Wikipedia:Wikipedia Signpost/2021-04-25/Recent research

< Wikipedia:Wikipedia Signpost | 2021-04-25

Recent research

Quality of aquatic and anatomical articles

By Tilman Bayer

A monthly overview of recent academic research about Wikipedia and other Wikimedia projects, also published as the Wikimedia Research Newsletter.

How aquatic scientists improve Wikipedia

An article in _Limnology and Oceanography Letters_ reports on the activities of WikiProject Limnology and Oceanography (WP L&O), founded in 2018 by "a group of early career aquatic scientists" concerned about "inadequate representation of aquatic information on Wikipedia":[1]

> "Unfortunately, limnology- and oceanography-related Wikipedia articles are in poor condition. For example, as of 05 May 2020, the English Wikipedia article for Hypolimnion was only six sentences long and contained no references yet had been viewed over 11,000 times per year on average since 2015. [...] Through recruiting new editors and hosting five "edit-a-thons" (focused editing time), WP L&O has added over 50,000 words to limnology- and oceanography-related Wikipedia articles; however, more than 60% of the articles assessed for the project remain in poor condition and lack references, have poor structure, or are missing crucial information."

The open access paper contains the following helpful overview of "ways to promote more equitable dissemination of aquatic scientific information through Wikipedia":

I am a:	And I want to:	I can:	With these tools:
Scientist, educator, student, scientific society, or work institution	Increase quality, accessible information	Edit directly	3, 4, 5, 18, 21, 22, 24, 28
		Recruit new editors	2, 5, 8
		Host edit-a-thons	8, 11
		Translate or promote translation of articles	6, 7, 23
		Incorporate editing into classrooms	1, 31
		Request training or projects involving Wikipedia	1, 29, 30
Scientist, educator, or student	Improve scientific communication skills	Edit directly	3, 4, 5, 18, 21, 22, 24, 28
		Request training or projects involving Wikipedia	1, 29, 30
Scientist, educator, student, scientific society, or work institution	Reduce barriers to editing	Host editing workshop(s)	5, 8
		Apply for funding to increase free information	2
		Encourage dual publication of journal and Wikipedia articles	12, 13, 25
		Require relevant Wikipedia articles be updated when journal article published	25
		Reward contributions	14, 16, 17, 26, 27
Scientist, educator, student, scientific society, or work institution	Assess impact of contributions	Track contributions of individuals or projects	10, 11, 15
		Assess article quality and popularity	11, 15, 19, 20, 21

Logo of WikiProject Limnology and Oceanography

See also an earlier paper involving some of the same authors: "Ripples on the web: Spreading lake information via Wikipedia"

Other recent publications

Other recent publications that could not be covered in time for this issue include the items listed below. Contributions, whether reviewing or summarizing newly published research, <u>are always welcome</u>.

"The Quality and Readability of English Wikipedia Anatomy Articles"

From the abstract:[2]

"Forty anatomy articles were sampled from English Wikipedia and assessed quantitatively and qualitatively. Quantitatively, each article's edit history was analyzed by Wikipedia X-tools, references and media were counted manually, and two readability indices were used to evaluate article readability. This analysis revealed that each article was updated 8.3 ± 6.8 times per month, and referenced with 33.5 ± 24.3 sources, such as journal articles and textbooks. Each article contained on average 14.0 ± 7.6 media items. [...] the articles had low readability and were more appropriate for college students and above. Qualitatively, the sampled articles were evaluated by experts using a modified <u>DISCERN</u> survey. According to the modified DISCERN, 13 articles (32.5%), 24 articles (60%), 3 articles (7.5%), were rated as "good," "moderate," and "poor," respectively. There were positive correlations between the DISCERN score and the number of edits (r = 0.537), number of editors (r = 0.560), and article length (r = 0.536). Strengths reported by the panel included completeness and coverage in 11 articles (27.5%), anatomical details in 10 articles (25%), and clinical details in 5 articles (12.5%). The panel also noted areas which could be improved, such as providing missing information in 28 articles (70%), inaccuracies in 10 articles (25%), and lack or poor use of images in 17 articles (42.5%)."

