

Readability Index and Text Optimization Model for Dyslexics Using NLP and AI

Esha Shah

*Chirec International School
Hyderabad, India
eshah.shah.2028@gmail.com*

Vinay Vishwakarma

*On My Own Technology
Mumbai, India
vinay.vishwakarma@omotec.in*

Abstract: This paper presents a new readability index to measure the reading difficulty level of text for dyslexic readers, and explores the effectiveness of an NLP model used to simplify input text to generate output text with a better readability score according to this index. The readability index for dyslexics is programmed in Python using NLP libraries such as NLTK, spaCy, and Textstat. It incorporates an algorithm that computes weighted parameters, which are accepted in the literature as impacting text readability for dyslexics. These include sentence length, syllable count, passive voice, sentence complexity, mirrored letters, double negatives, block letters, and Flesch reading ease. In addition, a model was made to generate text at a better readability score for dyslexics, using GPT-3, and text at varying levels of pre-determined difficulty was used as input to test and improve it. The text data for testing was collected from the Commonlit website, a freely available resource, to generate paragraphs according to people's reading level (i.e. 4th to 8th grade). The accuracy and effectiveness of this readability index and text simplification model were tested with dyslexic students at an English-medium Indian private school. The developed model demonstrates an enhancement in paragraph readability for individuals with dyslexia, resulting in a 10% increase in accuracy rate to 75%, compared to the original paragraph test. These are encouraging findings from the initial research. Further research can improve both the accuracy of the readability index for dyslexics and more sophisticated text optimization algorithms can be developed.

Keywords: *Dyslexia, Readability Index, Natural Language Processing, TextStat, Regex, GPT-3*

1. INTRODUCTION

Dyslexia is a language-based learning difficulty that affects an individual's reading skills by impairing their phonological awareness, verbal memory, and verbal processing speed¹. It affects 15-20% of the population and is the most common neuro-cognitive disorder. Despite this fact, there is a lack of readability indices (methods of measuring the reading difficulty level of a text) specifically tailored to provide an accurate measure of readability for dyslexics, as opposed to the general population.

There have been previous studies where researchers have tested models used to evaluate the cognitive accessibility of texts for readers with learning disabilities. In this context, there are several factors that determine the readability of a text, particularly content-related determinants and visual-related determinants. Previous research and literature, examined in the next section, have also explained the role these various determinants (such as typography, mirrored letters, word complexity, sentence structure, etc.) play in increasing or decreasing the readability of a text. This literature can be grouped into two distinct areas of study. The first is related to the specific techniques employed using natural language

¹ British Dyslexia Association. (n.d.). *What is dyslexia?*

<https://www.bdadyslexia.org.uk/dyslexia/about-dyslexia/what-is-dyslexia>

processing to investigate methods of improving text accessibility; the second is related to the effect of textual parameters (such as the length of a sentence, the use of active voice, or the presence of certain punctuation) in determining the readability of a particular text. However, none of the previous research attempted to create a comprehensive measure of readability, i.e., a specific readability score, as it relates to the accessibility of texts for dyslexic readers.

The goal of the research is to improve text readability analysis by using new factors including block letters, mirrored characters, word count, sentence length, syllable count, passive voice, and double negatives. An enhanced comprehension of text complexity is offered by this extended analysis, which enhances readability evaluations. The results are important for educators, content developers, and communicators since they provide information on how to modify texts for a range of audiences, including dyslexics, and guarantee the best possible comprehension on several platforms. This paper presents a comprehensive readability index for dyslexic readers, created in Python using libraries such as NLTK, spaCy, TextStat and Regular Expression, and explores how, using a GPT-3 transformer model, a given passage of text can be simplified to increase its comprehensibility. The readability index and model were trained on a corpus of text with passages about the same subject matter at different grade levels (third grade, eighth grade, and university level). The model was tested with dyslexic students in an Indian private school, and compared to reading parameters measured using Microsoft Team's Reading Progress feature in order to gauge its effectiveness.

