopedia in English that is still in production. It was first published in 1768 in Edinburgh, Scotland, its last 15th edition was released in 2010, and since then it continues development as a web-based edition. *Brockhaus Enzyklopädie*, the largest printed encyclopedia of present in German was first published 1796–1808. The last printed copies of the 21st edition were sold in 2014, and the content of its 300,000 articles stored digitally. The Croatian Encyclopedia (*Hrvatska enciklopedija*²) is a general encyclopedia available at the web site of The Miroslav Krleža Institute of Lexicography, and it is based on the printed edition published in 11 volumes during 1999–2009. The influence of traditional encyclopedias will be compared with contemporary web-based encyclopedias that function as knowledge portals.

The *Stanford Encyclopedia of Philosophy*³ project represents the link between web-based encyclopedic editions and peer reviewed original scientific papers in all areas of philosophy. This encyclopedia is exclusively web-based and was set up in 1995. An overview of *Wikipedia*⁴, the global multilingual^{web-based} edition, based on mass collaboration, initiated in 2001, will be analyzed. Its content is openly editable, meaning anyone is permitted to create and edit its content. The possibility of advanced information retrieval and interconnectivity is demonstrated through the example of Wikipedia's DBpedia⁵, ontology database generated from Wikipedia.

This paper analyzes epistemic characteristics of contemporary encyclopedias that include continuity (staying up-to-date), collaboration, boundlessness of scope, information retrieval, and interconnectivity. The continuity, boundless-

¹ <https://www.britannica.com/> (accessed Apr 5th 2017)

² <http://www.enciklopedija.hr/Default.aspx> (accessed Apr 5th 2017)

³ <https://plato.stanford.edu/> (accessed Mar 8th 2017)

⁴ https://en.wikipedia.org/wiki/Main_Page (accessed Apr 17th 2017)

⁵ <http://wiki.dbpedia.org/> (accessed May 15th 2017)

ness of scope, and collaborative character of the web-environment will be presented through a quantitative analysis of numerical indicators, demonstrating the epistemological upgrade when compared to traditional encyclopedias. For sake of brevity and keeping in mind the usually strict form of each of the encyclopedia-types (i.e. traditional, digital and web-based encyclopedia), not all aspects of the fore-mentioned encyclopedias were analyzed to be compared to the next.

3. Analysis of encyclopedias' particular epistemological features

3.1. Continuity and collaboration

Epistemic importance of encyclopedias staying up-to-date is reflected in how quickly the information in them out-dates, which reduces the epistemological effect of these works. Data can be outdated even at time of publishing and the time-line between updated editions is regularly several years⁶. The need to continuously keep track of reach and development of human activity is indicated in regular publishing of updated encyclopedic editions. Encyclopedia Britannica, existing for over 240 years, had since 1768 reached 15 editions before it was published online. The Brockhaus Enzyklopädie project began in 1796, and reached 21 editions. The editions differed in scope and number of volumes, thus limiting the number of articles and the content of each. Development of means of digital storage, along with appearance (1970s and 1980s), and later increase in number of personal computers (1990s) permitted encyclopedia's scope growth. In 1986 the Grolier's *Academic American Encyclopedia* was published on a CD-ROM, titled *Electronic Encyclopedia*⁷, and became the first digital encyclopedia to be distributed on a digital portable media. The printed version was made up of 21 volumes with 10,000 pages of text, while the CD-ROM was a text-only data base, with over 9 million words in 30,000 articles. The use was enabled by a software for search and retrieval of information, allowing search by article title, or entire text. After digital editions of mostly smaller encyclopedias and encyclopedias for young adults, larger encyclopedias gradually decided to try out in digital publishing. In 1994 Encyclopedia Britannica joined in and started publishing digitally, first as *Britannica online*⁸, and in 1995 the same version was published on a CD-ROM⁹. The first digital version

⁶ For example, the initial volume of Encyclopedia of Technology, published by The Miroslav Krleža Institute of Lexicography, was printed in 1963, and hasn't been updated since, meaning that for decades there was no encyclopedic update of achievements and reach in technology.