"Wikipedia Edit-a-thons as Sites of Public Pedagogy"

From the abstract:[3]

"Through the analysis of interviews with 13 edit-a-thon facilitators [...] we find motivations for running edit-a-thons extend far beyond adding content and editors. In this paper, we uncover how a range of personal and institutional values inspire these event facilitators toward fulfilling broader goals including fostering information literacy and establishing community relationships outside of Wikipedia. Along with reporting motivations, values, and goals, we also describe strategies facilitators adopt in their practice. Next, we discuss challenges faced by facilitators as they organize edit-a-thons. [...] Finally, we suggest new ways in which edit-a-thons, as well as similar peer production events and communities, can be understood, studied, and evaluated."

(The paper also cites two earlier reports by the Wikimedia Foundation related to the topic [1] (https://upload.wikimedia.org/wikipedia/commons/8/8b/Wikimedia_Movement_Organizers_Study.pdf) [2] (https://meta.wikimedia.org/wiki/Learning_and_Evaluation/Evaluation_reports/2015/Editathons).)

"How trust in Wikipedia evolves: a survey of students aged 11 to 25"

From the abstract:[4]

"we analyse the answers given by 841 young people [in France], aged eleven to twenty-five, to a questionnaire. To our knowledge, this is the largest study ever published on the topic. It focuses on (1) the perception young people have of Wikipedia; (2) the influence teachers and peers have on the young person's own opinions; and (3) the variation of trends according to the education level. [...] Results. Trust in Wikipedia depends on the type of information seeking tasks and on the education level. There are contrasting social judgments of Wikipedia. Students build a representation of a teacher's expectations on the nature of the sources that they can use and hence the documentary acceptability of Wikipedia. The average trust attributed to Wikipedia for academic tasks could be induced by the tension between the negative academic reputation of the encyclopedia and the mostly positive experience of its credibility."

From the "Results and findings" section:

"When young people consider their past experience on Wikipedia use, they mostly provide positive judgments about the quality of information available on Wikipedia, whatever their education level. The major point of view is that most often the collaborative encyclopaedia enables them to access information they qualify as useful (93%), understandable (92%), and accurate (92.7%). Indeed, in open questions about the

perception of Wikipedia, only one of the 841 respondents reported that he found mistakes in Wikipedia. [...] the higher the education level, the more participants attribute to their teachers a negative image of the collaborative encyclopaedia [...]. While only 31.7% of collège pupils feel that their teachers' opinion about Wikipedia is low (or very bad), 67.2% of lycée pupils, 68.2% of bachelor's students and 76.7% of master's students answer this way."

See also our review of a related earlier paper involving one of the authors: "Wikipedia: An opportunity to rethink the links between sources' credibility, trust, and authority"

"How do academic topics shift across altmetric sources? A case study of the research area of Big Data"

From the paper (preprint version):[5]

"Taking the research area of Big Data as a case study, we propose an approach for exploring how academic topics shift through the interactions among audiences across different altmetric sources. [...] we attempt to investigate the semantic similarity between topics from publications [about Big Data] and those from the discussions of audiences mentioning and disseminating [these] publications across different altmetric sources, including Blogs, News, Policy documents, Wikipedia and Twitter. [... To obtain topics, considering] the short titles of Wikipedia articles, we choose to use the first sentence in the summary which is a condensed explanation of an event, and is equivalent to the titles of blogs, news and policy documents in part. [...] By comparison [with blog and news publications], policy documents and Wikipedia entries have a more limited focus on Big Data publications with fewer publications mentioned. Specifically, the high-frequency terms in these two groups suggest a quite a different concern of topics on these platforms. Wikipedia entries are more oriented towards the research and application of technologies on internet and web, while policy documents have an obvious orientation to more general issues related to social progress like "EU law" and "Climate change". [... Based on a cluster analysis of all terms,] publications from Wikipedia are more oriented towards academic, technical and more theoretical topics (e.g., "university", "cloud computing", or "theory")."

IMGpedia: Enriching Wikimedia Commons images with metadata from DBpedia and Wikidata

The IMGpedia (http://imgpedia.dcc.uchile.cl/) dataset "brings together descriptors of the visual content of 15 million images [from Wikimedia Commons], 450 million visual-similarity relations between those images, links to image metadata from DBpedia Commons, and links to the DBpedia resources associated with individual images [as well as links to Wikidata, in a later version]. It allows people to perform visuo-semantic queries over the images." It is the topic of several academic publications:

A 2017 conference paper[6] and thesis,[7] which - as summarized in the abstract

"defines the process of analysis of 15 million images from Wikimedia Commons in order to build the knowledgebase. First, the visual descriptors must be calculated; later, we propose an efficient strategy to compute similarity links between them; afterwards, we define a method for obtaining relations between the images and DBpedia resources of the Wikipedia articles that use them; and finally, the data is published as an RDF graph ready to be queried through the SPARQL endpoint service mounted for IMGpedia."

