What Does a "Reliable Source" Mean? Types and Structure of References in Polish Wikipedia Articles About Historical Persons

Bartłomiej Włodarczyk
ORCID 0000-0001-9229-4656
Department of Bibliography and Documentation,
Faculty of Journalism, Information and Book Studies,
University of Warsaw, Poland

Abstract

Purpose/Thesis: The paper aims to characterize references to different types of sources featuring in select Polish Wikipedia articles from the category of people related to the Austrian Partition and its subcategories.
Approach/Methods: The data sample comprised references from 50 randomly selected articles from the Polish Wikipedia, including 1007 citations and 758 references. The references have been collected, processed, and analyzed with the use of R language. After they have been categorized, descriptive statistics were provided and analyzed.
Results and conclusions: The study shows that the majority of sources used in the research sample were primary sources. Consequently, it demonstrates that the analyzed articles about historical persons can be considered to be a product of research rather than simple derivative work, if only to a certain extent. Polish Wikipedians relied mainly on government directories and newspaper or magazine articles, often available in digital libraries. Secondary sources, on the other hand, consisted mostly of books, webpages, and book sections. The references were variously structured, and bibliographic descriptions sometimes lacked important elements. The findings confirm previously observed difficulties of analyzing sources in Wikipedia. Moreover, they support the argument that it is necessary to research different editions and subject areas of the largest online encyclopedia.
Research limitations: Due to the exploratory character of research, focusing on references from selected articles about historical persons from Poland, one cannot readily extrapolate its results to other parts of the Polish Wikipedia. Additionally, the research sample comprised only citations and references collected at one specific point of time.
Originality/Value: Most of the studies on the sources used in Wikipedia articles have focused on its English edition. Moreover, articles about historical persons in this encyclopedia have not been analyzed from the perspective of sources used, their types, and reference patterns. The paper broadens the understanding of the way the sources are used in Wikipedia by focusing on the Polish edition of the encyclopedia.

Keywords

Citations. Historical persons. Historical sources. Polish Wikipedia. References.

Received: 01 October 2020. Reviewed: 15 October 2020. Revised: 20 November 2020. Accepted: 30 November 2020.

1. Introduction

Wikipedia is one of the most popular Web resources regardless of its language edition. It held the thirteenth place in Alexa global traffic ranking in June 2020 (*Top Sites*, 2020), only below the websites owned by the biggest and most important companies such as Google, Alibaba Group, and Facebook. However, its popularity varies across countries. For instance, it occupied eighth place in the United States (*Top Sites in United States*, 2020) and ninth position in Poland (*Top Sites in Poland,* 2020). This amount of traffic makes Wikipedia the most significant collaboratively-made online reference work. The popularity of articles from different domains also differs to some extent. According to a laboratory experiment on temporal information searching behavior, Wikipedia is a basic source of past-related information (Joho et al., 2015) designed and implemented temporal-aware systems and solutions, understanding of people's temporal information searching behaviour is still limited. This paper reports the findings of a user study that explored temporal information searching behaviour and strategies in a laboratory setting. Information needs were grouped into three temporal classes (Past, Recency, and Future).

Due to Wikipedia's significance as primary online reference material, its reliability is an important issue. Regardless of the edition, Wikipedia community tries to ensure that the encyclopedia remains reliable by following a set of rules, beginning with the so-called five pillars which should guide the authors. The 'neutral point of view' policy requires that authors include different views on objects and phenomena described in the encyclopedia. According to this rule, Wikipedians should present all points of view basing on published sources (*Wikipedia:Pięć...*, 2020). Other policy intends to guarantee verifiability, which means that Wikipedians must locate reliable sources to support their statements, i.e., provide references and citations (*Wikipedia:Weryfikowalność*, 2020). The principle is strengthened by the ban on the original research and own interpretations (*Wikipedia:Nie ...*, 2020).

Polish Wikipedia guidelines indicate which types of sources should be cited in articles, and how they should be described. First of all, an article should list all sources that have been consulted by its author. Wikipedians have to provide all the details needed to find a source, e.g., ISBN or the relevant page numbers. For instance, a description of a Web page should include a link, a full bibliographic description, and an access date. Even if the link is not available, the reader should have a chance to find the resource (*Wikipedia:Bibliografia*, 2020). Wikipedia's help pages contain general instructions regarding citations (*Pomoc...*, 2020).

Wikipedia also provides more detailed guidelines regarding sources. According to these guidelines, the sources must be published, reliable, and current. Polish materials have priority in the Polish edition of Wikipedia (*Wikipedia:Weryfikowalność*, 2020). However, Polish Wikipedia rules are not as elaborate as those in the English edition. One of the websites devoted to the reliability of the sources currently has the status of a proposal, not a standard that needs to be implemented (*Wikipedia:Wiarygodne...*, 2020).

This paper characterizes references to different types of sources cited in select Polish Wikipedia articles in the category of "People related to the Austrian Partition" and its subcategories. The analysis shows what sources are used by Polish Wikipedians writing historical articles. This exploratory study seeks to answer the following research questions:

(1) What types of sources are used in Polish Wikipedia articles about historical persons living in the Austrian Partition (18th–20th centuries)? (Section 4.1)
(2) What is the structure of references in these articles? (Section 4.2)

The remainder of the article is organized as follows: Section 2 reviews the literature on the citations, references, and historical persons in Wikipedia. Section 3 presents the methodology employed in the research. The next section contains the results of the analysis and answers the above-mentioned research questions. Section 5 discusses the results. The paper closes with a summary of conclusions to be drawn from the research.

2. Literature review

As the paper analyzes references in history articles, the review will focus on the studies on the use of sources in such articles, with cursory investigation of articles in different disciplines. A number of studies addressed a range of issues related to citations and references in Wikipedia articles. Firstly, it should be emphasized that bibliographic references in Wikipedia are rarely standardized. Pooladian & Borrego (2017, 459), who studied the use of citations to articles from library and information science journals, stressed that "The degree of completeness of the references varies from entry to entry, even for a single article." They provided examples of incomplete references to scientific articles present in the English version of Wikipedia.