⁷ The Electronic encyclopedia, KnowledgeSet Corporation and Grolier Electronic Publishing Inc., 1986.

⁸ Then at: <http://www.eb.com/eb.htm>

⁹ Britannica CD, BCD, Version 1.0., Chicago: Encyclopedia Britannica 1994

of the German Brockhaus encyclopedia, *Der Brockhaus Multimedial*¹⁰, was published in 1998 in an abbreviated form¹¹. In 2005 the complete digital version of the entire 21st edition of Brockhaus encyclopedia¹² was published on a USB memory card – *Brockhaus Enzyklopädie Digital*¹³. It was to be internet-accessible free of charge since 2008, but that was never realized. The first comparisons between printed and digital encyclopedias highlighted their physical differences, organization of data, (un)limited scope, and (un)changeable content. Digital editions were updated more frequently, usually annually, and along with the advantages of more information retrieval options and added multimedia, a significant reduction in cost of publishing was also achieved.

Contemporary web-based encyclopedias are continuous projects, issued daily by updating existing and adding new entries, making them a more relevant source of knowledge. Further development is aimed at information retrieval capabilities, web design, setting up metadata, and so on. The largest project of such type is Wikipedia. It is based on mass collaboration, allowing anyone to create or modify content, with chronological storage of previous versions. Free access entails a certain risks, such as reduced accuracy in relation to professional works as well as reduced relevance, objectivity, and other components of the encyclopedic concept¹⁴, which are the basis of the encyclopedic work. Nevertheless, such a global, and relatively reliable source of information offers great opportunities for finding information. The Stanford Encyclopedia of Philosophy is a continuous project facilitating constant collaboration between authors and editors working in universities and institutes around the world. The editorial board, made up of experts in each field of philosophy, selects authors, designates articles, and also reviews their work.

The project strives to be up to date, so articles are intended to be revised and updated by the author every 3 to 5 years. Since the content is constantly changing, every update is archived to avoid problems of citing the articles. This allows the project not to out-date as it represents a system for assimilation, processing, and dissemination of new information, and regular implementation in the content of existing articles. The reach of current knowledge in philosophy is thusly monitored, kept up-to-date and tailored to users' needs.

¹⁰ Der Brockhaus Multimedial, Mannheim: F.A. Brockhaus, 1998.

¹¹ Electronic digital edition Der Brockhaus in fünfzehn Bänden, Mannheim: F.A. Brockhaus, 1997

¹² Brockhaus Enzyklopädie, 21st Vol., Mannheim: F.A. Brockhaus, 2005.

¹³ Brockhaus Enzyklopädie Digital, Mannheim: F.A. Brockhaus, 2005.

¹⁴The encyclopedic concept, along with the stated properties of accuracy, relevance and objectivity, also includes: comprehensiveness, credibility, uniformity, complexity, organization, and keeping up to date. More in: Jecić, Z.: Enciklopedijski koncept u mrežnom okruženju. *Studia Lexicographica*, 7(2014) 2(13), str. 99–115.

This organization and cooperation of scientific circles around the world in the field of philosophy can be named a scientific encyclopedia¹⁵, because it directly provides results of scientific research or content at the level of review papers. Significant epistemological upgrade of such projects, provided on-line as public service free of charge, stems from their availability to anyone with an internet connection.

The Croatian Encyclopedia¹⁶ is a comprehensive project by The Miroslav Krleža Institute of Lexicography. Though it was primarily produced as a hard copy edition, it is now web-based and shares the advantages of such encyclopedias. It is constantly being updated by editorial staff to prevent falling out of date, and has a growing corpus – expansion of existing and addition of new entries. It is collaborative and open to all users through inquiries and comments to the editorial staff which often leads to new insights and the raise in quality of content. Table 1 shows the number of editorial interventions, article updates and added new entries, which highlights the continuity of work and longevity of the project. The number of user comments in 2015 was 822, declined in 2016 (435 comments), then rose again in 2017, when the number of comments in just the first four months numbered 1231. This also represents a measure of interactivity between users and editorial staff, the user's interest in specific topics, and indicates what users find relevant.