A 2018 conference paper titled "Querying Wikimedia Images Using Wikidata Facts"[8] reports on enhancements:

"IMGpedia [...] is a knowledge-base that incorporates similarity relations between the images based on visual descriptors, as well as links to the resources of Wikidata and DBpedia that relate to the image. Using the IMGpedia SPARQL endpoint, it is then possible to perform visuo-semantic queries, combining the semantic facts extracted from the external resources and the similarity relations of the images. This paper presents a new web interface (http://imgpedia.dcc.uchile.cl/query) to browse and explore the dataset of IMGpedia in a more friendly manner, as well as new visuo-semantic queries that can be answered using 6 million recently added links from IMGpedia to Wikidata ..."

The IMGpedia dataset has also been published,[9] but appears to have been removed since.

See also the "Structured data on Commons" project, a separate effort

"Inauthentic Editing: Changing Wikipedia to Win Elections and Influence People"

This is the topic of two blog posts published by the Stanford Internet Observatory earlier this year (on occasion of Wikipedia's 20th birthday)[10][11]:

> "Building on the work of Wikipedia editors catching politically motivated editing [citing a recent *Signpost* report by User:Smallbones], we investigate one case of 'inauthentic editing'—where individuals targeted the Wikipedia pages of two contending politicians during the 2020 British Columbia (BC) general provincial election. Through this deep-dive case study, we also show a process for investigating Wikipedia, and identify Wikipedia's strengths and weaknesses in dealing with inauthentic edits."

The studied example involved conflict-of-interest editing in the article about Canadian politician John Horgan.

In the second post, the student authors provide a detailed overview of how to investigate such cases based on Wikipedia's publicly available data and tools. As summarized (https://twitter.com/MilesMcCain/status/134976348 1092972548) by one of the authors:

> "we explore the key markers of suspicious pages—such as a high edit-to-pageview ratio, or a high number of edits over time from a single otherwise-inactive user. We also include a free program that anyone can use to assess hundreds of pages at a time [called "Wikipedia Scanner" (https://github.com/stanfordio/wikipedia-note book/blob/main/notebook.ipynb) and published in form of a Jupyter notebook for use on Google's Colaboratory platform]."

References

1. Kincaid, Dustin W.; Beck, Whitney S.; Brandt, Jessica E.; Brisbin, Margaret Mars; Farrell, Kaitlin J.; Hondula, Kelly L.; Larson, Erin I.; Shogren, Arial J. (2021). "Wikipedia can help resolve information inequality in the aquatic sciences" (https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lol2.101 68). *Limnology and Oceanography Letters*. **6** (1): 18–23. doi:10.1002/lol2.10168 (https://doi.org/10.1 002%2Flol2.10168). ISSN 2378-2242 (https://search.worldcat.org/issn/2378-2242).

2. Suwannakhan, Athikhun; Casanova-Martínez, Daniel; Yurasakpong, Laphatrada; Montriwat, Punchalee; Meemon, Krai; Limpanuparb, Taweetham. "The Quality and Readability of English Wikipedia Anatomy Articles". *Anatomical Sciences Education*. **0** (0). doi:10.1002/ase.1910 (https://d oi.org/10.1002%2Fase.1910). ISSN 1935-9780 (https://search.worldcat.org/issn/1935-9780). 🔓

3. March, Laura; Dasgupta, Sayamindu (2020-10-14). "Wikipedia Edit-a-thons as Sites of Public Pedagogy". *Proceedings of the ACM on Human-Computer Interaction*. **4** (CSCW2): 100–1–100:26. doi:10.1145/3415171 (https://doi.org/10.1145%2F3415171). 🔓 Deposited version (https://cdr.lib.unc. edu/concern/articles/cr56n873p?locale=en)

4. Sahut, Gilles; Mothe, Josiane (2018-03-15). "How trust in Wikipedia evolves: a survey of students aged 11 to 25" (http://informationr.net/ir/23-1/paper783.html). *Information Research*. **23** (1).

