

DysWebxia: A Model to Improve Accessibility of the Textual Web for Dyslexic Users

Luz Rello

NLP and Web Research Groups, Department of Information and Communication Technologies, Universitat Pompeu Fabra
C/ Tànger 122-134, 08018 Barcelona
luz.rello@acm.org

Abstract

The goal of this research is to make textual content in the Web --especially in Spanish and English-- more accessible to people with dyslexia. The techniques that we will use to make the Web more accessible are Natural Language Processing (NLP) for its content (text) and Web design guidelines for its layout. To find out which solutions tackle better our purpose we will test a diverse set of Web pages examples. The main methodology to evaluate these examples will be eye tracking using regular and dyslexic students. In the case that our findings show that there are strategies that make the Web more accessible for dyslexic users, we plan to develop an application which includes such results, transforming a regular Web site into a dyslexic friendly Web site.

Introduction

Broadly, this thesis addresses a real world problem: Web accessibility for dyslexic users. Given that dyslexia is a learning disability that affects language, we can assume that accessibility can be approached from two dimensions: (1) form --HTML layout-- and (2) content --text--. Although there are applications that make the Web more accessible for dyslexic users, these only modify its design but not its content.

In particular, this Ph.D. thesis explores text accessibility for dyslexic users. This leads to diverse scientific challenges where some of them will be tackled for the first time. We will first study how dyslexia affects the different levels of language and find out which linguistic properties of the text are useful for our goal. Then, we will explore which of the existing and potential NLP techniques have a beneficial impact on accessibility. The results will be used for the development of a tool, DysWebxia, which will output friendly Web pages for dyslexic users.

Motivation

There are three main factors that led to my choice of topic: (1) its social relevance, (2) its scientific challenge, and (3) its usefulness.

First, dyslexia is a universal neurologically-based disorder [11] and, unfortunately, it is a fairly frequent condition (according to the National Academy of Sciences, 10-17% of the U.S. population has dyslexia [9]). We made an estimation of the presence of dyslexic texts in the Web to know the real impact of dyslexia and our results show that at least 0.28% of Web pages in English has dyslexic text [1]. Since dyslexia is a learning disability that affects the acquisition and the processing of language, learning periods become crucial to overcome it.

Second, our approach is novel considering the related work. Although there is already software (e.g., spell-checkers) designed specifically for dyslexics, most frequently visited Web pages currently offer no accessibility options for their dyslexic users. Moreover, most of the various studies that take into account dyslexia from the Web accessibility point of view are focused on the design characteristics of the Web page [8], e.g. special text formats for dyslexic users [7]. Although there are some recent studies that use NLP techniques (text simplification) for people with special needs [4], to the best of our knowledge, the use of such strategies considering dyslexic users together with eye tracking has not yet been approached. So our focus is on the content, not in the design of Web pages, although we will take into account the extensive literature to design the layout of our prototypes.

Finally, a tool that makes the Web more accessible for dyslexic users will be beneficial for all people, since dyslexic-accessible practices redress difficulties encountered by all Internet users [8]. There is a common agreement in related work that the application of dyslexic-accessible practices benefits also the readability for non-dyslexic users [10, 8], as well as other users with disabilities such as low vision [6, 7, 5]. The development of this tool would improve the access of dyslexic people to information content, being a help for overcoming dyslexia during the learning process.

Stages

Methodology and Techniques

A Theoretical Framework

- | | | |
|---|--------|---|
| 1- State of the Art | —————→ | Literature review |
| 2- Motivation
(relevance of Dyslexia in the Web) | —————→ | Information retrieval and Web mining
(Baeza-Yates & Rello, 2011a, 2011b) |
| 3- Description of Dyslexia | —————→ | Linguistics methodology |

B Experimental Studies

- | | | |
|-------------------------------|---------------|--|
| 4- Generation of Web Examples | ↙ Design ———→ | Information architecture and usability strategies inspired by work on human-computer interaction |
| | ↘ Text ———→ | NLP techniques based on linguistic properties |
| 5- Testing of Web Examples | —————→ | Eye-tracking complemented by Amazon Mechanical Turk and user engagement measures |

C Development

- | | | |
|---|--------|-------------------|
| 6- Development and Evaluation of the Selected NLP Modules | —————→ | NLP Methodologies |
| 7- Development of a Beta Version of the A-Dyslexic Web tool | | |

Figure 1. Ph.D. summary.

Methodology

Our research include threes main stages (see Fig. 1). Stage A includes the state of the art; the motivation and a linguistic description of dyslexia to find out the linguistic components that potentially improve accessibility. During Stage B a set of Web pages examples will be created taking into account different NLP techniques, linguistic knowledge and design strategies.

