

A typology of Wikipedia readers

Analyze Wikipedia practices through time¹

Laurent Mell² et Nicolas Jullien³

This survey was supported by the Brittany Region and the PEPR eSEMBLE (CONGRATS targeted project), which is working on the future of digital collaboration.

Introduction

With more than 700 million daily visitors worldwide, Wikipedia occupies an important place in the process of knowledge construction (Jemielniak and Aibar, 2016). Interrogating the place and role of Wikipedia in reading practices becomes eminently interesting, as the value of the information offered is regularly questioned and its reading is relegated to the level of legitimate cultural practices. What does it mean to read Wikipedia? What is the value of the content offered on this platform? Is its use part of a dynamic of 'ordinary' reading practices (Thompson, 2021; Thumala Olave, 2018) and, more broadly, of cultural practices (Bennett et al., 2009; Donnat, 2009), or does it renegotiate modes of access to information as well as the construction of knowledge?

To answer these questions, a survey of Wikipedia users was conducted in the summer of 2023, coordinated by the public research centre Marsouin⁴. Data was collected through an online questionnaire between June and July 2023. The link to the questionnaire was distributed through a banner published in 8 languages on the Wikipedia website. The survey included 200 questions about: what people were doing on Wikipedia before clicking on the link to the questionnaire; how they use Wikipedia as readers (professional and personal uses); their opinion on the quality, coverage and importance of the encyclopaedia; the creation of Wikipedia (how they think it is created, whether they have contributed to it and how); their social, sporting, artistic and cultural activities, online and offline; their socio-economic characteristics, including their political beliefs; and their propensity to trust the encyclopaedia. More than 200,000 people opened the questionnaire, 100,332 started answering it, making

¹ Marsouin project, 2023

² CREAD, IMT Atlantique

³ LEGO, IMT Atlantique

⁴ Find out more about Marsouin at: <https://www.marsouin.org>

up our dataset, and 10,576 completed it. These 10,576 responses make up our sample of respondents⁵.

Data Processing Method

First, it's important to note that our sample is what is called a 'convenience sample'. It tells us nothing about the actual distribution in the population that uses Wikipedia in the eight survey languages. On the other hand, we assume that the respondents represent the full range of profiles and uses of the encyclopaedia. We will therefore analyse the statistical relationships between uses (in this case, contributions) and the sociodemographic, cultural, professional, and other characteristics of the people who use them.

In order to identify factors that differentiate Wikipedia reading practices, we decided to create a typology based on a selection of variables describing these practices: access to Wikipedia; mode of access to content; evaluation of information; quality of article; professional or study (or personal) needs requiring access to written information or knowledge; use of Wikipedia to meet these professional or study (or personal) needs; relationship to Wikipedia sources. The profiles or classes derived from this classification are then interpreted in relation to a set of illustrative variables covering: socio-demographic characteristics; digital competences; digital uses; reading practices; cultural practices; discussions about Wikipedia; external perceptions of Wikipedia; modes of access to media information; perceived reliability of these information sources; and contribution practices.

We have thus identified five classes of Wikipedia readers (see Table 1 below), based on their uses. We have no way of knowing the proportion of each of these classes in the population of Wikipedia readers. For this reason, we do not report the proportion of each class in our sample: it gives no information about the actual distribution. However, as indicated above, we believe that they represent fairly well the different types of readers that can be found in the overall population of Wikipedia readers.

Table 1: A Typology of Wikipedia readers

	Name
Class 01	The Struggling readers
Class 02	The Conformists
Class 03	The Ambivalent
Class 04	The Over-informed
Class 05	The Confidants

The presentation of each class below is divided into three parts. The first, in italics, presents their use of Wikipedia (in other words, the active variables used to construct the different classes). The second describes their socio-economic profile, their digital competences, their

⁵ Survey presentation:

https://meta.wikimedia.org/wiki/Research:Surveying_readers_and_contributors_to_Wikipedia

reading practices and their access to media and information (these variables are not included in the construction of the classes, and the aim is to see if certain modalities of these variables are over-represented in the constructed classes). The third part is devoted to a synthesis of the results presented above and, above all, to our interpretation of the class from an overall perspective.

