Analyzing Persian Wikipedia's citations to discover the effectiveness of Persian scientific papers: applied web mining techniques

Forough Rahimi and Farshid Danesh
Department of Information Management, Regional Information Center for Science and Technology, Shiraz, Islamic Republic of Iran

Received 1 July 2022
Revised 19 March 2023
15 May 2023
10 June 2023
Accepted 28 June 2023

Abstract

Purpose – The main objective of this study is to evaluate the impact of Persian Scientific Papers (PSPs) on Persian Wikipedia by studying Wikipedia's citations to these papers.
Design/methodology/approach – The present study is applied research, which has been performed by the web-mining method, such as downloading web pages, extracting information (references), identifying papers, detecting peer-review journals and calculating the frequency rates. The statistical population included 10,000 Persian Wikipedia Pages (PWPs) that were analyzed in two rounds with a six-month interval.
Findings – The number of pages containing the Persian references section was 3,994 and 4,063 out of the 10,000 pages extracted in the first and second rounds. The ratio of pages that cited scientific sources (58 and 67 pages) to the pages extracted from the PWP was equal to 0.58 and 0.67%. The ratio of pages that cited scientific sources to pages with Persian references in each round was equal to 1.45 and 1.64%. The number of references extracted from the PWP in each round equaled 30,441 and 35,891. Eight titles from reputable Persian journals had received at least three citations from Wikipedia.
Originality/value – The present study has determined the extent of interaction between science and society (knowledge flow) in the form of citations from Wikipedia articles to articles in peer-reviewed journals. The study of this issue in Persian Wikipedia in more than 2000 Persian peer-reviewed journals shows the originality of the present paper. Studying citation reliability in a collaborative and openly editable platform is another originality of the work.

Keywords Wikipedia, Web-mining, Citation analysis, Persian journals, Societal impact, Web citation
Paper type Research paper

1. Introduction

Wikipedia is an extensive encyclopedia created and launched with the voluntary participation of individuals in more than 329 languages of the world, of which 318 are active and have more than 57 million papers (Wikipedia List [1], 2022). This encyclopedia has become one of the largest and most popular free reference websites without business advertisements since its launching of Wikipedia in 2001. ALexa ranking [2] has rated Wikipedia among the top 15 websites by 2021 (Wikipedia [3], 2021). Wikipedia was provided to the world by Jimmy Wales and Larry Sanger with the goal set to create and publish a free encyclopedia worldwide in all living languages of the world, which the Wikimedia Foundation [4] currently administrates. The International Federation of Library Associations

This article is extracted from a research project entitled "Impact Assessment of Persian scientific papers citations on the web: A citation analysis of Persian Wikipedia references." The article's authors acknowledge the chancellery of ISC for approving and supporting the current research project.

Performance Measurement and Metrics
Vol. 24 No. 2, 2023
pp. 85-100

1467-8047
DOI 10.1108/PMM-07-2022-0021

and Institutions (IFLA) Wiki data Working Group was also founded in late 2019 by Library and Information Science (LIS) experts aimed at discovering and extracting data.

With the knowledge that Wikipedia pages are good starting points for meeting users' information needs, search engines have contributed to the increasing visibility of Wikipedia on the web by indexing the pages of this encyclopedia (De la Calzada and Dekhtyar, 2010).

The growing popularity of Wikipedia is on the rise among the general public and academic communities (Lim, 2009). However, Wikipedia's reputation has been constantly debated despite its popularity. On the one hand, some have praised it for its free publication, free editing features, impartiality policy and wide range of titles. On the other hand, critics have questioned Wikipedia's reputability and credibility due to some factors, including the freedom of editing (anyone can edit), prone to be damaged by vandals, spammers and pranksters, the consensus preference over credibility in editing papers (disagreements over content changes are usually resolved by consensus). Despite the proponents (Huggett, 2012; Murley, 2008) and opponents (Miller and Murray, 2010; Cohen, 2007) of using Wikipedia as a credible source of information in the field of research, Wikipedia is still utilized and cited by researchers. Researchers have concluded their studies that the sabotage and entry of misinformation generally occur transiently (Viégas *et al.*, 2004). Besides, Wikipedia seems relatively as accurate as other encyclopedias, including Britannica (Greenstein and Zhu, 2018; Rosenzweig, 2006; Giles, 2005), which has put it on the list of reference sources in libraries. Interestingly, the findings have revealed that despite the annual rise in citations to other free and online encyclopedias in Scopus-indexed papers, the number of research papers citing Wikipedia appears higher than in other encyclopedias (Huggett, 2012; Park, 2011).

Due to the emergence of the social web and the feature of recording all activities and any use of information resources in the web environment, one question comes to mind: Are the PSPs studied by Persian Wiki writers and PWP users? The answer to this question may reflect the impact of reputable PSPs on users and wiki writers. Regarding English papers, various Altmetric service providers, including Altmetric, Crossref and PlumX, recognize and monitor Wikipedia citations to articles as Altmetric metrics (Ortega, 2018). However, there is no available ready-made tool for reputable PSPs.

Although the majority of the references found in Wikipedia are related to non-scientific sources such as web pages, reports, news and blogs (Serrano-López *et al.*, 2017), the participants in generating English Wikipedia content also use scientific papers (Priem, 2014; Kousha and Thelwall, 2017b). For instance, about 6% of the papers in Public Library of Science (PLOS) journals have received a link from Wikipedia (Lin and Fenner, 2014). The impact of scientific papers on web content, including Wikipedia, is reflected in its "References," suggesting the influence of valid and reputable PSPs on the web in various areas, some of which can be mentioned as follows: Effects on education and the transfer of knowledge to users (Van Vught and Ziegele, 2011), impact on society through social communication (Holbrook and Frodeman, 2011), creating and strengthening scientific participation, impact on culture by providing insights on how to interact with other communities and cultures and proving information and awareness concerning the values, culture and art of the People of a land (Donovan, 2008). Since the author found no research on the citation rate of reputable PSPs in Wikipedia, this issue was evaluated in this article using a review of the PWPs and their references.