Table 1: Interventions to articles of the Croatian Encyclopedia

Year	Number of updated articles	Number of new articles
2014	14,438	248
2015	5,900	200
2016	4,193	176

3.2. Boundlessness of scope

Due to cost of print and practicality of use, traditional encyclopedic form is limited in scope, having abundance of abbreviations that crowd the text and at times having to omit interesting information. Contemporary works, however, have no such limitations, allowing the use of more natural text flow and mention of any and all relevant information while keeping great information density, characteristic of encyclopedic works. While digital encyclopedia is partly boundless in scope (digital storage is finite), web-based encyclopedia, practically speaking, is not. Boundlessness of scope of web-based encyclopedia is

¹⁵ The term scientific encyclopedia refers to some vocational or special encyclopedias that gather and process the materials of science, art, or specific field. E.g. by The Miroslav Krleža Institute of Lexicography Medical Encyclopedia (1st edition 1957–65, 8 vol.), Encyclopedia of Technology (1963–97, 13 vol.), or personal Krležijana (1993–99, 3 vol.) dedicated to Miroslav Krleža, Institute's founder and first director.

¹⁶ Hrvatska enciklopedija (The Croatian Encyclopedia). 11 vol. Zagreb: The Miroslav Krleža Institute of Lexicography, 1999–2009.

manifested in its dynamic growth and development through expansion of existing articles and addition of new ones. Printed editions are for the better part planned in advance in scope and number of volumes. Scope ceases to be crucial to work organization in digital editions, and work deviates from the strict lexicographical form, but it remains concise and highly informative. The Croatian Encyclopedia's corpus, during 2014–2016 period, grew by 27,577 lines of text¹⁷, including updates to existing articles and addition of new entries. The same, by each year, is shown in Table 2. The overall corpus growth is indicative of the mentioned boundlessness of scope and, if printed, these additions alone would equal around half of one printed volume of a multi-volume edition.

Table 2: Corpus growth of the Croatian Encyclopedia

Year	Total interventions (lines of text)
2014	9,052
2015	7,581
2016	10,944

Another example of an unlimited encyclopedia is Wikipedia. There are currently¹⁸ 295 active editions of Wikipedia in different languages, with a total of 44,092,120 articles. A complete printed edition would require an estimated 16,539 volumes. The largest edition by far is in English, with 5,363,179 articles, making up 12.2 % of Wikipedia, while the Croatian Wikipedia has 172,727 entries, or 0.4 %. The English Wikipedia increases in size each day for as many as 800 new entries.

3.3. Information retrieval and content connectivity

The importance of efficient and simple information retrieval is fully apparent in large, comprehensive encyclopedic works, especially where one article covers a multitude of terms. Tools for thorough, complex information retrieval provide an epistemological upgrade of the work, making it more efficient through better availability of information.

The basis of traditional encyclopedic content organization, along with the standard alphabetical order of articles, is the index - a list of all mentioned terms (titles), concepts, or keywords listed in some order (most often alphabetically), with reference to the article(s) where each is mentioned. Index is especially important when articles cover several terms or names making it less significant in encyclopedic dictionaries and lexicons and a necessity in large general encyclopedias and encyclopedic atlases, allowing a specific concept (term) to be found even in large review articles that can cover several or even dozens of terms. In addition to its organizational function, the index plays a role in comparing dif-

¹⁷ A standard line of text contains 60 characters (with spaces).

¹⁸ https://meta.wikimedia.org/wiki/List_of_Wikipedias (accessed Apr 17th 2017)

ferent encyclopedic works, whereby the number of terms in the index, rather than the number of articles themselves, is often a good indicator of the extent of scope. In printed works, the index system is mundane - branched linear organization of knowledge, such as found in libraries and archives.