The Classes

Class 01 (the Struggling readers)

Whether in a personal, professional, or academic context, the Struggling Readers express a moderate need for education, culture, and information. When this need exists, Wikipedia is used little or not at all. This piecemeal use of Wikipedia is also evident when it comes to using Wikipedia texts for documents to be written for work or study.

The Struggling Readers are characterized by a high proportion of older people, with almost half of the members of this class aged over 55. This is the least educated class, and also the one with the highest proportion of lower socio-professional categories. They express the lowest sense of financial comfort of the five classes. The Struggling readers have limited digital competences, resulting in a level of digital fluency well below that of the other four classes, but consistent with their limited use of Wikipedia. The Struggling readers present the image of the 'least informed' class. They are low information consumers, with the lowest levels of print and digital reading, and the majority of non-readers across all content. In addition, they get most of their information from television and search engines, while developing a relatively low sense of the general reliability of information. Their cultural, sporting, artistic and social practices are relatively weak.

Interpretation. Overall, the Struggling readers are readers who identify Wikipedia as just another website and use it sporadically. Their low consumption of information - on Wikipedia and beyond - is linked to low needs (lower than other classes) or limited competences in reading, information retrieval, but also in the use of digital devices. They correspond to the "weak readers" identified in traditional surveys (Bourdieu, 1993; Suarez & Woudhuysen, 2010) and tend to come from working-class backgrounds. Inscribed in an ordinary working-class relationship to knowledge and written form (Lahire, 2015), the Struggling readers are more likely to question the reliability of information on Wikipedia and, more broadly, its legitimacy as an institution for the construction of encyclopaedic knowledge.

Class 02 (the Conformists)

Like the Struggling readers, the Conformists express the need to research out of curiosity, as well as the need to train or educate themselves for professional or academic purposes. But unlike the former, Wikipedia partially satisfies these needs. Personal needs are more important and diverse: to educate oneself, to inform oneself, to learn about politics, and, more moderately, to make a choice and settle a disagreement. But again, Wikipedia is used selectively. Moreover, the Conformists are among the least likely to cite the sources to which they refer on Wikipedia.

The Conformists are characterized by a remarkably high proportion of women compared to other classes. They are also the oldest class, with a significant proportion of retired people. The Conformists have a particularly strong feeling of financial ease, and a significant proportion of members of this class have a little or a lot of money and a significant amount of free time. They are distinguished from the Struggling readers by their relative digital affluence, although they retain fairly 'traditional' digital practices, with a more pronounced consumption of television and radio. They show a selective infatuation with the paper format, concentrating particularly on novels or current affairs. They are regular consumers of information, but use the media selectively to access information. They correspond to one of the two classes with the highest level of trust in the media in general, but the one with the highest level of trust in 'traditional' media (TV, radio, press). In terms of cultural and artistic practices, the Conformists show little interest, but develop other forms of involvement in sports and social activities.

Interpretation. The Conformists identify with a group of active retirees who recognize Wikipedia as an institution of encyclopaedic knowledge. Wikipedia is selectively mobilized to meet personal or professional needs, as a supplement to other sources. Conformists are part of an 'information browsing' dynamic facilitated by the large amount of free time available. Wikipedia is only one source of information among many, yet it is recognized as a legitimate knowledge institution. As such, this class adopts a conformist behaviour of knowledge consumption in an institution that a priori seems more accessible than 'traditional' knowledge and cultural institutions.

Class 03 (the Ambivalents)

The Ambivalents are characterized by a remarkable diversification of their professional or study needs, surpassing those of the Struggling readers and the Conformists. Their uses include education or personal enrichment, the search for diverse information, but also the precise verification of facts or data. Compared to the first two classes, Wikipedia is more suitable for these purposes, although it does not cover all of them. In this context, Wikipedia content is sometimes mobilized and sources are occasionally cited. Personal needs also vary, as in the case of conformists. Again, Wikipedia is used selectively, with less emphasis on making decisions or following political news.