Due to the increased number of Persian peer-reviewed journals to more than 1200 titles, the number of PSPs has also enhanced as the most important channel for knowledge transfer and the latest research achievements. Accordingly, users' daily searches on the web to meet their information needs have also been on the rise. Users often visit web pages and search engines to search for information. This paper examines two topics related to PSPs on the web platform. Firstly, it investigates their popularity rate and secondly, it explores how much they have been cited in the Persian Wikipedia as a platform that contributes to science communication to a global audience.

The impact of the PSPs [5] on Wikipedia can be assessed through the PWPs and the number of references found on these pages, some of which have referred to the PSPs. In this article, Wikipedia is referred to as Persian Wikipedia and scientific sources refer to Persian peer-reviewed journals.

2. Literature review

Advanced searches on the subject of citation to PSPs in the Persian Wikipedia found no literature published in this field. The number of studies investigating the citation of English papers in Wikipedia and the impact of scientific articles on the encyclopedia has been growing in recent years. This trend was revealed by a search of the Web of Science Core Collection (WOSCC), Scopus and Google Scholar. Some of the previous studies are reviewed below. Wikipedia, along with the traditional citation index, may be seen as an excellent indicator of the impact of psychological articles. Banasik-Jemielniak *et al.* (2022) focused on the 1,360 titles of reputable psychology journals to study the impact of these journals on Wikipedia articles. They found that only 357 of 1,350 reputable journals are cited at least once in the entire Wikipedia, not well-cited across psychology-related Wikipedia entries but also well-cited across the whole of Wikipedia.

Tattersall *et al.* (2022) extracted 6,454 citations of White Rose Universities' Research on Wikipedia from 1922 to 2019 from Altmetric.com and Unpaywall Application Programming Interface (API). According to Wikipedia, the highest citations occurred in life, medical and health sciences. In addition, the research found that more than half of the sample was not publicly available on Wikipedia. Davoudi and Noruzi (2022) studied Wikipedia citations to Iranian English-language prestigious journals. The findings indicated that Wikipedia cited 78 scientific journals with 464 citations.

Moreover, 165 Wikipedia citing articles were on English pages. Among the journals cited by the Wikipedia articles, 27 journals in the field of basic sciences, 19 journals in the field of technical and engineering, 17 journals in the field of agriculture and natural resources, nine journals in the field of humanities, four journals in the field of veterinary and only two journals in the field of art and architecture have been cited. Nicholson *et al.* (2021) searched identifiers such as Digital Object Identifier (DOI), PubMed IDentifier (PMID) and PubMed Central IDentifier (PMCID) in 1,923,575 papers available in English Wikipedia. According to the results, 824,298 papers were cited by Wikipedia. These papers have been cited more than 115 million times in the WOSCC.

Singh *et al.* (2021) extracted 29.3 million citations from 6.1 million pages of English Wikipedia. They classified citations as books, journal articles, or Web content. The results showed that 6.7% of Wikipedia articles cite at least one journal article with an associated DOI. Only 2% of all articles which Wikipedia cites were indexed in the WoS. By expanding the coverage of Wikipedia citations equipped with identifiers, distinguishing between academic and nonacademic sources will make Wikipedia's contents better structured and easier to query.

Although citations are at the heart of Wikipedia, there is little information on how users interact with them. For the first time, Piccardi *et al.* (2020) created a tool to analyze the readers' interaction with Wikipedia citations. The findings demonstrated that the overall interaction rate with citations is low: Out of every 300 views of Wikipedia articles, one view results in a click on the references. Studies also revealed that the "clicks" mainly occur on shorter pages and in lower quality, indicating that the user will refer to the "References" section only when Wikipedia does not contain the information the user wants.

Jemielniak *et al.* (2019) extracted 39,561 medical papers from Wikipedia by an API as the study population aimed at retrieving the citation rate for reputable scientific papers. The publication date of the articles was retrieved using the CrossRef technique. Among the medical articles, 11,325 Wikipedia articles had cited in 137,889 journal papers. According to

the results, Wikipedia cites more medical and multidisciplinary journals with a high Impact Factor, representing the strong foundation of medical articles on Wikipedia.

Wikipedia has more than 500 million daily visits. Users visit Wikipedia to meet simple information needs such as checking the statistics, facts, general searches and browsing. Singer *et al.* (2017) found that although their visits are not regular-based, they are more frequent on weekends. Wikipedia can be used to inform about the transfer from scholarly publications to popular and non-peer-reviewed publications, such as Web pages (news, blogs), popular magazines (science/technology) and research reports. Serrano-López *et al.* (2017) used web-mining techniques to extract many Wikipedia references. According to the results, less than 1% of articles (0.62%) related to wind energy were cited on Wikipedia, which revealed that the direct societal impact through Wikipedia is minimal for Wind Power research. The study by Teplitskiy *et al.*, 2017 showed that the impact factor of the journal and open access increases the likelihood of citing the journal on Wikipedia. Open-access policies significantly enhance the dissemination of science through intermediaries such as Wikipedia and its role in a global information economy.

Chen and Roth (2012) introduced a protocol for characterizing the referencing process in general article editing by extracting a sample of articles from the entire English Wikipedia. An article's reference density (number of references per unit length) measures the degree to which external sources are used to support and substantiate the article's content. The findings suggest that support and substantiation of articles by referencing external sources only occur after articles have reached a certain level of maturity.

Since Wikipedia is a public online encyclopedia liked by users, Thelwall (2016) extracted citations to astronomy and astrophysics journals from Wikipedia in his research using the Bing search engine and the webometrics method. The results showed that articles published before 2008 were less cited; this is true in general and in most languages of this encyclopedia. Since no research was found that has specifically reviewed and evaluated the citation effect of the PSPs in the web environment through the Persian Wikipedia References, the current topic was considered a research subject. Therefore, it was attempted to analyze the Wikipedia references using web mining techniques and clarify the rate of Persian Wikipedia citations to the PSPs.