The types of sources consulted and the number of references are different in Wikipedia than they are in published academic texts; in Wikipedia, they vary between different areas, fields, and language versions. In general, more references are provided for articles about objects and phenomena related to the country where the specific language edition originated, as exemplified by the entries on cities from five versions of Wikipedia (Lewoniewski et al., 2017). The study by Ford et al. (2013) examined types of sources used in English Wikipedia, with randomly selected 500 citations as its data sample. According to the authors, Wikipedians tend to rely on sources not regarded as reliable by scholars, such as governments – and associations-related sources, collaboratively-created websites, and other non-traditional sources. The authors also showed that, although Wikipedia rules prioritized secondary sources editors often referred to primary sources. The most commonly used primary sources included different data and statistics (Ford et al., 2013). The research conducted by Kousha & Thelwall (2017) showed that Wikipedians cited a larger part of monographs (33%) indexed in the Scopus base than of the articles (5%). The authors argued that "there were considerable disciplinary differences in the extent to which academic publications were cited in Wikipedia. Monographs were cited particularly often in the arts and humanities (48%) and in the social sciences (39%), probably due to the cultural or educational values of the books that were targeted at, or accessible to, students or a wider public" (Kousha & Thelwall, 2017, 775). However, the authors of Wikipedia entries do not limit themselves to monographs and journal articles. According to Huvila (2010), although they preferred online sources, they also used other resources, e.g., the literature they were familiar with. They also referred to personal experiences and reports from acquaintances as sources of information. Experts in specific domains, e.g., history, also appeared among the sources. Other studies demonstrated the prevalence of references to online sources

in Wikipedia articles. Kelly (2018) showed that digital library items, mainly documents, were among the most cited resources from cultural institutions in her study of the use of sources from Louisiana Digital Library – a library of the consortium of institutions from the state of Louisiana. Sport- and government-related content, and digitized newspapers were the most popular objects among single-item records and collection-specific hyperlinks.

Outside these English-focused studies, scholars have analyzed other language versions of Wikipedia. For instance, Noč & Zumer (2014) examined a sample of featured articles from Slovene Wikipedia to determine the type and language of cited sources. Since the majority of Slovene articles were adapted from the English Wikipedia, the majority of sources was in English. Popularity of different types of source depended on the topic of an article. For instance, articles on sports mostly cited newspaper articles, while articles about scientific and historical topics cited books. Whatever the topic of article, webpages were evenly used by different entries in Slovene Wikipedia.

The position of history among other fields of research becomes apparent in citation analysis. Torres-Salinas et al. (2019) created a co-citation map of the humanities as featured on Wikipedia. According to the researchers, "History is presented as the main knowledge Domain from a social point of view. It concentrates the largest number of citations of individual journals (531) and scientific articles (11661), and the highest number of total citations (15969). Co-citation analysis also places it in a central position, connecting specialties" (Torres-Salinas et al., 2019, 802). The study relied on the information provided by Altmetric.com and Scopus (Torres-Salinas et al., 2019), which cannot be used to analyze citations in the Polish Wikipedia as these bases index mainly English language resources.

Scholars have also compared citations in history articles published on Wikipedia with the citations in articles published by academic journals. Luyt & Tan (2010) those that are suffer from the choice of references used. Many of these are from only a few US government Websites or news media and few are to academic journal material. Given these results, one response would be to declare Wikipedia unsuitable for serious reference work. But another option emerges when we jettison technological determinism and look at Wikipedia as a product of a wider social context. Key to this context is a world in which information is bottled up as commodities requiring payment for access. Equally important is the problematic assumption that texts are undifferentiated bearers of knowledge. Those involved in instructional programs can draw attention to the social nature of texts to counter these assumptions and by so doing create an awareness for a new generation of Wikipedians and Wikipedia users of the need to evaluate texts. Hence, citations analyzed 50 randomly selected English Wikipedia entries about history of different countries and compared them with the articles from *Journal of World History* (JWH). The comparison showed that the citations per hundred words ratio was significantly lower in Wikipedia entries (0.3/100 for Wikipedia to 1/100 for JWH). Furthermore, only a small proportion of statements in Wikipedia included any citations (4.86%). Moreover, Wikipedians' references were primarily Web-based, which differed from the practice followed by professional historians in the history journal. The sources were mostly in English (91%). The study showed a distinctive characteristic of citations and references in Wikipedia entries.

Researchers have also explored Wikipedia articles about specific historical persons. Callahan & Herring (2011) examined sixty entries about famous persons in the English and Polish Wikipedias to assess the effect of cultural differences on the articles' character

and content. Their analysis revealed the presence of cultural biases in the analyzed entries. According to the authors, they can be attributed to "the recent political and economic histories of the United States and Poland, which shape the contributors' values in systematic ways" (Callahan & Herring, 2011, 1913). In addition, as the authors suggested in their analysis of Wikipedia policies on information content and sources referenced in entries, "what constitutes «significant views» and «reliable sources» may vary across cultures" (Callahan & Herring, 2011, 1913).

Many other studies have discussed the quality of references and citations in Wikipedia articles on historical topics. Rector (2008), who analyzed articles on the English Wikipedia, found examples of unattributed quotes and plagiarism. Moreover, she showed that the authors of entries sometimes used outdated or non-credible sources. For instance, the biographical article about William Kidd referred to a children's book. Similarly, Rosenzweig (2006) noticed that, although English Wikipedia historical articles included references, they were not selected carefully enough, i.e., there were better sources that were not consulted. The author used as an example the bibliography attached to the entry on Haym Salomon which included only two books, one of which contained significant mistakes. The focus of the specific sources and the emphases in the common understanding of an event occasionally had an undue influence on the content of the entire article. For instance, an entry about the Spanish-American War devoted a significant space to the cause of Maine's sinking, a topic investigated by the *National Geographic* and described by media. However, it did not refer to Kristin L. Loganson's book entitled *American Manhood: How Gender Politics Provoked the Spanish-American and Philippine-American Wars*, which according to Rosenzweig, presented new arguments about gender-related causes of the war. The author claimed that this second source was more important for professional historians (Rosenzweig, 2006).