5. Lyu, Xiaozan; Costas, Rodrigo (2020-05-01). "How do academic topics shift across altmetric sources? A case study of the research area of Big Data" (https://doi.org/10.1007/s11192-020-03415- 7). *Scientometrics*. **123** (2): 909–943. arXiv:2003.10508 (https://arxiv.org/abs/2003.10508). doi:10.1007/s11192-020-03415-7 (https://doi.org/10.1007%2Fs11192-020-03415-7). ISSN 1588- 2861 (https://search.worldcat.org/issn/1588-2861). 🔓 Preprint version (https://arxiv.org/abs/2003.105 08)

6. Ferrada, Sebastián; Bustos, Benjamin; Hogan, Aidan (2017). "IMGpedia: A Linked Dataset with Content-Based Analysis of Wikimedia Images". In Claudia d'Amato; Miriam Fernandez; Valentina Tamma; Freddy Lecue; Philippe Cudré-Mauroux; Juan Sequeda; Christoph Lange; Jeff Heflin (eds.). *The Semantic Web – ISWC 2017*. Lecture Notes in Computer Science. Cham: Springer International Publishing. pp. 84–93. doi:10.1007/978-3-319-68204-4_8 (https://doi.org/10.1007%2F978-3-319-68 204-4_8). ISBN 9783319682044. 🔓

7. Ferrada Aliaga, Sebastián Camilo (2017). "IMGPEDIA: A large-scale knowledge-base to perform visuo-semantic queries over Wikimedia Commons images" (http://repositorio.uchile.cl/handle/2250/1 46633). *Repositorio Académico - Universidad de Chile*. (thesis)

8. Ferrada, Sebastián; Bravo, Nicolás; Bustos, Benjamin; Hogan, Aidan (2018). "Querying Wikimedia Images Using Wikidata Facts" (https://doi.org/10.1145/3184558.3191646). *Companion Proceedings of the The Web Conference 2018*. WWW '18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. pp. 1815–1821. doi:10.1145/3184558.3191646 (https://doi.org/10.1145%2F3184558.3191646). ISBN 9781450356404. Author's copy (http://aidanhogan.com/docs/imgpedia-wikidata.pdf) (CC BY 4.0)

9. Ferrada, Sebastián (2017-05-10). "IMGpedia Dataset" (https://figshare.com/articles/dataset/IMGpedia_Dataset/4991099/2). *Figshare*. doi:10.6084/m9.figshare.4991099.v2 (https://doi.org/10.6084%2Fm9.figshare.4991099.v2). Retrieved 2021-04-25. (apparently no longer available as of 2021-04-25)

10. Al Fahim, Maha; Gallagher, Sean; McCain, Miles; Rubin, Nick (2021-01-07). "Inauthentic Editing: Changing Wikipedia to Win Elections and Influence People" (https://cyber.fsi.stanford.edu/io/news/wikipedia-part-one). *Stanford Internet Observatory*. Retrieved 2021-04-25.

11. Al Fahim, Maha; Gallagher, Sean; McCain, Miles; Rubin, Nick (2021-01-14). "How to Investigate Wikipedia: Our Process" (https://cyber.fsi.stanford.edu/io/news/wikipedia-part-two). *Stanford Internet Observatory*.

Supplementary references and notes:

1. https://wikiedu.org/ .
2. https://meta.wikimedia.org/wiki/Grants:Start.
3. https://en.wikipedia.org/wiki/File:WikiProject_L%26O_Quick_Start_Guide_v1.pdf .
4. https://en.wikipedia.org/wiki/Help:Introduction .
5. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Limnology_and_Oceanography .
6. https://en.wikipedia.org/wiki/Wikipedia:Translation .
7. https://www.translatorswithoutborders.org/blog/the-wikipedia-project-update/ .
8. https://en.wikipedia.org/wiki/Wikipedia:How_to_run_an_edit-a-thon .
9. https://tools.wmflabs.org/pageviews .
10. https://xtools.wmflabs.org/ec .
11. https://outreachdashboard.wmflabs.org/ .
12. https://en.wikiversity.org/wiki/WikiJournal_of_Science .
13. http://topicpageswiki.plos.org/wiki/Main_Page .
14. https://blog.wikimedia.org/2011/04/06/tenure-awarded-based-in-part-on-wikipedia-contributions/ .
15. https://xtools.wmflabs.org/ .
16. https://en.wikipedia.org/wiki/Wikipedia:Barnstars .
17. https://en.wikipedia.org/wiki/Template:The_Limnology_and_Oceanography_Barnstar .
18. https://commons.wikimedia.org/wiki/Commons:How_to_take_pictures_for_Wikimedia_Commons .
19. https://www.mediawiki.org/wiki/ORES/FAQ .
20. https://en.wikipedia.org/wiki/Wikipedia:Peer_review/guidelines .
21. https://tools.wmflabs.org/wikipedia-readability/ .
22. http://dispenser.info.tm/~dispenser/view/Altviewer .
23. https://en.wikipedia.org/wiki/Wikipedia:Multilingual_coordination .
24. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000941 .
25. http://wikiambassador.jiscinvolve.org/wp/2014/03/28/publishing-scholarly-wikipedia/ .
26. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003242 .
27. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Molecular_Biology/Computational_Biology/ISCB_competition_announcement_201? .
28. https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Accessibility .