The literature review indicated that the researchers have studied and analyzed a subject area, for example, Banasik-Jemielniak *et al.* (2022), the subject of psychology and Jemielniak *et al.* (2019) have analyzed medical articles cited in Wikipedia. Also, most research has focused on citations and articles in the English language of Wikipedia. In addition, in previous research, attention has been paid to the received citations of articles indexed in Scopus and WOSCC and their comparison with Wikipedia citations. For example, we can refer to the research of Singh *et al.* (2021) and Nicholson *et al.* (2021).

3. Main purpose and research questions

This paper aimed to examine the impact of PSPs on Persian Wikipedia by reviewing Wikipedia references. Otherwise, studying Wikipedia references was tied to evaluating scientific sources' citation rate and frequency, including PSPs. The following questions were answered to realize this objection:

RQ1. What is the ratio of pages with Persian Scientific References to all PSPs and all pages with Persian references?

RQ2. What is the ratio of Persian Scientific References to the total References extracted from the PWPs and the total Persian References?

RQ3. Which of the main categories of Wikipedia have more citations been given to Persian peer-reviewed Journals?

RQ4. Which Persian peer-reviewed journals have had the most citation effect on the Persian Wikipedia?

4. Methodology

The initial idea to design the method of this article, from (Serrano-López *et al.*, 2017). The article mentioned that web mining techniques and Python had been used to extract Wikipedia references.

4.1 Data collection

The present study applied research performed by the web-mining method with an analytical approach. Following the extraction of web pages, the pages were analyzed and mapped one by one between the PWRs and the titles of reputable Persian peer-reviewed journals. The statistical population consisted of two ten thousand pages of PWPs. Every ten thousand pages were extracted in two time periods with an interval of six months from each other. Due to its variable nature, six months were selected for the data retrieval according to the changes in Wikipedia pages.

Moreover, the Persian Python Library was utilized to extract the PWPs, which allows searching and receiving articles, links and images and is generally designed for simple applications. Then, ten thousand PWPs were extracted from Persian Wikipedia in November 2020 using the Wikipedia library based on web-mining techniques and script-writing in Python. The titles of the same pages were extracted again from Wikipedia in April 2021 to compare the data changes over time.

4.2 Data analysis

After extracting the pages, the steps below were followed to identify the references contained in the Wikipedia pages: Removing the texts and images in the PWPs; Eliminating the pages without references and thereby identifying pages with references; Extracting the references of pages and completing the references with the help of footnotes on each page; Deleting pages without Persian references and Separating scientific references from other sources with different identifiers such as International Standard Serial Number (ISSN), DOI, Quarterly, Monthly and Journal.

4.3 Data mapping

For mapping the Wikipedia references with the titles of reputable journals, the following steps were performed in the next stage:

(1) Accelerating the references mapping process with journals by processing the references and standardizing them using the following steps: reducing the search scope; eliminating the author's name; converting dashes to spaces; converting two spaces to one space and converting half-spaces to spaces.

(2) The "Exact Match" approach was used to do mapping between references and the journals' titles. There were some limitations in this stage, such as the summarization made in Wikipedia's references, not-mentioning the full name of the journals and the one-word name of some journals. A similarity-finding approach was applied to determine the similarity between the references and journal titles aimed at overcoming these limitations.

(3) The Jaccard index was used to determine the similarity between the scientific references of the Persian Wikipedia and the titles of peer-reviewed journals because of its higher accuracy in this study and its frequent usage in text similarity-finding studies with similarity rates of 70% (Jain *et al.*, 2017; Niwattanakul *et al.*, 2013).

(4) Finally, the pages that used reputable PSPs in their references were identified. Since each page of Wikipedia falls into several categories, to have stability in the conducting method to repeat it, the first category related to the paper based on the thematic relevance was considered to determine the category of the extracted PWPs. It should be noted that there is no uniformity in the PWPs in terms of mentioning the categories. For example, some articles have a section similar to the image indicated in Figure 1, which can be viewed, discussed and edited, in which the subcategories related to the article are mentioned (Figure 1). However, some other PWPs do not have such categories (see Figure 2).

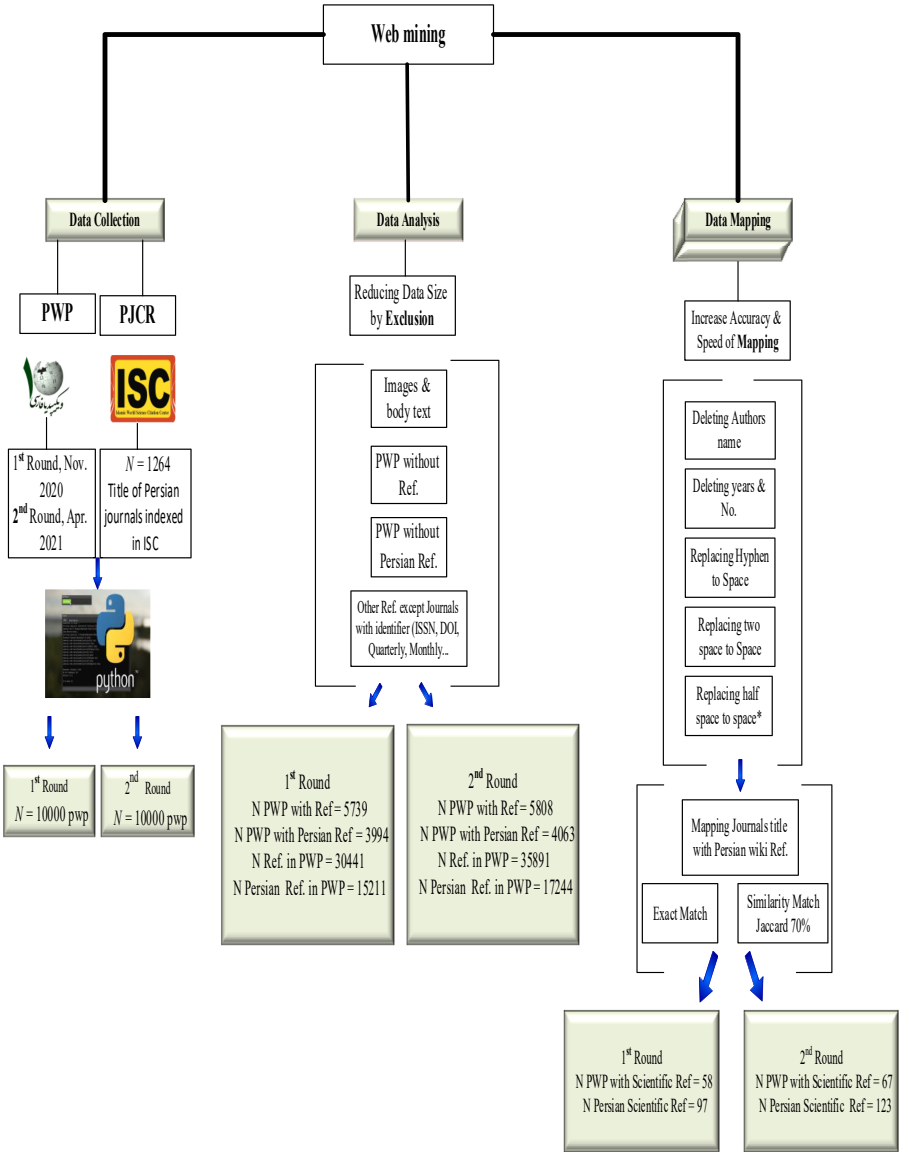
Microsoft Excel 2019 (especially function writing), Statistical Package for the Social Sciences (SPSS) (version 24.0) and the one-dimensional chi-square test was applied for data analysis. The process was time-consuming, helping researchers to confront ambiguities in some cases. In order to resolve these ambiguities, the researchers sometimes needed to refer to the mentioned pages in Wikipedia. The Islamic Science Citation (ISC) database was also employed to determine the subject area of Persian journals and their impact factors.