A fundamental question is: which sources are perceived as reliable by the authors of historical articles. Luyt (2015, 441) analyzed selected discussions from Wikipedia talk pages concerning Vietnam War "to comment on how the wider epistemological context of Wikipedia affects the nature of the debate about sources and through that, the actual choice of sources." He emphasized the shallowness of these debates which departed from the topic of the value of sources used to support the content of entries. Regarding the discussion about North Vietnamese land reform on Wikipedia, Luyt (2015, 449) pointed out that Wikipedia "changes the criteria we use to judge expertise [...] without replacing them with much that could be construed as progressive." Although Luyt made clear that his conclusions could not be used to generalize about all the articles from Wikipedia, they nevertheless provided valuable insight into the use of sources in history entries. The sources regarded as reliable by professional historians are not necessarily recognized as such by Wikipedians.

The studies discussed above give us a better understanding of the nature of citations and references in the largest online encyclopedia. However, they are clearly limited. Most of the studies of sources used in Wikipedia history articles have only considered its English edition. Moreover, discussions of the articles about historical persons have not focused on the sources used, their types, and reference patterns.

3. Methods

The research presented in the current paper had several stages. The first step was to gather all the articles related to historical persons from the selected period of Poland's history (1795–1918). The Polish Wikipedia includes a specific category called "Historia Polski" (History of Poland) divided into subcategories devoted to fixed periods, including the one between the Third Partition of Poland (1795) and regaining independence in 1918. The first subcategory considered was "Ludzie związani z zaborem austriackim" (Persons related to the Austrian Partition). R packages "rvest" (Wickham, 2019) and "WikipediR" (Keyes & Tilbert, 2017) were used to collect the titles of all articles from this category and its subcategories, together with the names of all assigned categories. The cut-off date was June 17, 2020, 8 a.m. The data was cleaned manually to exclude all articles unrelated to persons. The starting data set consisted of 2244 articles. Then, 50 numbers between 1 and 2244 were randomly selected to choose the sample of articles. The entries discussed persons of different types, such as politicians and rulers, officials, soldiers, landowners, teachers, and members of various organizations from different towns and cities.

Polish Wikipedia articles have two sections listing sources used in writing of the entries: citations and references. These sections should include all materials the entry has been based on. However, different entries are structured in different ways. Some articles contain both citations and references, some – only citations, and some – only references. The sources listed in one section are rarely the same as the sources listed in the other. Descriptions from both sections were automatically collected for further analyses. The name of the article, full bibliographical citations and/or references, as well as Web addresses were recorded. Internal Wikipedia links and links that did not refer to full texts, e.g., Worldcat links, were excluded. Next, the data was cleaned. For instance, if a citation included several sources, it was cut into several independent records. Author-data citations were replaced by descriptions from the reference section. Then, the sources were separated into unique and non-unique. For instance, if a source A is cited in an article twice, it is counted as one unique source. However, if the source B is used in two independent articles, it is counted twice. The selection of unique materials took into account the completeness of bibliographical description i.e., if several descriptions of the same source were present in an article, the fullest description was selected for further analysis. The term "citation" is used in the next part of the paper to refer to both unique and non-unique sources, while "reference" denotes a unique source. In total, 1007 citations and 758 references were collected.

The next stage was categorizing the data. Most of the work was done manually with the use of Excel spreadsheet software. Other tools used will be indicated when appropriate. The language of the sources was determined with the use of the package "textcat" (Hornik et al., 2018). The data obtained was corrected manually. All sources containing Web addresses were categorized as electronic (Web-based), and all the others as traditional (paper-based). The presence of link rot was determined with URLitor service (http://www.urlitor.com), used to gather HTTP statuses. Web addresses were divided according to the individuals or institutions managing content such as an association, a government, or a museum. Additionally, some categories were also divided into Polish and foreign (other). Digital libraries were distinguished from other websites. Historians use sources

of two kinds: primary and secondary. The primary sources are all the materials produced during the studied period, whereas the secondary sources are based on primary sources, produced later, and consist of all research works about the period. This division is widely accepted in historical research, but it should be noted that it is context-related to a great extent (Cullen, 2009). The distinction between primary and secondary resources was factored into the current study. Next, types of sources cited, such as a book, a periodical article, a manuscript, and a webpage were distinguished. Additionally, primary sources were divided into subtypes such as a government directory, a personal paper, or a report. Lastly, reference patterns, i.e., the sequence and types of elements, were listed. The data sample was analyzed by means of descriptive statistics, using R language, e.g., employing package "dplyr" (Wickham et al., 2020). It is published and available online (Włodarczyk, 2020). A detailed description is provided in the appendix.

4. Results

4.1. Types of sources used in references

References in the selected Wikipedia articles can be divided according to different features. Firstly, they can be separated into references to traditional and electronic sources. The data set includes 283 traditional, i.e., paper-based sources (37%), and 475 electronic, i.e., Web-based sources (63%). The cited Web-based materials were managed by different owner types, as shown in Table 1.

Tab. 1. Owner types of Web-based references

Owner type	Number of references	Percentage of electronic references
Library (PL)	239	50.3%
Association	95	20.0%
Library (other)	56	11.8%
Other	33	6.9%
Personal	22	4.6%
Government (PL)	10	2.1%
Local government	8	1.7%
Government (other)	5	1.1%
Museum	4	0.8%
Cemetery	3	0.6%
Total	475	

The data shows that the most popular sources of references in the data set were library and association websites. The first group included mainly Polish digital libraries such as Jagiellonian Digital Library, Sanok Digital Library, Wielkopolska Digital Library, Podkarpackie Digital Library, and others. Foreign digital libraries, such as ANNO – AustriaN

Newspapers Online, ALEX, and Hungaricana, featured less often. Association-related resources occupy the second position in terms of prevalence; they consisted mainly of periodicals digitized by Małopolskie Towarzystwo Genealogiczne. Other owner types provided only a low number of materials. Taken together, the data shows that Polish Wikipedians tended to rely on digital libraries for most of the electronic references.