Retrieving information in a digital encyclopedia means access to the entire content promptly and efficiently, encompassing article titles and/or entire text. To maximize use and potency, data should be tagged, that is set up in a metadata network that the computer can process. Such a system's example is DBpedia, an ontological database and a semantic representation of Wikipedia. It represents a knowledge base that comprises of individuals (objects), classes (collections or object types), attributes (related properties, appearances, characteristics, or parameters that an object can have or distribute), and relations (the way objects are related to each other). Ontology is a formal representation of concepts with well-defined relationships between these concepts. Using the infoboxes within Wikipedia articles as structured content, DBpedia extracts metadata describing a large number of entities (persons, places, music, film, organizations) and it contains RDF triplets¹⁹ extracted from various multilingual editions of Wikipedia. It covers many domains, is updated alongside Wikipedia, is multilingual, and available on the web. Via search interface, the system provides exhaustive information retrieval for a given query as it includes the entire Wikipedia and not just the article titles and text that corresponds to the query's characters (letters).

Traditional encyclopedic content is internally interconnected through referrals and index, but not to outside content (other than with references). Digital works, other than having hypertext links, are no different while the contemporary web-based encyclopedia's content can be connected to any other information on the web. This greater content connection brings about a new epistemological value, making web-based encyclopedia a part of the internet's immense data-base. By structuring content, tagging it and setting up meta-data, connection on multiple levels²⁰ is enabled. Publishing such content on the web that computers can process facilitates connection to other databases on semantic level.²¹

Further advancement can be made using machine learning algorithms and tools for natural text processing, which have become quite efficient in analyzing raw

¹⁹ The RDF model is based on statements about resources known as triplets. Each is presented in the form of a subject-predicate-object statement, and can be constructed as a graph with two nodes (subject and object) connected with the predicate.

²⁰ DBpedia extracts metadata, or structured content from Wikipedia articles in form of info-boxes. By setting relationships between objects and associated attributes, a network of terms is created that represents Wikipedia content. Searching such a database provides a deep overview of the entire content.

²¹ Such kind of interconnectivity represents the Linking Open Data project: <http://lod-cloud.net/> (accessed Mar 8th 2017)

text and allow for the human-created ontologies to be omitted, but are yet to be further implemented in encyclopedistics.

5. Conclusion

This research shows that contemporary web-based encyclopedia is characterized by greater epistemological value regarding its update possibilities (work's continuity), collaboration potential, having unlimited scope, far more possibilities in information retrieval, and content connectivity options.

Increase in human cognizance and knowledge emphasizes the continuous need to organize, catalog, and synthesize knowledge, formatting it into the encyclopedic form, making encyclopedia a highly conductive knowledge-manual. Encyclopedia's overview and capacity for retrieving information provides a number of epistemological benefits: comprehensiveness in a given field, high expertise and relevance for a broad user circle, objectivity, and high accuracy, making it a reliable source of information. Encyclopedia's digitalization, and foremost its new web-based form, changed not only the scope of lexicographer's work, but also its epistemological features. Contemporary encyclopedias, being based on information technology of today, epistemologically surpass traditional encyclopedic form. The way users gather new knowledge has changed as well because web-based encyclopedia is now surrounded with a multitude of data circling the web. Web-based encyclopedia can now take on a role of a digital database. Software tools allow thorough search of content, interconnectivity with other knowledge sources, hypertext links, far more frequent updates, etc., making them an effective public service. Encyclopedia has changed and continues to change, or develop. This phenomenon can surely be named development since digital and especially web-based encyclopedias carry a number of benefits, including the quality of printed editions, thus becoming a place for easy access and exchange of trusted and connected content. Digitalization and globalization lead to networked and interconnected knowledge, eliminating obstacles to fast information retrieval. These features make encyclopedia the epistemological support in digital information. The future of web-based encyclopedia is in effective interconnectivity of its own content and linking to outside sources of knowledge. The responsibility of encyclopedia as an information source in the epistemological sense becomes greater, since they are now available to an even greater extent to anyone with an internet connection.

