Vision (WikiGaze)

October 30, 2019

The success of Wikipedia and the relative high quality of its articles seem to contradict conventional wisdom. Recent studies have begun shedding light on the processes contributing to Wikipedias success, highlighting the role of editors and moderators. Very few studies focus on the importance of other end of spectrum i.e readers of Wikipedia articles. [1] The users of Wikipedia can be divided into two sets; the production team members(editors, moderators) and the passive team members(readers). Most of the researches performed to analyze various characteristics of Wikipedia revolve around understanding the behavior of the production teams, i.e. editors, moderators and their collaboration dynamics. A study [2] performed on 477 Wikipedia based researches revealed that 42% of these reseachers were centered on understanding characteristics of contributors involment or article quality evaluation [3, 4, 5]. Only 20% of the studies related to readers in Wikipedia and the usage of Wikipedia. Less than 1% of the reviewed studies looked at users' reading preferences.

For a particular article, multiple users generate their personalized summaries. We ask the users to share their summaries for analyzes purpose. Due to privacy reasons, some users might not be willing to share their data. As an incentive mechanism we implement a special feature to generate "recommended summary". To unlock this feature, a user must share their data with the central analyzes team. Each user is given a special ID to maintain anonymity. We develop a cross-platform standalone application to generate personalized summary and to recommend summaries based on user's past reads.

1 Relation between Article Readability and Summary

Eye movement patterns characteristic of reading, extensively studied in the psychology literature [6], open the door for an automated analysis of web browser reading.

While readers intuitively may think that their eye gaze follows a continuous left-to-right motion, eye tracking studies show that eye motion advances in discrete chunks across the page. A reader's eyes will actually stop, or fixate, on a set of characters for about 250 ms. This fixation is followed by a saccade, an eye movement of about 10 characters to the right, where the eyes will stop at the next fixation. A regression, or backwards eye movement in the text, is a sign that the reader is having difficulty understanding the material.

A high value of regression can be an indication of low readability of the text. The summary is being created by observing the reading pattern of the article. The summary generated for an article with low readability will contain so many sentences with jumbled sentence order. This can indicate be a nice way of research.

2 Relation between article quality and gaze pattern

In this thread, I plan to explore if there exists any relation between how people read an article and how it develops. Whether the articles which are read in sequential manner (i.e. linearly from one to another section) have more chances to become a "higher quality" article. Here quality indicates one of the quality labels from stub to FA.

I also plan to explore if the articles which are viewed more number of times and which has more "reading proportions" i.e. users are reading a good amount of the article when they refer it; tend to rank up in the quality leader or not.

References

[1] D. Beymer and D. M. Russell, "Webgazeanalyzer: A system for capturing and analyzing web reading behavior using eye gaze," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, (New York, NY, USA), pp. 1913–1916, ACM, 2005.

[2] C. Okoli, M. Mehdi, M. Mesgari, F. Å. Nielsen, and A. Lanamäki, "The peoples encyclopedia under the gaze of the sages: A systematic review of scholarly research on wikipedia," *Available at SSRN 2021326*, 2012.

[3] D. Wilkinson and B. Huberman, "Assessing the value of cooperation in wikipedia, digital libraries (cs. dl)," *Computers and Society; Physics and Society*, 2007.

[4] A. Kittur and R. E. Kraut, "Harnessing the wisdom of crowds in wikipedia: quality through coordination," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 37–46, ACM, 2008.

[5] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, "Information quality work organization in wikipedia," *Journal of the American society for information science and technology*, vol. 59, no. 6, pp. 983–1001, 2008.

[6] K. Rayner, "Eye movements in reading and information processing: 20 years of research.," *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.