The Ambivalents are characterized by a very young profile, and a significant proportion of members of this class are employed. This class also has the highest proportion of people in education, illustrating a strong commitment to building a professional future or developing a course of study. This dual commitment - to work or to study - is reflected in a limited amount of free time. This class is characterized by a good level of digital literacy, but develops moderate digital uses for work or study as well as for entertainment. When it comes to reading, the Ambivalents, like the Conformists, take a selective approach. While their interest in paper reading is moderate, their interest in digital reading is more developed. The same is true for leisure activities. While cultural and artistic activities remain relatively moderate, sports and social activities are preferred. Like the Struggling readers, they have one of the lowest levels of trust in Wikipedia, but unlike the Struggling readers, Wikipedia partially meets their needs (both professional and personal).

Interpretation. The Ambivalents are characterized by a real need to educate and inform themselves, but take a critical stance toward the information they find on Wikipedia. They seem to have occasional difficulties with the encyclopaedic format. This class's approach to Wikipedia reflects an ambivalent attitude: a pragmatic use of information to accomplish a task (professional, academic, personal), but difficulties in appropriating the content and transforming it into useful knowledge, especially for a class that suffers from limited free time.

Class 04 (the Over-informed)

The Over-informed are characterized by particularly diverse and important needs in both profession-al and academic contexts (education, training, information, but also verification of information or settlement of disputes). In these contexts, Wikipedia can satisfy many of these needs. On a personal level, the needs are also numerous and varied, exceeding those of the Conformists and the Ambivalents. Members of this class use Wikipedia selectively, seeing it as a relevant but not exclusive supplement. This class is illustrated by their strong tendency to cite sources from Wikipedia in professional or academic contexts.

Readers who belong to this class have striking socio-demographic characteristics that reflect a privileged social position. This class is the most highly educated, with the highest proportion of individuals holding a master's or doctoral degree. This distinction is accompanied by a strong representation of higher socio-professional categories (directors, executives, managers, intellectual or scientific professions). The feeling of financial ease is also particularly pronounced. It also has the lowest proportion of women of all the classes studied. However, the Over-informed face a notable constraint: free time. They are among those with the least time for leisure activities. Like the Ambivalents, they have a good level of digital literacy, but they mobilize Wikipedia selectively according to their needs. In this class, the practice of reading on paper is selective. While interest in this medium remains moderate, digital reading is much more developed here, in line with highly diversified digital uses. This is the class that receives the most information across all media. The frequent use of scientific publications is particularly noteworthy.

Interpretation. This class embodies a privileged socio-economic profile, with a high level of qualification, professional success and significant financial affluence, but they have to deal with time constraints. The Over-Informed are characterized by a constant search for information, supported by a variety of digital practices and a strong trust in the media, especially 'traditional' media. This high level of information consumption is underpinned not only by the recognition of Wikipedia as a reference institution, but also by confidence in their ability to select and interpret information, in contrast to the Ambivalent. This pronounced attitude to citing sources could be explained by the high concentration of contributors among the Over-informed, who are aware of the time required to structure knowledge on the platform.

Class 05 (the Confidants)

The Confidants appear to be the class whose personal, professional, and learning needs are the most significant and diverse. These needs cover a wide range of objectives, such as

resolving a disagreement, verifying information, educating oneself, informing oneself, making a choice, or learning about political issues. In this context, Wikipedia occupies a central position. This class is the one that most mobilizes Wikipedia and where Wikipedia is most responsive. Like the Over-informed, the Confidants are quick to cite Wikipedia sources.