Islamic World Science and Technology Monitoring and Citation Institute (ISC) provides access to targeted content and power tools to search, track, measure and collaborate in science, social science, art and humanities. The multidisciplinary research platform lets users search numerous databases via one interface, such as Journal Citation Report (JCR). The JCR aims to review, evaluate and rank reputable journals. The quantitative method is based on the authors' use of journals and, more importantly, journals' use of other journal articles. This website is available in Persian, Arabic and English at https://jcr.isc.ac/.

By referring to the above link and searching for the title of the journal, the impact factor, the quartile and the subject area of that journal were retrieved.

Figure 1.
Categories section in one of the Persian articles on Wikipedia

Source(s): Figure courtesy of Wikipedia (2021)



Source(s): Figure by authors

Figure 2.
An illustration of the research steps

5. Findings

5.1 Calculating and comparing the ratio of pages with PSRs to the total PWPs and pages with Persian references

Examining the PWPs regarding possessing the references section demonstrated that 5,731 pages (57.31%) out of 10,000 extracted pages had the references section in the first round. Moreover, 3,994 pages (39.94%) had Persian references; however, only 58 pages cited

scientific references. Over half of the pages (6006) lacked a references section or Persian references (Table 1). Therefore, the ratio of pages with PSRs to the total PWPs was 0.58%, and the ratio of pages with PSRs to the total pages with Persian references seemed equivalent to 1.45%.

In the second round, 5,808 pages (58%) of the ten thousand PWPs had a references section. Furthermore, 4,063 pages (40.63%) had Persian references; however, only 67 pages cited scientific references. Therefore, more than half (5,937 pages) did not have a references section or did not include Persian references (Table 1). Therefore, the ratio of pages with PSRs to the total PWPs was 0.67%, and the ratio of pages with PSRs to the total pages with Persian references seemed equivalent to 1.64%.

According to Table 1, compared to 58 PWPs in the first round, 67 papers in the second round had cited the country's reputable journals. Among these 68 articles, nine new papers had been retrieved, suggesting a trivial difference between the ratio of pages with PSRs to the total PWPs in the first and second rounds of data extraction (0.58 and 0.67).

5.2 Calculating and comparing the ratio of PSRs to the total PSRs extracted from the PWPs and the total Persian references

In the following, the total PSRs extracted from the pages with references section were counted in the first round. Out of 5,431 pages with a references section, 30,441 references were extracted, and among them, 24,915 references were extracted from pages with at least one Persian reference. Among the total references, 15,211 references (equal to 50%) were allocated to the Persian references, and only 97 PSRs were identified in this round. Therefore, the ratio of the PSRs to the total PWRs was equal to 0.31%, and the ratio of the number of the PSRs to the total Persian references was equal to 0.63% (Table 2). Six out of every 1000 Persian references used on Wikipedia pages had cited the scientific references.

The number of references extracted from the pages with references section in the second round was also counted. Among 5,808 pages with the references section, 35,891 references were extracted. Among this number, 28,959 references were extracted from pages with at least one Persian reference. Moreover, 17,244 references (equal to 48%) were allocated to the Persian references. Moreover, 123 PSRs were identified in the second round. Accordingly, the ratio of the PSRs to the total PWRs was equal to 0.34%, and the ratio of the number of the PSRs to the total Persian references was equal to 0.71% (Table 2). Seven out of every 1,000 Persian references used on Wikipedia pages had cited the scientific references.

5.3 Determining the main categories of Wikipedia with the most citations to peer-reviewed journals

The subject categories of Wikipedia and its subcategories were first identified to answer the second question of the present research. Wikipedia contains 41 known general categories. Each article in Wikipedia may fall into several categories as well. In addition, the categories in

Table 1.
Number of reviewed pages on Wikipedia in the first and second rounds

		Total number of studied pages in wikipedia	Number of pages with references (Persian or non-Persian)	Number of pages with at least one Persian ref	Number of pages with scientific references
1st rounds	Freq	10,000	5,731	3,994	58
	%	100	57	39.94	0.58
2nd rounds	Freq	10,000	5,808	4,063	67
	%	100	58	40.63	0.67

Source(s): Table 1 by authors

the PWP are not clearly determined. Therefore, the relevant pages on Wikipedia were searched to specify the category of each of the 58 and 67 pages of Wikipedia that had cited the PSRs. Table 3 indicates the pages' titles extracted in the first and second rounds that cited the PSPs.

The category of each page was identified based on the categories contained in Wikipedia. Table 4 lists the categories that have given the most citations to the PSPs.