The analysis also accounts for the phenomenon of link rot. Substantial majority (465, i.e., 98%) of Web resources used as references was still available at the time the current research was conducted. Eleven Web resources had been archived. The remaining addresses yielded failure HTTP final statuses such as 404 (5 resources), site not available (3 resources), 403 (1 resource), and connection time out (1 resource). Out of 10 unavailable resources, only two cannot be found in the Internet Archive Wayback Machine (http://web.archive.org).

Native sources are preferred in the Polish Wikipedia, and consequently, the bulk of references is in Polish. Table 2 presents a more detailed distribution of references according to language.

Tab. 2. Distribution of references according to language

Reference language	Number of references	Percentage of references
Polish	689	90.9%
German	53	7.0%
Latin	10	1.3%
Ukrainian	5	0.7%
Russian	1	0.1%
Total	758	

As the analyzed articles came from the category of historical persons who lived in the Austrian Partition, German publications constituted an important part of references alongside the Polish sources. They comprised official government publications and newspaper articles. Latin sources consisted of parish registers. Lastly, Ukrainian and Russian references included encyclopedias, a dictionary, and published primary sources. The language use followed the Wikipedia recommendations.

Polish Wikipedians used both primary and secondary sources. The data set in this study consists of 530 primary sources (70%) and 228 secondary sources (30%). Majority of primary sources was Web-based (369, 70%); paper-based (122, 54%) and Web-based (106, 46%) secondary sources were more or less equally prevalent.

The analysis distinguished between different types of primary and secondary sources. The composition of primary sources is presented in Table 3.

The most popular primary sources among Polish Wikipedians were periodicals and articles published therein. Together, they accounted for 88% of primary sources and 61% of all sources, so they made up the vast bulk of all references. The cause of their popularity might have been their accessibility as Web resources. Other types of sources were used only occasionally.

The primary sources can also be divided into subtypes. Table 4 shows the division into subtypes.

Tab. 3. Types of primary sources

Type of primary source	Number of primary sources	Percentage of primary sources	Percentage of all sources
Periodical	320	60.4%	42.2%
Periodical article	146	27.5%	19.3%
Book	33	6.2%	4.4%
Manuscript	22	4.2%	2.9%
Webpage	6	1.1%	0.8%
Other	3	0.6%	0.4%
Total	530		

Tab. 4. Subtypes of primary sources

Subtype of primary source	Number of primary sources	Percentage of primary sources	Percentage of all sources
Directory, government	215	40.6%	28.4%
Newspaper/magazine article	136	25.7%	17.9%
Report	44	8.3%	5.8%
Directory, military	37	7.0%	4.9%
Legal document	36	6.8%	4.7%
Record, church	15	2.8%	2.0%
Other	15	2.8%	2.0%
Directory, other	10	1.9%	1.3%
Personal paper	9	1.7%	1.2%
Gazetteer	5	0.9%	0.7%
Record, school	4	0.8%	0.5%
Record, other	2	0.4%	0.3%
Record, city	1	0.2%	0.1%
Record, government	1	0.2%	0.1%
Total	530		

Government directory was the most popular subtype, with 215 references, i.e., 40.6% of all cited primary sources. This category contained different official staff yearbooks published by the government. Different editions of *Schematyzm Galicyjski* were the most frequently cited publications of this kind. Governorship in Lviv published different volumes of this yearbook between 1849 and 1914. They contained basic information about officials, clergy, and members of different institutions, associations, and societies (Kramarz, 2007). The second most prevalent type of sources included newspaper, magazine, and journal articles, which accounted for 25.7% of all primary sources. Wikipedians frequently cited daily and weekly newspapers such as *Gazeta Lwowska*, *Nowa Reforma*, *Nowości Illustrowane*, and *Tygodnik Ziemi Sanockiej*. The most often cited articles were short news from towns and

cities of the Austrian Partition. Less popular were magazines and journals, which included *Śpiewak* (a monthly magazine about singing), *Przewodnik Kółek Rolniczych* (a monthly magazine about agriculture), and *Przegląd Lekarski* (a monthly journal about medicine). The next position in terms of citing is occupied by school reports and, to a lesser extent, different associations. Legal documents included statutes and military commands from the sources such as *Monitor Polski*, a Polish official journal. The remaining sources included military and other directories providing personnel lists. Archival materials consisted mostly of parish registers, such as registers of marriages and burials from the Parish of the Transfiguration in Sanok. Personal papers such as diaries, journals, or letters were not used much.

Tab. 5. Subtypes of primary sources divided according to types

Type of primary source	Subtype of primary source	Number of primary sources	Percentage of the primary source type
Periodical	Directory, government	213	66.6%
	Report	42	13.1%
	Directory, military	34	10.6%
	Legal document	23	7.2%
	Directory, other	4	1.3%
	Other	4	1.3%
Periodical article	Newspaper/magazine article	135	92.5%
	Legal document	9	6.2%
	Personal paper	1	0.7%
	Other	1	0.7%
Book	Personal paper	8	24.2%
	Directory, other	6	18.2%
	Gazetteer	5	15.2%
	Other	5	15.2%
	Directory, military	3	9.1%
	Directory, government	2	6.1%
	Legal document	2	6.1%
	Report	2	6.1%
Manuscript	Record, church	14	63.6%
	Record, school	4	18.2%
	Record, other	2	9.1%
	Record, city	1	4.5%
	Record, government	1	4.5%
Webpage	Other	4	66.7%
	Newspaper/magazine article	1	16.7%
	Record, church	1	16.7%
Other	Legal document	2	66.7%
	Other	1	33.3%
Total		530	

The subtypes were also analyzed in the relation to their respective types. Table 5 supplements previous figures presenting subtypes of primary sources divided according to types.

In all cases, one specific subtype was cited significantly more often than others. The most common subtype of periodicals cited were the government directories. The most commonly cited subtype of periodical articles was a newspaper or magazine article. Other directories, i.e., those not directly connected with the government or army, constituted the most commonly cited subtype of books, church records – of manuscripts. However, each type comprised many other subtypes.