- 29. https://wikiedu.org/blog/category/wikipedia-professional-development/wiki-scientists/ .
- 30. https://www.aps.org/publications/apsnews/202005/wikipedia.cfm .
- 31. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Limnology_and_Oceanography/Workshops#Resources_for_incorpora

---

+ Add a comment (https://en.wikipedia.org/w/index.php?editintro=Wikipedia:Wikipedia_Signpost/Templates/Comment-editnotice&title=Wikipedia_talk:Wikipedia_Signpost/2021-04-25/Recent_research&action=edit&preload=Wikipedia:Signpost/Templates/Signpost-article-comments-end/preload)

THESE COMMENTS ARE AUTOMATICALLY TRANSCLUDED FROM THIS ARTICLE'S TALK PAGE. TO FOLLOW COMMENTS, ADD THE PAGE TO YOUR WATCHLIST (HTTPS://EN.WIKIPEDIA.ORG/W/INDEX.PHP?TITLE=WIKIPEDIA_TALK:WIKIPEDIA_SIGNPOST/2021-04-25/RECENT_RESEARCH&ACTION=WATCH). IF YOUR COMMENT HAS NOT APPEARED HERE, YOU CAN TRY .

- I often coach at edit-a-thons and am perennially annoyed that most have a theme of making new biographies. Twice wrong. First, biographies are already, to my mind, a disproportionate fraction of articles. Second, Wikipedia would be more improved by making existing articles, even biographies, better, than by adding more bad ones that few would read. My guess is, it's done this way because new biographies attract more prospective editors, which reminds me of the old adage that small minds talk about people; mediocre minds talk about events, and great minds talk about ideas. Better if we could find a way to stretch those minds to reach above pettiness and into mediocrity, and maybe even touch greatness occasionally. Pleasant to see that some efforts in a fairly small, poorly covered, and lightly read branch of science have been modestly successful, and I wish this could become a trend. Jim.henderson (talk) 01:15, 26 April 2021 (UTC)

  - So, so true. Partly the emphasis on biographies I think reflects the emphasis on the gender gap, and using editathons to encourage specifically female editors, who are (rather patronizingly perhaps (https://en.wikipedia.org/w/index.php?title=User_talk%3AIridescent&type=revision&diff=1020062709&oldid=1020060098)) assumed to want to write about women. Partly it has now become a habit. But there's no doubt new editors like to see a new article that is "theirs". Johnbod (talk) 14:05, 28 April 2021 (UTC)

    - I think part of the draw of bios (especially BLPs) is that they can feel like a lower level of technical expertise is needed (e.g. students are usually more comfortable writing a BLP than about some scientific topic) and the human angle is itself a draw in and of itself to participants. It's a pity because they're *so* much harder to write than other articles and can give new editors a negative experience. I'm actually in the process of organising a biography-writing editathon with my uni, but I'd like to follow it up with a editathons focused on non-bio articles. T.Shafee(Evo&Evo)^{talk} 09:59, 3 May 2021 (UTC)

- > more than 60% of the articles assessed for the project remain in poor condition and lack references, have poor structure, or are missing crucial information.

  How does this compare to the rest of the encyclopedia? About 58% of assessed, non-list articles are marked as stubs. I'd be curious about the evidence that Limnology/Oceanography coverage suffers any more than any other topic area. **CZAR** 02:28, 26 April 2021 (UTC)

WHAT DO YOU THINK OF *THE SIGNPOST*? SHARE YOUR FEEDBACK.