# ../../../../../../../../../../Users/andreacurley/Desktop/Screen%20

# MSc in Computer Science - Team Project

# Interim Report Template

# Team Name:

## Introduction

Dyslexia, a neurological disability, affects approximately 10% of the global population. This condition challenges individuals in multiple linguistic areas, such as phonology, orthography, morphology, lexicon, syntax, and discourse. Individuals with dyslexia often experience prolonged reading times and reduced comprehension levels, leading to the need for repeated readings. Despite the prevalence of dyslexia, there remains a lack of comprehensive guidance in document content and formatting that caters to this demographic. This gap highlights the need for technological advancements and improved document formatting strategies to enhance the accessibility of online materials.

The primary goal of this project is to develop a specialized website designed to address the unique needs of individuals with dyslexia, with a central focus on enhancing the readability of their documents. Recognizing the diverse range of documents people interact with daily, the platform will accommodate multiple file types, including docx, ppt,xls . After content is uploaded, the system will carefully analyze it to find any accessibility issues that could impair reader’s ability to comprehend it. The platform will proactively suggest actionable improvements, ensuring that the content aligns better with the reading preferences of those with dyslexia. Users are given the authority to make final decisions after the suggested improvements are presented. They can choose to adopt the adjustments selectively in accordance with their own preferences.

The document will be displayed in an interactive viewer on the platform after optimization. Users can further modify the document's appearance and organization using this dynamic interface to make sure it precisely suits their reading comfort. When users are satisfied with the changes, they can easily download the revised document ensuring they have a version tailored for optimal readability, which they can use across various platforms and devices.

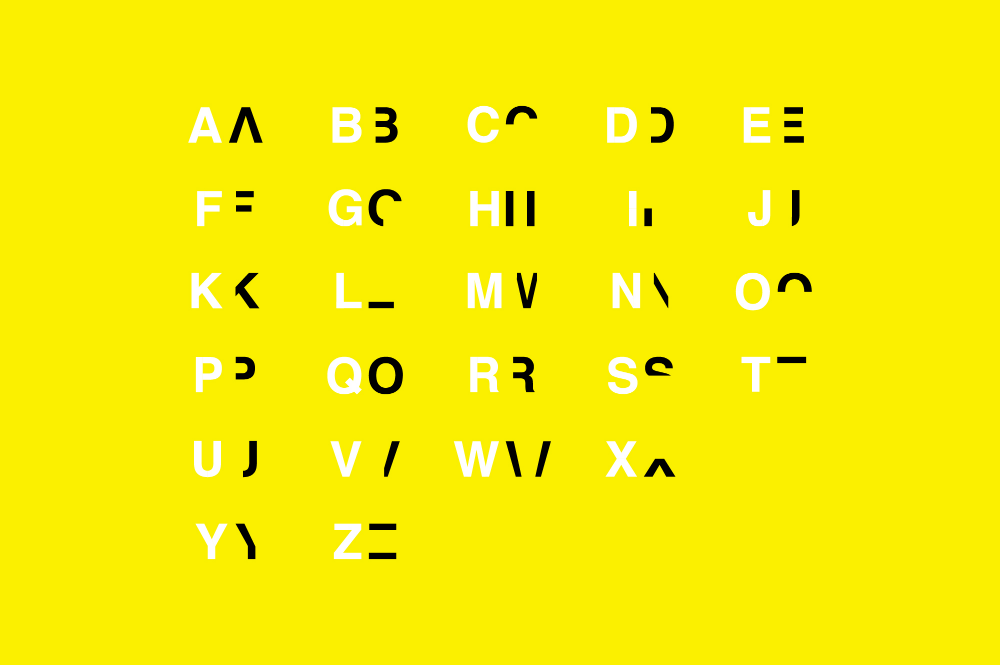
## 1. User Scenario: The Characters (500 words approx.)

* Who is your target user? User personas (diagram,description. How we identified them)
* Why are they important?
* What problem are you solving for them?

The two primary target user groups we identified are people with dyslexia and people who may know/care for them. These user groups help in acknowledging and comprehending the unique demands of individuals who have dyslexia as well as those who may know or assist them. Our designated user group consists of members from the Dyslexia Association of Ireland. We have maintained ongoing communication with them to ensure their active involvement through surveys and feedback, in the development and testing of our system.