The most frequently used types of secondary sources were different than the most frequently used types of primary sources. They are presented in Table 6.

Tab. 6. Types of secondary sources

Type of secondary source	Number of secondary sources	Percentage of secondary sources	Percentage of all sources
Book	104	45.6%	13.7%
Webpage	70	30.7%	9.2%
Book section	32	14.0%	4.2%
Periodical article	21	9.2%	2.8%
Other	1	0.4%	0.1%
Total	228		

As in the case of primary sources, there were two types of secondary sources more common than others; however, the difference between the numbers of references to the sources of second and third most common type were smaller. Books, book sections, and periodical articles, regarded as main sources of professional historians (Dalton & Charnigo, 2004; Kolasa, 2012; Mendez & Chapman, 2006), constituted 68.9% of all secondary sources in the data set. Webpages accounted for one-third of all secondary sources in the sample, suggesting they are an important source for Polish Wikipedians.

Secondary sources were more varied than primary sources. Wikipedians cited both contemporary and older monographs. For instance, they cited both Józef Buszko's book (1996) about Poles in the parliament in Vienna, and Antoni Kurka's book (1930), about the police in Lviv during the Austrian rule. Entries cited many chapters from an edited book about history of Sanok (Kiryk, ed., 1995). The entries cited a variety of websites, including Web biographical dictionary on the Austrian Parliament website (https://www.parlament.gv.at/WWER/) and an online grave locator of Rakowicki Cemetery (http://www.rakowice.eu).

4.2. Structure of references

It is a fundamental rule that, within a given text, references should be formatted according to one standard. This paper analyzes the structure of references in relation to the type of the source, regardless of whether it was primary or secondary. Bibliographical descriptions of periodicals mainly consisted of a periodical title (320 references; 100% of all periodical references), page numbers (301, 94%), a publication date (218, 68%), and a publication place (197, 62%). Other elements such as a publisher (37, 12%) or a periodical issue or

volume (24, 8%) were rare. The most frequently used pattern was as follows: periodical title, publication place, publication date, and page numbers (145 occurrences, 45%). Another common pattern (97, 30%) included only periodical title and page numbers.

Periodical article descriptions contained more elements than descriptions of periodicals. There, the most often used elements were publication date (165, 99% of all periodical article references), periodical title (164, 98%), article title (161, 96%), page numbers (159, 95%), and periodical issue or volume (157, 94%). The absence of the name of the article's author might be explained by the character of the sources, which were mainly anonymous articles. The position of the issue's number and page number changed across citations.

References to books were structured according to a different pattern than those discussed so far. The most often used elements were publication date (130, 95% of all book references), title (129, 94%), publication place (116, 85%), author (108, 79%), and page numbers (97, 71%). Authors did not always feature in a citation, as the books did not always directly indicate them. Other more frequently used elements were publisher (71, 52%) and ISBN (45, 33%). The remaining elements, such as edition, language tag, or OCLC number, were rare. Some references identified the publisher, but not the publication place, constructed as following: Authors-Title-Publisher-Publication date. Note that this description does not include page numbers. However, the data set also includes a description containing only the title of the publication. In general, bibliographical descriptions of books tended to include all elements needed to identify a publication.

References to individual book sections differ from the references to entire books as they need to include the title of collective work alongside the title of the section cited. The most frequent elements in the sample included title of a chapter or section (31, 97% of all book section references), collective book title (31, 97%), publication date (30, 94%), author (29, 91%), page numbers (29, 91%), and publication place (26, 81%). Other less frequent elements were editor (15, 47%) and publisher (11, 34%). Overall, most references included all required elements; however, as was the case with references to books, there were some gaps in the descriptions.

References to webpages require the least number of parts. The most prevalent elements were Web address (76, 100% of all webpage references), webpage title (71, 93%), access date (65, 86%), and domain name and/or website name (63, 83%). Wikipedians' references to websites were more consistent than their references to sources of other types. However, certain references only included a website name.

The last type of a source analyzed in this section is a manuscript. Most descriptions included three basic elements: title (22, 100% of all manuscript references), page numbers (20, 91%), and place of creation or storage (19, 86%). Three references did not identify the place of storage or the reference code. One reference to archival materials only included a title, one – a title and a creation date, and one – a title and page numbers. The Polish Wikipedia has only four templates for referencing and citing: webpage, book, journal, and the "universal" model, which makes the editors' choice very limited, especially compared with the English Wikipedia, which provides more models.

5. Discussion

The article presents the types and structure of references used in select Polish Wikipedia articles about historical persons living between the 18th and the 20th century in the Austrian Partition. Some of the findings are similar to those from the previous investigations into the sources used in Wikipedia articles. The majority of sources in the data sample is in Polish, in a situation analogous to that in the study of Luyt & Tan (2010), who found that the sources cited most commonly in the articles published on the English Wikipedia were in English as well. However, as Noč & Zumer's (2014) study of Slovene Wikipedia has shown, the language version of the encyclopedia does not have to determine the language of the sources used in the entry. Other studies have confirmed the popularity of Internet-based sources (Huvila, 2010; Luyt & Tan, 2010). This study has shown that the most commonly used sources are materials shared by Polish and foreign digital libraries. According to Kelly (2018), digital library items were the most popular sources from Louisiana Digital Library used in Wikipedia.

The distinction between primary and secondary sources is especially significant for historians. The investigation by Ford et al. (2013) showed that the authors of entries published on the English Wikipedia made extensive use of primary sources. The secondary and tertiary sources (encyclopedias and other reference works) accounted for 66% of all sources, with primary sources constituting 34% of all materials. However, their data was randomly selected from the entirety of the English Wikipedia. This paper shows that the number of references to primary sources (70% of all references) is substantially higher than the number of references to secondary sources. It proves that the analyzed articles about historical persons can be regarded as a product of research rather than simple imitative work, if only to a certain extent. The use of primary sources involves a careful analysis of their reliability, and therefore, it is more complicated than work with secondary sources, especially for a person not trained as a historian. The differences between professional historians and students' reading practices were mentioned by Luyt & Tan (2010), who emphasized the importance of appropriate information literacy training.

