

A Visualization Bookmarklet for Assessing the Readability of Texts on the Web

Frode E. Sandnes

Oslo Metropolitan University, 0130 Oslo, Norway

Evelyn Eika

Oslo Metropolitan University, 0130 Oslo, Norway

Readability is an important goal of texts on the web. Manual inspection of readability can be time consuming. Several readability indices have been proposed but their calculations often require text to be copied into specific tools. We therefore developed a bookmarklet which allows the readability of websites to be visualized directly in the browser. The background of different passages of text is marked according to their readability with heat-map colors, using established readability indices. Frequent words and difficult words are marked with colors.

CCS CONCEPTS • Human-centered computing~Visualization~Visualization techniques • Human-centered computing~Accessibility~Accessibility design and evaluation methods

Additional Keywords and Phrases: Readability, Accessibility, Web, Readability score, Reading, Reading level, Visualization, Heat map, Language teaching

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 Introduction

A textual message needs to be readable to successfully reach an audience. Several factors affect readability such as how familiar the reader is with the words and the complexity of sentence structures. Several readability scores have been proposed that are typically based on word complexity (number of letters of syllables) and sentence lengths. Many tools are available for computing such readability scores including web-based tools where users can paste texts, or word processors with built-in readability score computation. Such tools can be useful for content owners, but they may be more cumbersome for other stakeholders such as teachers who need to vet textual material to use in class.

A bookmarklet was therefore implemented which allows readers to rapidly obtain readability information about texts on the web without having to copy text into other tools. A bookmarklet is a JavaScript program installed in the browser. Example use cases include extracting metadata from webpages [9] and assessing

accessibility [10]. Bookmarklets are activated by applying a bookmark to the current page loaded in a browser. The current bookmarklet creates a text heat-map allowing the readability of various contents to be visually inspected in context of the webpage without the user having to read the texts in detail. Word level details are provided by highlighting frequent words, ordinary words, and difficult words.

2 Related work

The literature on text readability, readability indices [1], and web readability [2, 3] is vast. Although most studies focus on textual readability, the readability is also affected by legibility of presentation [13, 14]. Several studies have explored means of visualizing readability [4, 6, 8]. It must also be noted that readability is limited to the visual domain as low readability also can affect screen reader users [5, 11, 15]. Other works have looked at the effects of grammar checking tools on readability [7].

3 The readability bookmarklet

The readability bookmarklet was implemented in JavaScript. It starts by traversing the Document Object Model (DOM) to identify all passages of text within the body of the html-document. A html span element (a generic inline container) with a unique identifier is injected around each passage of text. Also, span elements with unique id are injected around each word in each passage. These span elements allow easy subsequent access to the text and the words for performing text passage analysis, word level analysis, and altering text and background colors.

Passages of text are analyzed using the median reading grade reported by several well-established readability measures as this aggregated measure gives a more robust measure than any of the individual measures (using an implementation by Simo Ahava). The readability score was normalized to fit the heat-map hue range anticlockwise on the HSL color wheel from green representing easy to read texts (120 degrees hue angle) to purple representing hard to read texts (300 degrees hue angle). The background color of the entire text passage was set to the given hue fully saturated with a lightness of 75% to allow sufficient contrast between the text and the background (see Figure 1). The background of the text consumes most of the text real-estate thereby attracting the viewer's attention.

The individual words of the passage were also analyzed, and the text color marked using three categories (see Figure 2), namely frequent words (green), words defined as difficult (red) and other words (black). For this analysis a stem of each word was applied using Porter's stemming algorithm [12]. A list of the 3,000 most frequent words in the English language was used as a source. A list of 300 difficult SAT words was used as a source for difficult words. Words not occurring in either list were classified as other words. These other words were less frequent but not necessarily more difficult. Tooltips added to each passage allow the user to inspect details of the different readability indices.

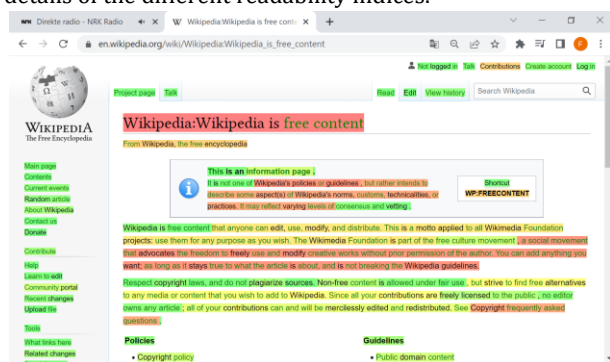


Figure 1. Readability visualisation of a Wikipedia page.