People with dyslexia make up a substantial section of the online user base due to the ubiquitous nature of the disorder. By giving accessibility for users who are dyslexic priority, we help to create a more democratic online environment. It makes sure a larger audience can access information and services, promoting inclusivity and equitable opportunity for everyone.

Participants in a study by Grusky, Taft, Naaman, and Azenkot at Cornell Tech, particularly those who had dyslexia, used practices of skipping and skimming sections of text that posed comprehension challenges, as well as the act of re-reading to solidify understanding. The survey participants with dyslexia consistently reported higher frequencies of these behaviors in comparison to their counterparts without dyslexia.[5]



[2] <https://danielbritton.info/dyslexia/>

The consequences of dyslexia extend beyond academic difficulties. Individuals with dyslexia often experience prolonged reading times and reduced comprehension levels, leading to the need for repeated readings. This struggle can lead to heightened levels of stress, anxiety, and depression. Moreover, the societal stigma surrounding dyslexia can exacerbate emotional and psychological burdens, affecting self-esteem and motivation We want to significantly improve this population's reading experience using various document optimization techniques

Our website provides a user-friendly platform for individuals to upload documents. Our system thoroughly analyzes the document after uploading it to find any accessibility issues that may be there. The relevant adjustments are then automatically made for optimal readability.

## 2. Technical Problem: The Setting (1,000 words approx.)

* Why does your system exist?

In accordance with the definition provided by the International Dyslexia Association (IDA) [14], dyslexia are identified as a specific neurobiological-based learning disability. It leads to difficulties in recognizing words accurately and smoothly and in spelling and decoding. These issues are mostly caused by a phonological language impairment which causes individuals with dyslexia to face challenges in understanding what they read and may have a less fulfilling reading experience. It is estimated that around 10% of the population is affected by dyslexia (dyslexia association Ireland).

Even though there are many accessibility-enhancing systems in the market, all of them cater to real-time reading enhancements. One major difference between the existing systems is that our application provides the user with the ability to export the reformatted document which they can re-read at their own convenience. Most of the existing systems offer design-based formatting like font, layout etc. Our system makes use of Data Science techniques to transform the content of the document as well as make it more optimized for readers with dyslexia.

* What is the core technical problem? (provide an example, an image or a diagram that describes the technical components)

The core technical problem is to convert common document formats (such .doc, .docx, .ppt, .pptx, .xls, and .xlsx) into more accessible formats which are constructed according to the unique requirements of people with dyslexia.

The following is a list of features to be built in the Accessibilator:

1. Document Parser: Understand and extract content from different proprietary or semi-proprietary formats.
2. Content Extraction: Identify and classify various elements in the document (e.g., headings, paragraphs, images, tables, charts).
3. Accessibility Transformation: Modify the document to make it more accessible, I.e, in adherence to the accessibility standards (e.g., alt text for images, structured hierarchy for headings, readable fonts (like Lexand, Comic Sans MS), and maintaining good colour contrast.
4. Output to Accessible Format: Choose a universally readable format (like PDF, or the original format of the document) and encode the transformed content into it.

To illustrate with an example:

1. Input: A .docx document containing text, an image without alt text, and a table.
2. Parsing: The tool identifies the document as a .docx format and uses a relevant parser to read it.
3. Content Extraction: The tool extracts the text, recognizes the image and table, and reads any embedded styling or metadata. This data is then used for further transformation.
4. Accessibility Transformation: The tool adds a placeholder alt text to the image (or prompts the user for one), ensures the table has proper headers, and adjusts any problematic color contrasts or font sizes or font colour. It also suggests some other accessibility changes which are optional in nature.
5. Output: An accessible PDF or a document (with formats like .doc, .docx, .ppt, .pptx, .xls, .xlsx) that includes the revised content.