		Total number of studied ref. In wikipedia	Number of Persian and non-Persian references on pages with at least one Persian ref	Number of Persian references on pages with at least one Persian ref	Number of PSRs
1st round	Freq	30,441	24,915	15,211	97
	%	100	81.84	50	0.63
2nd round	Freq	35,891	28,959	17,244	123
	%	100	80.68	48	0.34

Source(s): Table 2 by authors

Table 2. The number of studied references in Wikipedia

No	Article's title on Wikipedia	No.	Article's title on Wikipedia
1	Ibn Sina**	30	*Syria*
2	Apostasy in Islam	31	*Seyyed Mohammad Khatami*
3	Khuzestan Province	32	*Simin Daneshvar*
4	Isfahan	33	*Shahnameh*
5	Sugar Addiction	34	*Rotational Shaping*
6	Iran	35	*City of Talesh*
7	Iraj Mirza	36	*Citizenship*
8	Abadan	37	*The proposed plans for administrative division in Iran*
9	Arman Hosepian	38	*Abbas Mirza*
10	Drug liberalization	39	*Ayyaran*
11	Manichaeism	40	*Sexual Fetishism*
12	Blaise Pascal	41	*Rasht Municipality Palace*
13	Bandar Lengeh	42	*Cameroon*
14	Birjand	43	*Kamyaran*
15	Cultural Development	44	*Book of Ahmad*
16	Wealth	45	*Kelardasht*
17	Republic of Azerbaijan	46	*Cambodia II*
18	Jahrom	47	*The 28 Mordad coup d'état*
19	Sefidab Spring	48	*Quality of life*
20	Hamid Reza Sheshjavani	49	*Georgia*
21	Heiran Alishah	50	*Lahijan*
22	God	51	*Majāles-e Sab'a (Seven Sessions)*
23	Damavand	52	*Mashhad*
24	Herat quarter	53	*Meymand (Kerman)*
25	International Regimes	54	*Naser al-Din Shah*
26	Research gate	55	*Najib al-Din Jorfadeghani*
27	Persian language	56	*Exchange rate*
28	Saadi	57	*Neyshabur*
29	Sanandaj	58	*Varamin*

Nine new titles were retrieved in the second round of extracting pages that had cited the PSRs: Ecuador, Algeria, Ahvaz, Bagh-e Malek, Bushido, Paraguay, Tehran, Geography of Iran, Saqez

Note(s): **Pages belonging to the category of names are displayed in italic

Source(s): Table 3 by authors

Table 3. Titles of Wikipedia pages with citations to the PSPs

As shown in Table 4, the number of pages related to the names of places and people, i.e., the two categories of geography and people, have the highest number of pages with scientific references on Wikipedia. Moreover, the Humanities category is in third place.

We examined ten thousand Wikipedia articles to determine whether they were geographical (places) or human articles. According to the results, 29% of the articles were related to the names of famous people, characters and figures, while 39% belonged to the names of places and the geography category. This means that only 32% of the papers belong to other categories.

5.4 Determining the main categories of wikipedia with the most citations to peer-reviewed journals

The findings concerning the third research question demonstrated that 67 and 76 journals had been cited on Wikipedia pages in the first and second rounds, respectively. The table below lists journals cited at least three times on Wikipedia. Among them, the journals of 1-*Amirkabir Journal of Civil Engineering*, 2-*Heritage of Mirror* (Ayeeneh-ye Miras), 3- *Ganjineye Asnad*, 4- *Macroeconomic Research Letter*, 5-*Journal of History of Iran*, 6- *Applied Researches in Geographical Sciences*, 7- *Population* and 8-*Rahbord-e-Tousee* in the first round, have been cited more than other journals and at least three times in Wikipedia. In the second round, the journals of 1- *History of Iran*, 2- *Historical Studies*, 3- *Population*, 4- *Ganjineye Asnad*, 5-*Amirkabir Journal of Civil Engineering*, 6- *Applied Researches in Geographical Sciences*, 7-*Heritage of Mirror (Ayeeneh-ye Miras)*, 8- *Macroeconomic Research Letter* have been cited more than other journals and cited between 7 and 3 times in Wikipedia (Table 5).

According to the results, 50 and 18 titles of journals have been cited in the first round once and twice or more, respectively. In the second round, 55 titles and 21 titles were cited once and twice or more, respectively. We added nine new journal titles to Wikipedia's cited journal list in the second round.

6. Conclusion and discussion

Approximately 1.5% (1.45 and 1.64%) of pages with PSRs in both rounds have cited the PSRs, whilst the ratio of pages containing PSRs in both rounds to the total pages is almost equal to 0.5%. Moreover, between 6 and 7 out of every 1000 PSRs used on Wikipedia pages had cited the PSRs. This finding is somewhat consistent with the results of Serrano-López *et al.*, 2017, suggesting that a low percentage of the Wikipedia references have cited the articles related to the "Wind Power" research. They concluded that despite the trivial social impact of "Wind and Wind Power" articles on Wikipedia, Wikipedia can still be used as a tool for informing the transfer of information from peer-reviews journals to public and non-scientific publications such as web pages (news, blogs). This finding is not in line with the

Table 4.
Frequency distribution of the subject category of Wikipedia pages

Main subject	Freq	Main subject	Freq
Geography	23	Government	2
Persons	14	Education	1
Humanities	3	Awareness and knowledge	1
Philosophy	2	Language	1
Policy	2	Human Behaviors	1
Economy	2	Industry	1
History	2	Science	1
Society	1		