2. LITERATURE REVIEW

Rello [1] created a model called DysWebxia to improve the accessibility of text (primarily in Spanish language) for dyslexics through content and presentation-related optimizations. An automated lexical simplification algorithm was integrated into this model to generate simplified synonyms for words in the text. This algorithm, CASSA (Context Aware Synonym Simplification Algorithm), was compared to a frequency-based algorithm, which generates synonyms based on their frequency in the most common *sense* (context). To create CASSA, the researchers modified the Spanish OpenThesaurus to create a list of unique senses, each with a target word, and established their frequency in the Web - which resulted in a list of synonyms for each unique sense, and their frequencies. In order to identify the sense of a target word, CASSA uses the word's context and selects the three most frequent synonyms from the most frequent sense, from the n-gram of its (word, context) pair. The experiment conducted to evaluate the quality of synonyms generated by CASSA for dyslexics suggested that CASSA generated simpler and more accurate synonyms than the frequency-based algorithm did. This facilitated the text modification aspect of DysWebxia, through automated lexical simplification, allowing it to be successfully implemented in tools such as Text4All and DysWebxia Reader. Saggion et al. [2] investigated the impact of verbal paraphrasing in Spanish on the readability and comprehension of text for people with and without dyslexia. The computational process for lexical simplification as opposed to manual simplification was done using the Badele.3000 database. Paraphrased phrases for the simplification, consisting of a support verb and lexical verb with a noun collocation, were extracted from this database - from this, the lexical verb has been considered to be the simplified version. In order to study this they performed an eye-tracking study which combined reading tests to measure readability, comprehension tests to measure understandability, and semi-structured interviews to gauge the participant's preference in terms of which text was more readable. The eye-tracking data from the study consisted of the average fixation duration and the total visit duration of the text passages to gauge the reader's processing load and reading speed (and thereby the text's readability). According to these results, lexical simplification through verbal paraphrasing did not improve readability or comprehension of text for dyslexics.

Rello et al. [3] proposed the Real Check system which identifies real-word spelling errors (spelling errors that form a real word, but not the word originally intended) in text written by people with dyslexia. Diagnosed dyslexic volunteers from Barcelona and Madrid were asked to procure texts written by them in

order to obtain a sample of real-world errors. Using the Levenshtein Automaton dynamic algorithm, the researchers generated Confusion Sets from a Spanish lexicon and identified any potential anagrams for the dictionary's words. An algorithm was devised where n-grams were extracted and compared from the original sentence and what was called the candidate sentences (sentences where each word has been changed for the words in its confusion set). Real word errors were detected from the original sentence if the 'candidate words' in the candidate sentence were more frequently present according to the n-gram corpus, and these candidate words became suggested corrections. The BerkeleyLM (language model) was used to filter out ungrammatical corrections. Other erroneous suggestions were filtered out according to a set of rules using a joint dependency parser, lemmatizer, part-of-speech tagger and morphological tagger. The results suggest that Real Check is able to more accurately detect and correct real word errors, and aid dyslexic readers in evaluating their written work. Fournery et al. [4] studied the cognitive accessibility of web pages to investigate questions related to their readability and relevance for dyslexic and non-dyslexic readers. By rating the top 20 search results from 10 different web searches on the Likert scale, the researchers determined whether the website was readable and whether it was relevant to the web search for each participant. The researchers then aimed to establish a correlation between these scores and the web pages' lexical and aesthetic features. An automated web browser was used to run custom scripts in the page's JavaScript coding to identify the web page's aesthetic features. The text within the web page was separated into sentences by doing an in-order traversal of the DOM tree, and their readability was computed using The GunningFog Index which estimates the grade level of a student in the U.S. who would be able to read that particular text. All features related to a web page such as the average line length, and sentence-to-text ratio, are associated with a web page's readability - for both dyslexics and non-dyslexics. Zabkar et al. [5] aimed to create a machine-learning model to identify dyslexic readers by using a readability application to assess their fluency in oral reading. They compared a group of dyslexic readers with a control group by transcribing the recordings of their oral readings and identifying the most frequent errors made while reading. Using these error types, parameters for assessing the text's readability were determined from the test battery (i.e. exam groups) of the Slovenian National External Assessment of Knowledge. On the basis of these parameters, machine learning models were trained and built to identify whether a person had dyslexia or not. A naive Bayesian classifier, a decision tree, and a FreeViz projection were used to predict the target outcome based on the parameters and establish a two-dimensional scatterplot with the project data. According to the results, the control group exhibited shorter reading times and paused before more difficult words while the dyslexic group demonstrated longer reading times and pauses - hence, the models were successfully able to differentiate between people with reading-related disabilities and the control group. Chakraborty et al. [6] aimed to create a readability tool for the Bengali language. They did this by adopting readability indexes used for English in accordance with the U.S.-based education system for Bengali with an appropriate age-to-age comparison. This readability tool also provided an analysis of written Bengali to give insights into its complexity. Using a pronunciation dictionary of the Bengali language, the researchers experimented with formula-based approaches, particularly Automated Readability Index (ARI), Flesch Reading Ease (FE), Flesch-Kincaid (FK), Gunning Fog (GF), SMOG, and Dale-Chall, to gauge the most appropriate one. ARI was found to be the most accurate one and was used to predict the age range for the readability of the input text. They created a Bengali Readability Analysis Tool which included, counting consonant conjunct, and extending the BiLSTM model by adding global average pooling and global max-pooling layers for their ablation experiments. The results suggest the maximum accuracy was a result of a combination of BiLSTM with pooling, CL, CC, and embeddings from the LaBSE model.