This is one of the classes with the highest percentage of women. It is also one of the youngest classes, representing a population in full employment or pursuing studies. With a relatively high level of education, a significant proportion in higher socio-professional categories and a marked sense of financial ease, the Confidants reflect a young, qualified and, for some, economically stable image, but faced with the temporal demands of a rhythm of life that limits their available free time. This class is the most comfortable with digital technology and has the most diverse uses. This class is characterized by a unique relationship to reading and information. Paper reading is selective, reflecting a moderate interest in this traditional medium. On the other hand, they excel at digital reading, making them the class in which this practice is most developed. They keep themselves informed on a regular basis, using multiple channels. Search engines play a central role in this approach, as this class is the one that uses them most to access information. This class is also characterized by a high level of trust in the media in general and in 'digital' media in particular (especially scientific publications, search engines and social networks).

Interpretation. The Confidants embody a networked and curious profile, relying on digital resources and online media in their search for information. This class embodies intensive, Wikipedia-centric use, using its resources intensively to meet multiple and diverse needs.

Conclusion

This typology of Wikipedia readers highlights differences and reveals deeply structured usage dynamics according to individuals' socio-demographic characteristics, cultural practices, and information consumption patterns. The five classes identified illustrate both the diversity of uses of the online encyclopaedia and the differentiated relationships to knowledge and information.

The Struggling readers represent a class that consumes little information, has limited digital and reading practices, and is far removed from regular or in-depth use of Wikipedia. Inscribed in the ordinary working-class relationship to legitimate knowledge and written form, they struggle to appropriate encyclopaedic knowledge. At the other end of the spectrum, the Confidants embody a young, networked class that places Wikipedia at the centre of their personal and professional use of information. This group, which makes extensive use of the encyclopaedia to meet their needs, reflects an assertive and confident relationship with digital technology, digital reading, and online media. In between these two profiles, the Conformists, the Ambivalents, and the Over-informed adopt heterogeneous uses of Wikipedia that reflect their specific positions. The Conformists use Wikipedia selectively, as a complement to other traditional media, which they value more. The Ambivalent, while distrustful of Wikipedia and sometimes struggling to appropriate its content, pragmatically integrate it into their information processes. Finally, the Over-informed, with their diverse digital practices and constant appetite for information, recognize Wikipedia as an essential knowledge institution.

More generally, this typology raises questions about the way Wikipedia structures its content and presents information, given the diversity of ways in which readers of the platform appropriate it. These initial findings show that the process of social differentiation of reading practices also operates on the platform, and that access to information on Wikipedia refers to 'ordinary' reading practices. However, this typology of Wikipedia readers also highlights specific modes of information appropriation that deviate from social representations of reading. These preliminary findings pave the way for thinking about the connections between information needs, digital literacy, and differentiated relations to knowledge. Long criticized and marginalized by the proponents of legitimate knowledge (schools, academia, information professionals, political actors, etc.), Wikipedia is gradually tending to adopt their values. But doesn't following in their footsteps limit access to knowledge for those who have the most difficulty with the written word and literate culture?

References

- Bennett, T., Savage, M., Silva, E., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. Routledge.
- Bourdieu, P. (1993). *The Field of Cultural Production*. Columbia University Press.
- Donnat, O. (2009). French Cultural Participation in the Digital Age Summary 1997-2008. *Culture études*, No 5(5), 1-12.
- Jemielniak, D., & Aibar, E. (2016). 'Bridging the gap between Wikipedia and academia '. *Journal of the Association for Information Science and Technology*, 67(7), 1773-1776.
- Lahire, B. (2015). Cultural dissonances: the social in the singular. Laurie Hanquinet; Mike Savage. *Handbook of the sociology of art and culture*, Routledge, p. 312-323.
- Lombardo, F., Daly, M. (2021). 'Analyzing suicide life stories on Wikipedia with Highway_star and other textual visualization tools'. *SN Soc Sci* 1, 268.
- Suarez, M. F. & Woudhuysen H.R. (2010), *The Book: A Global History*. Oxford university press.
- Thumala Olave, M. A. (2018). 'Reading Matters: Towards a Cultural Sociology of Reading'. *American Journal of Cultural Sociology*, 6(3), 417-454.
- Thompson, J. B. (2021). *Book Wars: The Digital Revolution in Publishing*. Cambridge: Polity Press.