Source(s): Table 4 by authors

Row	Cited journal by wikipedia, round 1	URL	No. Citation, round 1	Cited journal by wikipedia, round 2	No. Citation, round 2
1	مهندسی عمران امیرکبیر Amirkabir Journal of Civil Engineering	https://ceej.aut.ac.ir/?lang=en	5	تاریخ ایران Journal of History of Iran	7
2	آینه میراث Mirror of heritage	http://www.am-journal.ir/?lang=en	4	جستارهای تاریخی Historical Studies	7
3	گنجینه اسناد Ganjine-Ye-Asnad	http://ganjineh.nlai.ir/?lang=en	4	جمعیت Population Journal	6
4	پژوهشنامه اقتصاد کلان Macroeconomic Research Letter	http://jes.journals.umz.ac.ir/?lang=en	3	گنجینه اسناد Ganjine-Ye-Asnad	6
5	تاریخ ایران Journal of History of Iran	https://irhj.sbu.ac.ir/	3	Amirkabir Journal of Civil Engineering مهندسی عمران امیرکبیر	5
6	تحقیقات کاربردی علوم جغرافیایی Journal of Applied researches in Geographical Sciences	https://jgs.khu.ac.ir/index.php?slc_lang=en&sid=1	3	تحقیقات کاربردی علوم جغرافیایی Journal of Applied researches in Geographical Sciences	3
7	جمعیت Population Journal	https://populationmag.ir/en	3	آینه میراث Mirror of Heritage	3
8	راهبرد توسعه Rahbord-e-Tousee	http://rahbord-mag.ir/en/Page/Info	3	پژوهشنامه اقتصاد کلان Macroeconomic Research Letter	3

Note(s): *Since the peer-reviewed journals were Persian, the Persian and English names of the journals and their "URLs" are given in the table above
Source(s): Table 5 by authors

Table 5.
Names of journals cited in Wikipedia in the first and second rounds*

results of Kousha and Thelwall (2017b), indicating that 5% of the reviewed papers extracted from Scopus have a citation in Wikipedia that suggests the increased level of the social impact of research, especially in the field of environmental sciences.

The results of the study conducted by Barbic *et al.* (2016) also revealed that the papers in the field of emergency medicine had received the highest rate of social citations on Wikipedia and Facebook after Twitter. The knowledge of emergency medicine has found its way to the community through social media, including Wikipedia, an issue that has not been observed in the field of Persian articles to this extent.

The findings presented in studies related to the increased motivation of Wiki writers have clarified those social rewards far greater than material rewards can effectively enhance the Wiki writers' incentive to improve Wikipedia. Moreover, receiving feedback on collaborating with and promoting Persian Wikipedia is the strongest motivator for improving Wikipedia (Asadi *et al.*, 2013). Therefore, granting special awards for active and accurate Wiki writers by the Academy of Persian Language and Literature and associations active in the field of Persian language and literature protection and institutions related to the monitoring and evaluation of science can be helpful in this context. These efforts have made the rapid growth of PWPs and upgraded the Persian Wikipedia rank, among others, in recent years. PWPs are

built upon published knowledge that helps remove barriers to access to research for global audiences beyond academia.

On the other hand, such rewards could be considered an indicator to improve the quality, comprehensiveness and sustainability of Persian Wikipedia. Creating the world's largest reference work in Persian requires the cooperation of society and the university and constructive interaction between them. Communicating with society through Wikipedia can enhance users' awareness, including school and university students (especially undergraduates), housewives, teachers and other readers. Moreover, translating sources, discovering vague facts and taking licensed and free images are among the tangible and specific issues that have met the information needs of users of Wikipedia. In other words, Wikipedia serves as a public library and seems a window into knowledge.

The analysis of PWP in terms of the number of PSRs in them in the first and second rounds indicated that more than half of the reviewed pages in Wikipedia either have no references section or lack Persian references. In other words, the share of pages with Persian references is less than non-Persian references in the Persian Wikipedia. Since English is the language of science (Ammon, 2001) and the majority of the information on the web is in English, it can be claimed that it would be easier and simpler to cite non-Persian references. Besides, studies have indicated that the distribution of participation of different countries in editing Persian Wikipedia is such that 30% of the participants in this encyclopedia are from Europe and 25% from the United States (Yasseri *et al.*, 2012).

The findings revealed that the two categories of geography and individuals have the highest number of citations to PSRs. A glance at the 10,000 papers extracted from Wikipedia also confirms that most of the information in Persian Wikipedia relates to places (geography) and people (individuals). The results of this section of the present article seem consistent with the results of Kittur *et al.* (2009). The names of individuals, artists, athletes and politicians accounted for nearly 30% of their research data. A significant issue concerning the categories in the PWP is to pay attention to thematic confusion and the fact that an article belongs to more than one subject category, which confuses users and troubles the process of understanding. Therefore, the first category related to the page was considered for the same page to resolve this problem.

Improving the content of Persian Wikipedia and increasing the diversity of its categories beyond the categories of places and people requires the interaction between the community and the university and the support of academics to provide easy-to-understand scientific topics to the public.

The analyses have also demonstrated that Wikipedia has cited less than a hundred titles of Persian peer-reviewed journals. Seventy-five percent of journals have been cited only once, while twenty-five percent have been cited more than twice and up to seven times. The review of journals cited more than three times by Wikipedia suggests that those journals often have a history of more than 15 years of publication and have an Impact Factor (IF) in the ISC. This could also indicate that reputable journals with a high impact factor have succeeded in finding their way into the community and users recognize them more. The evidence suggests that citing these journals on Wikipedia reflects the impact of these journals and their articles on society.

Wikipedia is a reference that sometimes appears more comprehensive and useful than other sources. However, besides the original article, its references need to be reviewed to trust its articles because publishing articles on Wikipedia aims to state facts and disseminate scientific content. In other words, stories that merely contribute to the richness of cultural heritage will have no place in Wikipedia articles (Ayers, 2018).

Wikipedia aims to include recorded facts that are scientifically vetted, not anecdotes or the type of storytelling that gives richness to our cultural heritage — and as a consequence and

because of mirroring the biases of past sources of knowledge, a vast part of the human experience is left out of Wikipedia entirely.

Suppose librarians, researchers and university professors adopt a positive approach to Wikipedia and turn into Wiki writers and a part of the advocating community of open access to high-quality and inclusive knowledge and actively generate and edit Wikipedia articles; it can change the attitudes toward Wikipedia and prevent the entry of factually incorrect information. Writing encyclopedia articles is a perfect example of academic activity. If the publication of articles falls within the scope of expertise of researchers and professors, it brings them reputational points and faculty members can benefit from the points of publishing specialized encyclopedia articles when changing their academic rank. In this regard, if university professors recommend writing, editing and publishing articles on Wikipedia as a class assignment and if the work on Wikipedia is further encouraged by academics and researchers, it would allow the transfer of knowledge flow from the university to the community, bring more reliability in the data of this free encyclopedia, make it possible to learn more and even replace the "Doctor Google" in the future (Jemielniak, 2018).