The types of primary sources used by Wikipedians are of particular interest. Wikipedians favor periodicals and periodical articles. They tend to rely on government directories and short news from newspapers and magazines. Different volumes of *Schematyzm Galicyjski* are commonly cited. Kramarz (2007) stressed that these publications are regarded as reliable and valuable sources by contemporary historians. However, they are not infallible: Kramarz mentioned that government periodicals contained errors, e.g., in family names. Some Wikipedians have noticed these errors, which suggests that they are capable of analyzing sources. For example, one of the authors of the article about Ludzimił Trzaskowski noticed that Trzaskowski was mentioned twice in *Schematyzm* from 1911, under two different first names (*Ludzimił ...*, 2020). Other types of primary sources used, such as different types of directories, provide similar content. They simply list items alongside basic information. Information from newspaper articles is more difficult to assess. Personal papers were not commonly cited in the articles from the sample. Apart from the published sources, the articles also cited some archival materials, but only rarely. Only some are available online in a digital library or on a website providing descriptions and scans from Polish archives (https://www.szukajwarchiwach.gov.pl/en/strona_glowna). This too suggests that Wikipedians use similar sources as professional historians.

The secondary sources constituted the minority of references in the analyzed articles. Books, webpages, and book sections were the main secondary sources (90.4% of this type of sources) used by Polish Wikipedians writing articles about historical persons from the Austrian Partition. To an extent, this result confirms the findings of Luyt & Tan (2010), who showed that books are the most popular materials among non-Internet based references. The popularity of Web resources has been discussed above.

The Polish Wikipedia provides some guidelines regarding correct and full bibliographic description of sources, but not very elaborate. The analysis shows that these guidelines were not always followed. Most book descriptions consisted of title, publication date, page numbers, and publication place, but lacked ISBN, which is identified as a required element of a full description in the Wikipedia documentation (*Wikipedia:Bibliografia*, 2020). Webpage descriptions also often failed to follow the guidelines in their entirety: they tended to include webpage title, Web address, and access date, and, rarely, domain or website name. The elements of descriptions were not always given in the same sequence, e.g., in references to articles, page numbers were given before or after periodical issue or volume. Such inconsistencies slow down the construction of citation databases, as many elements need to be corrected manually. Another challenge is posed by the sheer variety of the primary sources cited, as they require different descriptions for which the guidelines do not always provide a model.

The current study was limited in several ways. Due to the exploratory character of presented research, focusing on references from selected articles about historical persons from Poland, the conclusions should not be taken as a basis for extrapolating about articles in other categories of the Polish Wikipedia. Additionally, the research sample comprised only citations and references, which were collected at a specific point of time, rather than over a longer period.

Further studies may explore references in articles about different types of objects studied in history, such as events. It would also be interesting to compare the sources used in these areas across different language editions of Wikipedia. Finally, the differences in the use of sources related to history by different groups of Wikipedians can be examined.

6. Conclusion

This paper gives a picture of references in a selected area of the Polish Wikipedia related to the discipline of history, which academic literature has not yet offered. Previous studies have tended to focus on the English edition of the encyclopedia as it is the largest version worldwide. However, investigations into other editions may present a different image of sources used by Wikipedians. Referencing practices may also vary across different areas of Wikipedia, even within a single discipline such as history. For instance, sources used in articles about historical persons may be different from sources used in articles about countries. Conclusions yielded by studies of the most popular version cannot be simply extended to other editions. As analyzed articles rely mostly on primary sources, bibliometric analysis of these materials is not easy. This is also caused by the construction of citations and references. The lack of clear guidelines on what can be included in reference and citation sections does not help the matters. It is an area where the Polish Wikipedia can still improve.

References

Buszko, J. (1996). *Polacy w parlamencie wiedeńskim: 1848–1918*. Warszawa: Wydaw. Sejmowe.

Callahan, E. S., Herring, S. C. (2011). Cultural Bias in Wikipedia Content on Famous Persons. *Journal of the American Society for Information Science and Technology*, 62, 1899–1915, https://doi.org/10.1002/asi.21577

Cullen, J. (2009). *Essaying the Past: How to Read, Write, and Think about History*. Chichester: Wiley-Blackwell.

Dalton, M. S., Charnigo, L. (2004). Historians and Their Information Sources. *College & Research Libraries*, 65(5), 400–425, https://doi.org/10.5860/crl.65.5.400

Ford, H., Sen, S., Musicant, D. R., Miller, N. (2013). Getting to the Source: Where Does Wikipedia Get Its Information From? In: *Proceedings of the 9th International Symposium on Open Collaboration* (1–10). New York: Association for Computing Machinery, https://doi.org/10.1145/2491055.2491064

Hornik, K., Rauch, J., Buchta, Ch., Feinerer, I. (2018). *Package 'textcat'* [online]. The Comprehensive R Archive Network, [20.06.2020], https://cran.r-project.org/web/packages/textcat/textcat.pdf

Huvila, I. (2010). Where Does the Information Come From? Information Source Use Patterns in Wikipedia. *Information Research* [online], 15(3), [20.06.2020], http://informationr.net/ir/15-3/paper433.html

Joho, H., Jatowt, A., Blanco, R. (2015). Temporal Information Searching Behaviour and Strategies. *Information Processing & Management,* 51(6), 834–850, https://doi.org/https://doi.org/10.1016/j.ipm.2015.03.006

Kelly, J. E. (2018). Use of Louisiana's Digital Cultural Heritage by Wikipedians. *Journal of Web Librarianship*, 12(2), 85–106, https://doi.org/10.1080/19322909.2017.1391733

Keyes, O., Tilbert, B. (2017). *Package 'WikipediR'* [online], The Comprehensive R Archive Network, [20.06.2020], https://cran.r-project.org/web/packages/WikipediR/WikipediR.pdf

Kiryk, F., ed. (1995). *Sanok: dzieje miasta: praca zbiorowa*. Kraków: Secesja.