* Can you review other existing systems or products that address this problem? (how do they meet or fail to meet the needs of your target users) (compare existing products)

## 3. Technical Solution: The Plot (1,000 words approx.)

* What does your system do?
* How does it work? (System diagram)
* Front-end: Technologies, User interface components including interface mock-ups
* Back-end: Technical components

Java:

Since our application is primarily web-based, Node.js was the natural choice for backend as this would establish uniformity between the layers and due to the ease of development which NODE.JS offers. But while researching available document parsing libraries, we found that the choices were limited in Node.js. For each document type like docx, ppt etc we would have to use libraries and some of them did not have active support. So, we made the informed choice to pivot to Java owing to the support of many efficient document parsing libraries. We have currently chosen Apache POI as the document parsing library owing to its ability to parse different document types like docx,ppt,excel, and its rich feature set that allows you to read and edit documents efficiently.

Apache POI has a large and active community of developers, which means you can find support, tutorials, and resources online. Additionally, the library has extensive documentation and examples to help you get started.

SpringBoot Framework:

Amazon DynamoDB:

Amazon S3:

* Data: What data resources are you going to use and how will you access, collect, and store them?

## 4. Evaluation: The Reviews (500 words approx.)

* What does success look like for your system?
* How will you evaluate the system that you built?

## 5. Conclusion: The Plan (500 words approx.)

* What is your project management strategy?
* What are the biggest challenges you are currently facing?
* How will you use the time remaining to achieve a successful outcome?

## 6. References and Key Resources

* List of resources (software, papers, tutorials, books, stats, business indicators)

**References**

1. British Dyslexia Association. (2018). *Dyslexia friendly style guide*. British Dyslexia Association. <https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide>
2. Britton, D. (2019, July 22). *Dyslexia*. Daniel Britton. <https://danielbritton.info/dyslexia/>
3. de Santana, V. F., de Oliveira, R., Almeida, L. D. A., & Baranauskas, M. C. C. (2012). Web accessibility and people with dyslexia. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A ’12*. <https://doi.org/10.1145/2207016.2207047>
4. Galliussi, J., Perondi, L., Chia, G., Gerbino, W., & Bernardis, P. (2020). Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: testing text readability in people with and without dyslexia. *Annals of Dyslexia*, *70*(1), 141–152. <https://doi.org/10.1007/s11881-020-00194-x>
5. Grusky, M., Taft, J., Naaman, M., & Azenkot, S. (2020). Measuring and Understanding Online Reading Behaviors of People with Dyslexia. *In Submission*. <https://par.nsf.gov/biblio/10170285>
6. McCarthy, J. E., & Swierenga, S. J. (2009). What we know about dyslexia and Web accessibility: a research review. *Universal Access in the Information Society*, *9*(2), 147–152. <https://doi.org/10.1007/s10209-009-0160-5>
7. Rello, L., & Baeza-Yates, R. (2015). How to present more readable text for people with dyslexia. *Universal Access in the Information Society*, *16*(1), 29–49. <https://doi.org/10.1007/s10209-015-0438-8>
8. Rello, L., & Baeza-Yates, R. (2016). The Effect of Font Type on Screen Readability by People with Dyslexia. *ACM Transactions on Accessible Computing*, *8*(4), 1–33. <https://doi.org/10.1145/2897736>
9. Rello, L., Baeza-Yates, R., Bott, S., & Saggion, H. (2013). Simplify or help? *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A ’13*. <https://doi.org/10.1145/2461121.2461126>
10. Rello, L., Kanvinde, G., & Baeza-Yates, R. (2012). Layout guidelines for web text and a web service to improve accessibility for dyslexics. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A ’12*. <https://doi.org/10.1145/2207016.2207048>
11. Rivero-Contreras, M., Engelhardt, P. E., & Saldaña, D. (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: effects of visual support and word frequency. *Annals of Dyslexia*, *71*(1), 170–187. <https://doi.org/10.1007/s11881-021-00217-1>
12. Stark, Z., Franzen, L., & Johnson, A. P. (2021). Insights from a dyslexia simulation font: Can we simulate reading struggles of individuals with dyslexia? *Dyslexia*. <https://doi.org/10.1002/dys.1704>
13. Yoliando, F. T. (2020, December 3). *A Comparative Study of Dyslexia Style Guides in Improving Readability for People With Dyslexia*. [Www.atlantis-Press.com](http://Www.atlantis-Press.com); Atlantis Press. <https://doi.org/10.2991/assehr.k.201202.050>
14. International Dyslexia Association. (2002). *Definition of Dyslexia*. International Dyslexia Association. <https://dyslexiaida.org/definition-of-dyslexia/>