7. Future research

PWP has increased from 1000 to more than 800,000 papers in 2021, indicating that the number of Persian papers has increased higher than 80,000 folds during the studied years. Moreover, the rank of Persian Wikipedia has been upgraded from 38 to 18 among the Wikipedia pages in different languages, which reveals the welcoming of Wikipedia articles by Persian language Wiki writers and users. Despite the mentioned achievements and successes regarding Persian Wikipedia, the results suggest that less than 1% of scientific findings are used in Persian Wikipedia. Accordingly, some research topics are suggested for the future as follows:

(1) The impact rate of scientific outputs on the community of Persian-speaking users can be somewhat determined by calculating the Wikipedia citations to Persian articles as an alternative measurement. The attention drawn by an article and scientific output among Wiki writers and Wikipedia users can be considered an indicator. The articles that received citations from Wikipedia managed to attract the attention of Wikipedia users and Wiki writers, which can be read as an altmetric index. Examining these articles based on scientometric and bibliometric indicators can be the subject of future research.

(2) In a specific subject area, the journals in that domain can be searched on Wikipedia to determine the particular citation rate to the subject area journals. Wikipedia's citations to journals and articles mean that science has entered the boundaries of society. In order to investigate the impact of different subject areas on society, it is necessary to research to examine various subject areas.

(3) It is suggested that the amount of citations of Wikipedia to Iranian journals indexed in Scopus and Web of Science is studied and their statistical relevance is measured.

(4) Analyzing the citations to other Persian sources and references is possible. Therefore, the types of sources cited by Wikipedia, including papers, books, conference proceedings, websites, news agencies, blogs, newspaper articles and interviews should be specified separately.

(5) It is suggested to research the statistical significance of societal impact articles (articles that have received citations in Wikipedia) with the number of citations received in Scopus and WOSCC databases. One of the expected outputs of this proposal is the validation of Wikipedia citations.

(6) One of the manifestations of Wikipedia's societal impact is people's daily use of Wikipedia articles. Studying summarized and simplified articles will increase public awareness of society. We suggested that librarians of public libraries write wikis voluntarily to promote citizens' general awareness.

Notes

1. https://meta.wikimedia.org/wiki/List_of_Wikipedias
2. www.alexa.com/siteinfo/wikipedia.com
3. https://en.wikipedia.org/wiki/Wikipedia
4. The Wikimedia Foundation is an American non-profit and charitable organization with more than 350 employees generated US$104m revenues with US$81m expenses in 2019, meanwhile raising charitable contributions. The staff and administrators of the Wikimedia Foundation believe in free knowledge and contributing to the realization of this goal. The staff and administrators support values and policies that contribute to the advancement of free knowledge and volunteer communities.
5. Persian Wikipedia started a work, known as "Wikipedia: Free Encyclopedia" in December 2003. Persian Wikipedia's prominent growth rate (80,000%) over the years from 1,000 papers to more than 870,000 articles in March 2022, associated with Wikipedia ranking 11th in terms of number of active users and 18th in terms of number of administrators, reflect the welcoming of Wikipedia articles by the Persian language Wiki writers and users (https://en.Wikipedia.org/wiki/Persian_Wikipedia, 2022).

References

Ammon, U. (2001), *The Dominance of English as a Language of Science*, Mouton de Gruyter, Berlin.

Asadi, S., Ghafghazi, S. and Jamali, H.R. (2013), "Motivating and discouraging factors for Wikipedians: the case study of Persian Wikipedia", *Library Review*, Vol. 62 No. 4, pp. 237-252.

Ayers, P. (2018), "Wikipedia and libraries", in Reagle, J. and Koerner, J. (Eds), *(2020), Wikipedia@ 20: Stories of an Incomplete Revolution*, The MIT Press, pp. 89-106.

Banasik-Jemielniak, N., Jemielniak, D. and Wilamowski, M. (2022), "Psychology and Wikipedia: measuring psychology journals' impact by Wikipedia citations", *Social Science Computer Review*, Vol. 40 No. 3, pp. 756-774.

Barbic, D., Tubman, M., Lam, H. and Barbic, S. (2016), "An analysis of altmetrics in emergency medicine", *Academic Emergency Medicine*, Vol. 23 No. 3, pp. 251-268.

Chen, C.-C. and Roth, C. (2012), "{{Citation needed}} the dynamics of referencing in Wikipedia", *Proceedings of the eighth annual international symposium on wikis and open collaboration*, pp. 1-4.

Cohen, N. (2007), "A history department bans citing Wikipedia as a research source", *New York Times*, Vol. 21 No. 2, available at: http://www.nytimes.com/2007/02/21/education/21wikipedia.html

Davoudi, H. and Noruzi, A. (2022), "Analyzing Wikipedia citations to Iranian English-language journals approved by the Iranian ministry of science, research and technology", *Sciences and Techniques of Information Management*, Vol. 8 No. 1, pp. 15-44.

De la Calzada, G. and Dekhtyar, A. (2010), "On measuring the quality of Wikipedia articles", *Proceedings of the fourth workshop on Information credibility*, pp. 11-18.

Donovan, C. (2008), "The Australian Research Quality Framework: a live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research", *New Directions for Evaluation*, Vol. 2008 No. 118, pp. 47-60.

Giles, J. (2005), "Internet encyclopedias go head-to-head", *Nature*, Vol. 438 No. 15, pp. 900-901.

Greenstein, S. and Zhu, F. (2018), "Do experts or crowd-based models produce more bias? Evidence from Encyclopedia Britannica and Wikipedia", *MIS Quarterly*, Vol. 42 No. 3, pp. 945-959.

Holbrook, J.B. and Frodeman, R. (2011), "Peer review and the ex-ante assessment of societal impacts", *Research Evaluation*, Vol. 20 No. 3, pp. 239-246.