Kolasa, W. M. (2012). Specific Character of Citations in Historiography (Using the Example of Polish History). *Scientometrics*, 90(3), 905–923, https://doi.org/10.1007/s11192–011-0553–0

Kousha, K., Thelwall, M. (2017). Are Wikipedia Citations Important Evidence of the Impact of Scholarly Articles and Books? *Journal of the Association for Information Science and Technology*, 68(3), 762–779, https://doi.org/10.1002/asi.23694

Kramarz, H. (2007). Schematyzmy galicyjskie (1776–1914) jako c.k. rocznik sprawozdawczy dotyczący obsady kadrowej władz, urzędów, towarzystw i instytucji. *Rocznik Historii Prasy Polskiej,* 10(1), 5–29.

Kurka, A. (1930). *Dzieje i tajemnice lwowskiej policji z czasów zaboru austrjackiego: 1772–1918*. Lwów: Gubrynowicz.

Lewoniewski, W., Węcel, K., Abramowicz, W. (2017). Analysis of references across Wikipedia languages. *Communications in Computer and Information Science*, 756, 561–573, https://doi.org/10.1007/978-3-319-67642-5_47

Ludzimił Trzaskowski (2020, May 14). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Ludzimi%C5%82_Trzaskowski&oldid=59782094

Luyt, B. (2015). Debating Reliable Sources: Writing the History of the Vietnam War on Wikipedia. *Journal of Documentation*, 71(3), 440–455, https://doi.org/10.1108/JD-11-2013-0147

Luyt, B., Tan, D. (2010). Improving Wikipedia's Credibility: References and Citations in a Sample of History Articles. *Journal of the American Society for Information Science and Technology,* 61(4), 715–722, https://doi.org/10.1002/asi.21304

Mendez, M., Chapman, K. (2006). The Use of Scholarly Monographs in the Journal Literature of Latin American History. *Electronic Journal of Academic and Special Librarianship* [online], 7(3), [20.06.2020], http://southernlibrarianship.icaap.org/content/v07n03/mendez_m01.htm

Noč, M., Zumer, M. (2014). The Completeness of Articles and Citation in the Slovene Wikipedia. *Program*, 48(1), 53–75, https://doi.org/10.1108/PROG-12-2012-0069
Pomoc:Przypisy (2020, April 19). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Pomoc:Przypisy&oldid=59499337
Pooladian, A., Borrego, Á. (2017). Methodological Issues in Measuring Citations in Wikipedia: A Case Study in Library and Information Science. *Scientometrics*, 113(1), 455–464, https://doi.org/10.1007/s11192-017-2474-z
Rector, L. H. (2008). Comparison of Wikipedia and Other Encyclopedias for Accuracy, Breadth, and Depth in Historical Articles. *Reference Services Review*, 36(1), 7–22. https://doi.org/10.1108/00907320810851998
Rosenzweig, R. (2006). Can History Be Open Source? Wikipedia and the Future of the Past. *The Journal of American History*, 93(1), 117–146. https://doi.org/10.2307/4486062
Top Sites [online]. *Alexa*, [20.06.2020], https://www.alexa.com/topsites
Top Sites in Poland [online]. *Alexa*, [20.06.2020], https://www.alexa.com/topsites/countries/PL
Top Sites in United States [online]. *Alexa*, [20.06.2020], https://www.alexa.com/topsites/countries/US
Torres-Salinas, D., Romero-Frías, E., Arroyo-Machado, W. (2019). Mapping the Backbone of the Humanities Through the Eyes of Wikipedia. *Journal of Informetrics*, 13(3), 793–803, https://doi.org/https://doi.org/10.1016/j.joi.2019.07.002
Wickham, H. (2019). *Package 'rvest'* [online], The Comprehensive R Archive Network, [20.06.2020], https://cran.r-project.org/web/packages/rvest/rvest.pdf
Wickham, H., François, R., Henry, L., Müller, K. (2020). *Package 'dplyr'* [online], The Comprehensive R Archive Network, [20.06.2020], https://cran.r-project.org/web/packages/dplyr/dplyr.pdf
Wikipedia:Bibliografia (2020, March 5). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Wikipedia:Bibliografia&oldid=58977386
Wikipedia:Nie przedstawiamy twórczości własnej (2020, March 8). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Wikipedia:Nie_przedstawiamy_twórczości_własnej&oldid=59014800
Wikipedia:Pięć filarów (2020, February 18). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Wikipedia:Pięć_filarów&oldid=58835261
Wikipedia:Weryfikowalność (2020, June 9). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Wikipedia:Weryfikowalność&oldid=60041593
Wikipedia:Wiarygodne źródła (2018, January 25). *Wikipedia, wolna encyklopedia* [online] [20.06.2020], https://pl.wikipedia.org/w/index.php?title=Wikipedia:Wiarygodne_źródła&oldid=52295345
Włodarczyk, B. (2020). *Data for: What Does a "Reliable Source" Mean? Types and Structure of References in Polish Wikipedia Articles about Historical Persons* (Version 2) [online]. RepOD, [20.11.2020], https://doi.org/10.18150/A2VH9M

Appendix

The data set has been constructed partly automatically, partly manually. The details of the process of data collection are provided in the paper, in the section on methodology. The following section defines the attributes and values in each data file.

The first file, entitled "wiki-persons-data-1.csv," includes the following attributes:
1. Number – an ordinal number of the selected Wikipedia article.
2. WikipediaArticleTitle – a title of the selected Wikipedia article.
3. Lifetime – a lifetime of the selected historical person.
4. ModificationDate – the last modification date before collecting the data (before June 17, 2020, 8 a.m.).
5. URL – a Web address of the selected Wikipedia article.

The second file, entitled "wiki-persons-data-2.csv," includes the following attributes:
1. Number – an ordinal number of the selected Wikipedia article.
2. WikipediaArticleTitle – a title of the selected Wikipedia article.
3. WikipediaCategory – a category/categories of the selected Wikipedia article.

