

# Analysis of Reading Difficulty in Web Environment

Martin ŠTRBÁK\*

*Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
xstrbak@fiit.stuba.sk*

Software development is an ongoing process consisting of requirements gathering, design and implementation of individual modules of the system. Despite the best efforts the requirements all future users cannot be met. There is no best design or configuration. An important feature is to be able to understand users' needs and based on that, we can modify the content, appearance, and applications behaviour. For us, the most interesting are Web applications. Personalization of Web pages is done by collecting the characteristics of the individual, his hobbies and search queries, how he interacts and how long he has been on the page. Further information can be obtained using questionnaires, but they are perceived as annoying and also might not be objective.

Website also contains large amounts of text, which has a certain size, font and text density. One person prefer larger text written Serif font, another on the other hand dense writing, but more paragraphs. Reading analysis consists of recording the movement of the eye, from which we derive valuable information as subject reads. The process of reading is a cognitive activity, which can be recorded by EEG and processed. It can be assumed that a person reading the more difficult text demands a higher concentration than reading of a simple text. Analysis of measurements of multiple documents in different classes of complexity can provide us additional information to user modelling, which helps us to specify the content provided [1].

We assume that the amount of textual content on the Internet will only increase, and so will information congestion of readers. They do not know in advance whether the article he plans to read is written vaguely, if he expects scientific work. Different person might find the same article as satisfactory. If we can grade the text complexity, we could provide more accurate recommendation.

Textual complexity might be based on multiple factors [2]:

- Text structure

---

\* Supervisor: Marián Šimko, Institute of Informatics, Information Systems and Software Engineering

- Semantics
- Errors
  - Grammar
  - Typos

We will focus on errors, mainly typos, where few letters in words are mixed. Text will be simple by any other means, for example, using only frequent words with average length. As study [3] suggests, with EEG we can distinguish reading from other mentally not difficult activities. Next, they have find out that even different kinds of text content can be known. Some irregularities occurred due to face muscle induced noise. We assume, that errors distract human reader, and any focus needed to regain text context, will result in more cognitive load.

EEG is device used to measure electrical impulses generated by brain. Unfortunately, this offers poor signal-to-noise ratio. We can use band pass filter to obtain only frequencies in 1 – 30 Hz, which is known to be the most important part of the signal. However, some artefacts will remain:

- Artefacts of heart activity
- Artefacts invoked by eye movement
- Artefacts invoked by muscle contraction

To get additional information about reading, eye-tracking device is being used. This way we can map EEG signal to text being read. After this study, we will better understand how errors correlate with cognitive load and user experience.

## References

- [1] Sosnovsky, S., Dicheva D. Ontological Technologies for User Modelling. *International Journal of Metadata, Semantics and Ontologies* 5. Inderscience, pp. 32–71 (2010).
- [2] Biedert, R., Hees, J., Dengel, A., & Buscher, G. A Robust Realtime Reading-Skimming Classifier. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, pp. 123-130 (2012).
- [3] Kunze, K., Shiga, Y., Ishimaru, S. & Kise, K. Reading activity recognition using an off-the-shelf eeg - Detecting reading activities and distinguishing genres of documents. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 96-100 (2013).