Huggett, S. (2012), "The influence of free encyclopedias on science", *Research Trends*, No. 27, pp. 7-10.

Jain, A., Jain, A., Chauhan, N., Singh, V. and Thakur, N. (2017), "Information retrieval using cosine and Jaccard similarity measures in vector space model", *International Journal of Computer Applications*, Vol. 164 No. 6, pp. 28-30.

Jemielniak, D. (2018), "Wikipedia as A role-playing game, or why some academics do not like wikipedia", in Reagle, J. and Koerner, J. (Eds), *(2018), Wikipedia@ 20: Stories of an Incomplete Revolution*, The MIT Press.

Jemielniak, D., Masukume, G. and Wilamowski, M. (2019), "The most influential medical journals according to Wikipedia: a quantitative analysis", *Journal of Medical Internet Research*, Vol. 21 No. 1, e11429.

Kittur, A., Chi, E.H. and Suh, B. (2009), "What's in Wikipedia? Mapping topics and conflict using socially annotated category structure", *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1509-1512.

Kousha, K. and 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*, Vol. 68 No. 3, pp. 762-779.

Lim, S. (2009), "How and why do college students use Wikipedia?", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2189-2202.

Lin, J. and Fenner, M. (2014), "An analysis of Wikipedia references across PLOS publications", *altmetrics14: Expanding impacts and metrics An ACM Web Science Conference 2014 Workshop*, pp. 23-26.

Miller, J.C. and Murray, H.B. (2010), "Wikipedia in court: when and how citing Wikipedia and other consensus websites is appropriate", *John's Hopkins Law. Review*, Vol. 84 No. 2, pp. 633-656.

Murley, D. (2008), "In defense of Wikipedia", *Law Library Journal*, Vol. 100, p. 593.

Nicholson, J.M., Uppala, A., Sieber, M., Grabitz, P., Mordaunt, M. and Rife, S.C. (2021), "Measuring the quality of scientific references in Wikipedia: an analysis of more than 115M citations to over 800 000 scientific articles", *The FEBS Journal*, Vol. 288 No. 14, pp. 4242-4248.

Niwattanakul, S., Singthongchai, J., Naenudorn, E. and Wanapu, S. (2013), "Using of Jaccard coefficient for keywords similarity", *Proceedings of the International Multiconference of Engineers and Computer Scientists*, Vol. 1 No. 6, pp. 380-384.

Ortega, J.L. (2018), "Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data", *Scientometrics*, Vol. 116 No. 3, pp. 2123-2138.

Park, T.K. (2011), "The visibility of Wikipedia in scholarly publications", *First Monday*, Vol. 16 No. 8, pp. 1-16.

Piccardi, T., Redi, M., Colavizza, G. and West, R. (2020), "Quantifying engagement with citations on wikipedia", *Proceedings of The Web Conference 2020 (*WWW '20). *Association for Computing Machinery*, New York, NY, USA, pp. 2365-2376.

Priem, J. (2014), "Altmetrics", in Cronin, B. and Sugimoto, C.R. (Eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Performance*, MIT Press, Cambridge, MA, USA, pp. 263-288.

Rosenzweig, R. (2006), "Can history be open source? Wikipedia and the future of the past", *The Journal of American History*, Vol. 93 No. 1, pp. 117-146.

Serrano-López, A.E., Ingwersen, P. and Sanz-Casado, E. (2017), "Wind power research in Wikipedia: does Wikipedia demonstrate direct influence of research publications and can it be used as adequate source in research evaluation?", *Scientometrics*, Vol. 112 No. 3, pp. 1471-1488.

Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M. and Leskovec, J. (2017), "Why we read wikipedia", *Proceedings of the 26th International Conference on World Wide Web*, pp. 1591-1600.

Singh, H., West, R. and Colavizza, G. (2021), "Wikipedia citations: a comprehensive data set of citations with identifiers extracted from English Wikipedia", *Quantitative Science Studies*, Vol. 2 No. 1, pp. 1-19.

Tattersall, A., Sheppard, N., Blake, T., O'Neill, K. and Carroll, C. (2022), "Exploring open access coverage of wikipedia-cited research across the white Rose universities", *Insights*, Vol. 35 No. 1, pp. 1-13.

Teplitskiy, M., Lu, G. and Duede, E. (2017), "Amplifying the impact of open access: wikipedia and the diffusion of science", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 9, pp. 2116-2127.

Thelwall, M. (2016), "Does astronomy research become too dated for the public? Wikipedia citations to astronomy and astrophysics journal articles 1996-2014", *El Profesional de la Informacion*, Vol. 25 No. 6, p. 893.

Van Vught, F. and Ziegele, F. (2011), Designing and Testing the Feasibility of a Multidimensional Global University Ranking, Final Report, Consortium for Higher Education and Research Performance Assessment, Brussels.

Viégas, F.B., Wattenberg, M. and Dave, K. (2004), "Studying cooperation and conflict between authors with history flow visualizations", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 575-582.

Yasseri, T., Sumi, R. and Kertész, J. (2012), "Circadian patterns of Wikipedia editorial activity: a demographic analysis", *PloS One*, Vol. 7 No. 1, e30091.

Further reading

List of Wikipedia (2021), available at: https://meta.wikimedia.org/wiki/List_of_Wikipedias (access 18 January 2021).

About the authors

Forough Rahimi got his Ph.D. in Library and Information Science (LIS) from Kharazmi University. She also had BSc and M.A. in Library and Information Science (LIS) from Shiraz University. She is working as an assistant professor in the information management department at ISC. Her research interests are Altmetrics, Text mining, quantitative research in social sciences, and novel scientometrics indicators. Nowadays, she is the director of the information management department.

Farshid Danesh got his Ph.D. in Library and Information Science (LIS) from the Ferdowsi University of Mashhad in 2013. He is working as an assistant professor in the information management department at ISC. His research interests are Scientometrics, Text mining, quantitative research in medical and social sciences, and novel indicators to measure scientists, journals, universities, and countries. Farshid Danesh is the corresponding author and can be contacted at: farshiddanesh@gmail.com