We will explore the effectiveness of each strategy by testing the examples using eye tracking considering two groups of students: (i) regular native students --control group— and (ii) dyslexic students --with dysphonetic dyslexia [3]. To complement the findings of eye tracking we will use crowdsourcing (e.g. Amazon Mechanical Turk) and user engagement measures. Once we know which results help dyslexic users to read and access information, we will (stage C) develop the actual software that could take into account the strategies devised and integrate them into browsers.

Status of the Research

This PhD is on its early stage of research as I am in my first Ph.D. year. So far I am finishing the accomplishments of stage 1 (Section Methodology). I have reviewed the literature to find out the open problem and I have worked on the motivation of this thesis estimating the impact of dyslexia in the Web [1, 2]. At this moment I am working on the detailed linguistic description of the problems that people with dyslexia experience and the experimental design [3].

Envisioned Contributions

The expected contributions of my Ph.D. Thesis are:

- A description of the state of the art of dyslexia related to accessibility, linguistics and natural language processing.
- Studies about the presence of dyslexia in the Web using Web mining methodologies.
- Finding of features that contribute to make the text more accessible for dyslexic users.
- The adaptation and development of the NLP techniques that benefits accessibility for dyslexics.
- A beta version of a tool for making Web sites more accessible.

Doctoral Consortium

I expect to gain from the doctoral consortium: (1) a constructive criticism to my Ph.D. proposal which I hope to put into practice; (2) advise from experts in accessibility since the topic is fairly original and some questions have risen throughout my research so far; and (3) meet researchers in this field, which is new for me, who come from a NLP and linguistics background.

Acknowledgments

I thank my advisors, professors Ricardo Baeza-Yates and Horacio Saggion, whose support has been crucial so far. These Ph.D. studies are supported by a FI-DGR grant funded by the Government of Catalonia and the European Union.

References

- [1] R. Baeza-Yates and L. Rello. Estimating dyslexia in the Web. In Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A 2011), 2011.
- [2] R. Baeza-Yates and L. Rello. How bad do you spell?: The lexical quality of social media. In Proceedings of Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011) Workshops Publication, the Future of the Social Web (FOSW 2011), 2011.
- [3] E. Boder and S. Jarrico. The Boder's Test of Reading-Spelling Patterns: A Diagnostic Test for Subtypes of Reading Disability. Grune and Stratton, New York, 1982.
- [4] S. Bott and H. Saggion. An unsupervised alignment algorithm for text simplification corpus construction. In Proceedings of the Workshop on Monolingual Text-To-Text Generation, pages 20–26, Portland, Oregon, June 2011. Association for Computational Linguistics.
- [5] P. Brophy and J. Craven. Web accessibility. Library Trends, 55(4):950–972, 1964.
- [6] B. Caldwell, M. Cooper, L. Guarino Reid, and T. Vanderheiden. Web Content Accessibility Guidelines (WCAG) 2.0. MIT, ERCIM, Keio, 2008.
- [7] L. Evett and D. Brown. Text formats and web design for visually impaired and dyslexic readers-clear text for all. Interacting with Computers, 17:453–472, July 2005.
- [8] J. E. McCarthy and S. J. Swierenga. What we know about dyslexia and web accessibility: a research review. Universal Access in the Information Society, 9:147–152, June 2010.
- [9] H. Meng, S. Smith, K. Hager, M. Held, J. Liu, R. Olson, B. Pennington, J. DeFries, J. Gelernter, T. O'Reilly-Pol, S. Somlo, P. Skudlarski, S. Shaywitz, B. Shaywitz, K. Marchione, Y. Wang, P. Murugan, J. LoTurco, P. Grier, and J. Gruen. DCDC2 is associated with reading disability and modulates neuronal development in the brain. In Proceedings of the National Academy of Sciences, volume 102, pages 17053–17058, November 2005.
- [10] M. G. Paciello. Web accessibility for people with disabilities. CMP Books, Lawrence, Kansas, 2000.
- [11] E. Paulesu, J.-F. Démonet, F. Fazio, E. McCrory, V. Chanoine, N. Brunswick, S. F. Cappa, G. Cossu, M. Habib, C. D. Frith, and U. Frith. Dyslexia: Cultural diversity and biological unity. Science, 291(5511):2165–2167, November 2001.

About the Author:

Luz Rello is a Ph.D. student at Universitat Pompeu Fabra. She is founding member and associate editor of "*Diálogo de la Lengua*", Erasmus Mundus Course Representative, Google Anita Borg Scholar in 2011 and collaborates in the forthcoming Oxford Dictionary of Computational Linguistics.