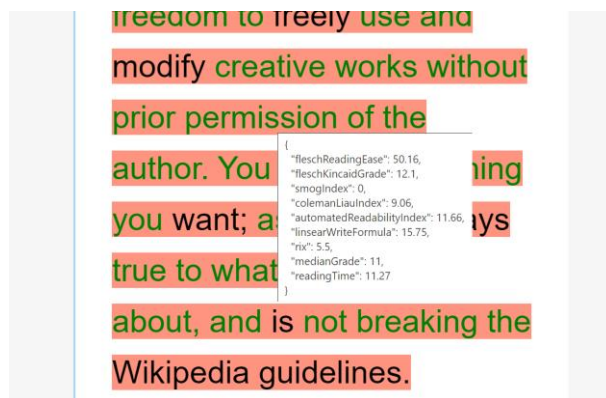


Figure 2. Most frequent English words in green. Tooltip lists detailed readability statistics.

4 Conclusion

A bookmarklet implementation of a readability visualization for texts of the web was described. Color is used to provide a text heat map allowing the viewer to get a rapid impression of the readability of the various text passages on a page. Once the bookmarklet is imported into the browser it can be easily applied to any websites with a few mouse clicks, providing rapid access to readability information. The bookmarklet may be useful for stakeholders that rapidly need to vet web texts, such as web accessibility auditors or teachers looking for reading material for their students. The bookmarklet is available at <https://github.com/frode-sandnes/READABILITY>.

References

- < bib id="bib1">< number>[1]</ number>Kevyn Collins-Thompson. 2014. Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics* 165, 97-135.</ bib>
- < bib id="bib2">< number>[2]</ number>Evelyn Eika. 2016. Universally designed text on the web: towards readability criteria based on antipatterns. *Stud. Health Technol. Inform* 229, 461-470.</ bib>
- < bib id="bib3">< number>[3]</ number>Evelyn Eika and Frode E. Sandnes. 2016. Assessing the reading level of web texts for WCAG2.0 compliance—can it be done automatically? *Advances in Design for Inclusion: Proceedings of the AHFE 2016 International Conference on Design for Inclusion*. Springer International Publishing, 361-371.</ bib>
- < bib id="bib4">< number>[4]</ number>Evelyn Eika and Frode Eika Sandnes. 2016. Authoring WCAG2.0-compliant texts for the web through text readability visualization. In *International Conference on Universal Access in Human-Computer Interaction*, Springer, Cham, 49-58.</ bib>
- < bib id="bib5">< number>[5]</ number>Bam Bahadur Kadayat and Evelyn Eika. 2020. Impact of sentence length on the readability of web for screen reader users. In *International Conference on Human-Computer Interaction*, Springer, Cham, 261-271.</ bib>
- < bib id="bib6">< number>[6]</ number>Saurav Karmakar and Ying Zhu. 2010. Visualizing multiple text readability indexes. In *2010 International Conference on Education and Management Technology*, IEEE, 133-137.</ bib>
- < bib id="bib7">< number>[7]</ number>Hitesh Mohan Kaushik, Evelyn Eika, and Frode Eika Sandnes. 2020. Towards universal accessibility on the web: do grammar checking tools improve text readability?. In *International Conference on Human-Computer Interaction*, Springer, Cham, 272-288.</ bib>
- < bib id="bib8">< number>[8]</ number>Hyoyoung Kim, Dongseop Lee, and Jin Wan Park. 2011. Textual visualization based on readability. In *SIGGRAPH Asia 2011 Posters (SA '11)*. Association for Computing Machinery, New York, NY, USA, Article 9, 1. <https://doi.org/10.1145/2073304.2073314></ bib>
- < bib id="bib9">< number>[9]</ number>Kun Ma and Ajith Abraham. 2015. Bookmarklet-triggered literature metadata extraction system using cloud plugins. In *Pattern Analysis, Intelligent Security and the Internet of Things*, Springer, Cham, 351-359.</ bib>
- < bib id="bib10">< number>[10]</ number>Andrea Mangiatordi and Harpreet Singh Sareen. 2011. Farfalla project: browser-based accessibility solutions. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '11)*. Association for Computing Machinery, New York, NY, USA, Article 21, 1-2. <https://doi.org/10.1145/1969289.1969317></ bib>
- < bib id="bib11">< number>[11]</ number>Prabin Parajuli and Evelyn Eika. 2020. A comparative study of accessibility and usability of Norwegian university websites for screen reader users based on user experience and automated assessment. In *International Conference on Human-Computer Interaction* Springer, Cham, 300-310.</ bib>
- < bib id="bib12">< number>[12]</ number>Martin F. Porter. 1997. An algorithm for suffix stripping. In *Readings in information retrieval*, 313-316.</ bib>
- < bib id="bib13">< number>[13]</ number>Frode Eika Sandnes. 2016. Understanding WCAG2.0 color contrast requirements through 3D color space visualization. *Stud. Health Technol. Inform*, 229, 366-375.</ bib>

< bib id="bib14">< number>[14]</ number>Frode Eika Sandnes and Anqi Zhao. 2015. An interactive color picker that ensures WCAG2. 0 compliant color contrast levels. *Procedia Computer Science*, 67, 87-94.</ bib>
< bib id="bib15">< number>[15]</ number>Usama Sarwar and Evelyn Eika. 2020. Towards more efficient screen reader web access with automatic summary generation and text tagging. In *International Conference on Computers Helping People with Special Needs*, Springer, Cham, 303-313.</ bib>