References

- Bizer, Christian; Heath, Tom; Berners-Lee, Tim. Linked Data – The Story So Far. (2009) <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (May 5th 2017)
- Featherstone, Mike; Venn, Couze. Problematizing Global Knowledge and the New Encyclopedia Project. *Theory, Culture & Society*. 23 (2006), 2–3, pg. 1–20.
- Jecić, Zdenko; Boras, Damir; Domijan, Darija. Prilog definiranju pojma virtualna enciklopedija. *Studia lexicographica*. 2 (2008), 1(2), pg. 115–126.
- Jecić, Zdenko. Enciklopedijski koncept u mrežnom okruženju. *Studia Lexicographica*. 7 (2014), 2(13), pg. 99–115.
- Kubelka, Ozren; Šoštarić, Petra. Wikipedija nasuprot Hrvatskoj enciklopediji, kvalitativan odnos slobodno i tradicionalno uređenoga enciklopedijskoga sadržaja a hrvatskome jeziku. *Studia lexicographica*. 5 (2001), 2(9), pg. 119–134.
- Thagard, Paul. Internet epistemology: Contributions of new information technologies to scientific research. (1997) <http://cogsci.uwaterloo.ca/Articles/Pages/Epistemology.html> (April 24th 2017)
- Allen, Colin; Jagodzinski, Cecile. *From SEP to SEPIA: How and why Indiana University is helping the Stanford Encyclopedia of Philosophy. Against the Grain*. 18 (2006), 4, pg. 42–43.
- Allen, Colin; Nodelman, Uri; Zalta, Edward N. The Stanford Encyclopedia of Philosophy: A Developed Dynamic Reference Work. *Metaphilosophy*. 33 (2002), 1–2, pg. 210–228.
- Hammer, Eric M.; Zalta, Edward N. A Solution to the Problem of Updating Encyclopedias. *Computers and the Humanities*. 31 (1997), 1, pg. 47–60.
- Kane, Gerald C.; Ransbotham, Sam. Collaborative development in Wikipedia. (2012) <https://arxiv.org/pdf/1204.3352.pdf> (Mar 21st 2017)
- Katz, William A. *Introduction to Reference Work, Volume I*. New York: Mc Grow Hill Book Company, 1978.
- Nodelman, Uri; Allen, Colin; Zalta, Edward N. Stanford Encyclopedia of Philosophy: A Dynamic Reference Work. *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*. New York: IEEE, 2003, pg. 380.
- Pasternack, Jeff; Roth, Dan. The Wikipedia Corpus. (2008) <http://cogcomp.cs.illinois.edu/papers/PasternackRo08.pdf> (Mar 7th 2017)
- Perry, John; Zalta, Edward N. Why Philosophy Needs a ‘Dynamic’ Encyclopedia. (Nov 1997) <https://plato.stanford.edu/pubs/why.html> (April 13th 2017)
- Tananbaum, Greg. I Hear the Train A Comin'. *Against the Grain*. 18 (2006), 1, pg. 84–85
- Torres, Diego; Molli, Pascal; Skaf-Molli, Hala; Diaz, Alicia. Improving Wikipedia with DBpedia. SWCS - Semantic Web Collaborative Spaces Workshop 2012 in 21st WWW Conference. Egyed-Zsigmond, E.; Gripany, Y.; Favre, C.; Langeron, C. (ed.). Lyon: ACM, 2012, pg. 1107–1112
- Voss, Jakob. Measuring Wikipedia. *Proceedings of the 10th International Conference on Scientometrics and Informetrics*. Ingwersen, P.; Larsen, B. (ed.). Stockholm: Karolinska University Press, 2005, pg. 221–231
- Zalta, Edward N. The Stanford Encyclopedia of Philosophy: A University/Library Partnership in Support of Scholarly Communications and Open Access. *College & Research Libraries News*. 67 (2006), 8, pg. 502–504.
- Zalta, Edward N. Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/prc-oct99.html> (Mar 10th 2017)