The third file, entitled "wiki-persons-data-3.csv," includes the following attributes:
1. Number – an ordinal number of the selected Wikipedia article.
2. WikipediaArticleTitle – a title of the selected Wikipedia article.
3. UniquenessWithinTheArticle – the uniqueness of a source within the Wikipedia article, i.e., whether a source is a reference (values: yes, no).
4. Citation – a source from the selected Wikipedia article.
5. CitationLanguage – a language of a source from the selected article (values: German, Latin, Polish, Russian, Ukrainian).
6. ReferencePattern – a sequence and types of elements in the bibliographical description of the source (values:
   A – an author,
   B – a unique work title,
   C – a collective work title,
   D – a periodical title,
   E – a Web domain and/or a website name,
   F – an editor,
   G – a publication place,
   H – a publisher,
   I – a printer,
   J – a publication or creation date,
   K – a periodical issue/volume,
   L – page number(s),
   M – a place of creation or storage,
   N – an access date,
   O – a signature or a reference code,
   P – a work volume,
   R – a series name,
   S – ISBN,
   T – ISSN,
   U – a language tag,
   V – an edition,
   W – a file format,
   X – OCLC number).
7. URL – a Web address of the selected Wikipedia article.
8. HTTPStatusFinalDestination – HTTP status of the URL (the attribute mentioned above) (vales: 200, 403, 404, connection time out, site not available).
9. WebsiteOwnerType – a type of a website owner and manager (values: association, cemetery, government (other), government (PL), library (other), library (PL), local government, museum, other, personal (other), personal (PL)).
10. PolishOtherDigitalLibrary – a designation of the source of a citation as a digital library (values: digital library (other), digital library (PL)).
11. ElectronicTraditionalSource – a designation of the source medium (values: electronic, traditional).
12. PrimarySecondarySource – a designation of the source as primary or secondary (values: primary source, secondary source).

13. SourceType – a type of a source (values: book, book section, manuscript, other, periodical, periodical article, webpage).
14. PrimarySourceSubtype – a subtype of a primary source (values: directory, government; directory, military; directory, other; gazetteer; legal document; newspaper/magazine article; other; personal paper; record, church; record, city; record, government; record, other; record, school; report).

Czym jest „wiarygodne źródło"? Typy źródeł bibliograficznych i struktura ich opisów w artykułach dotyczących postaci historycznych w polskiej Wikipedii

Abstrakt

Cel/Teza: Celem artykułu jest opisanie typów źródeł bibliograficznych i struktury ich opisów, cytowanych w wybranych artykułach z polskiej Wikipedii należących do kategorii „Ludzie związani z zaborem austriackim" i wszystkich kategorii podrzędnych.
Koncepcja/Metody badań: Dane badawcze składały się z opisów bibliograficznych z 50 losowo wybranych artykułów z polskiej Wikipedii, w tym z 1007 cytowań oraz 758 unikalnych pozycji bibliograficznych. Pozycje te zostały pobrane, przetworzone i przeanalizowane głównie z użyciem języka R. Po skategoryzowaniu, zostały przedstawione i przeanalizowane statystyki opisowe dotyczące wybranych elementów.
Wyniki i wnioski: Badanie pokazuje, że większość materiałów w próbie badawczej stanowią źródła historyczne. W rezultacie okazuje się, że analizowane artykuły na temat postaci historycznych mogą być w pewnym stopniu uważane bardziej za produkt badań niż prostą odtwórczą pracę. Polscy wikipedyści używają głównie spisów rządowych oraz artykułów z gazet i czasopism, często pochodzących z bibliotek cyfrowych. Z kolei pozostałe zasoby składają się głównie z książek, stron internetowych oraz rozdziałów z książek. Struktura opisów bibliograficznych jest zróżnicowana i często brakuje w nich istotnych elementów. Wyniki potwierdzają problemy związane z analizą źródeł wykorzystywanych w Wikipedii. Ponadto, wskazują potrzebę badania różnych edycji i obszarów tematycznych największej encyklopedii internetowej.
Ograniczenia badań: W związku ze wstępnym charakterem badań, które skupiają się na źródłach bibliograficznych z wybranych artykułów dotyczących postaci historycznych z terenu Polski, nie powinno się ekstrapolować ich wyników na inne części polskiej Wikipedii. Dodatkowo, próba badawcza zawierała wyłącznie opisy, które zostały zebrane w jednym, określonym punkcie czasowym.
Oryginalność/Wartość poznawcza: Większość badań dotyczących źródeł wykorzystywanych w artykułach z Wikipedii była dotychczas ograniczona do edycji angielskiej. Ponadto, artykuły dotyczące postaci historycznych z tej encyklopedii nie były analizowane z perspektywy wykorzystywanych źródeł, ich typów oraz struktury opisów bibliograficznych. Artykuł poszerza zrozumienie użytkowania źródeł w Wikipedii poprzez skupienie się na polskiej wersji encyklopedii.

Słowa kluczowe

Bibliografia załącznikowa. Cytowania. Polska Wikipedia. Postacie historyczne. Źródła historyczne.

BARTŁOMIEJ WŁODARCZYK, PhD is Assistant Professor in the Department of Bibliography and Documentation in the Faculty of Journalism, Information, and Book Studies at the University of Warsaw. His main research interest is knowledge organization. His publications include Topic Map as a Method for the Development of Subject Headings Vocabulary: An Introduction to the Project of the National Library of Poland (2013). Cataloging & Classification Quarterly, *51(7), 816–829, doi: 10.1080/01639374.2013.801061;*

with M. Roszkowski. Cytowania zasobów sieciowych w polskich czasopismach z zakresu bibliotekoznawstwa i informatologii: analiza aktualności adresów URL (2016). Zagadnienia Informacji Naukowej–Studia Informacyjne*, 54(1), 21–43; KABA Subject Headings and the National Library of Poland Descriptors in Light of Wojciech Wrzosek's Theory of Historiographical Metaphors and Different Historiographical Traditions (2020).* Knowledge Organization*, 47(1), 56–71, doi: 10.5771/0943-7444-2020-1-56.*

Contact to the Author:
bm.wlodarczyk@uw.edu.pl
Department of Bibliography and Documentation,
Faculty of Journalism, Information and Book Studies,
University of Warsaw,
Nowy Świat 69,
00-046 Warsaw, Poland