Baeza-Yates et al. [7] performed two eye-tracking studies to investigate whether more frequent synonyms and shorter words improve the readability and comprehensibility of a text. Participants in the two studies were made to read texts which were adapted to each condition: more/less frequent synonyms and longer/shorter words. To create synonym pairs based on their frequency (more or less), synonyms for target words within the text were obtained using a synonym dictionary, and their relative frequencies were recalculated using the advanced search of a major search engine. To find synonym pairs based on word

length (longer or shorter), the longest words from the Royal Spanish Academy Dictionary were selected and their shorter synonyms were found. The results suggest that participants with dyslexia read faster and had shorter fixation durations (pauses) when reading more frequent synonyms within the text. Shorter words also allowed dyslexic participants to increase their reading speed and better comprehend the text. Erlikhman et al. [8] investigate the legibility of mirror-reflected text, such as mirror-reversal errors when reading letters like 'b' and 'd', among other types of geometrically transformed text. Their paper reviews and evaluates current literature on this subject, and states that typical errors for text involving letters that are inversions, reflections or rotations of each other result in left-right mirror reversal errors, up-down mirror reversal errors and transposing the position of letters within words. The paper suggests that mirror-reversal errors in particular are normal for both children and adults when reading, but can persist longer for dyslexic readers. Ownby [9] conducted a study to investigate the influence of vocabulary, sentence complexity, and passive voice on the readability of mental health information on the Internet. Two types of searchers were performed on Google, Yahoo! and MSN to obtain the sample text on mental health - one that was more sophisticated ("geriatric depression") and one that was less sophisticated ("elderly sadness"). The samples were evaluated for their readability score using the Flesch Reading Ease Index and Flesch-Kincaid grade equivalent, and scores for their vocabulary complexity, sentence complexity and passive voices were also recorded. The results of the study suggest that vocabulary complexity contributes the most to differentiating a text with a higher readability from that with a lower readability. Sentence complexity and the use of passive voice (as opposed to active voice) also differentiate the least readable texts from the most readable ones but are not as consistent in differentiating the intermediate texts within the extremes of these two readability levels.

Collins-Thompson [10] did a survey of research on current and future computational methods of assessment for text readability, particularly focusing on algorithms used by these methods, their applications and future potential for research. The survey suggests that there are multiple factors involved in computational readability assessments including lexico-semantic features (ambiguous words), morphological features, syntactical features (grammar-related), discourse-related features (cohesion, transitions, argument structure etc.), higher-level semantics (idioms, implicit connotations etc.), pragmatic features (contextual language), and user-oriented factors. The combination of these features within machine learning models that can use them, enhances computational readability assessments. On the whole, readability measures that employ analysis of various linguistic factors within a text produce higher correlations for a text's readability. These readability measures were also determined to more accurately differentiate between readability scores for text at lower grade levels than higher grade levels. The applications of such assessments include gauging the readability of text for foreign language speakers, and most relevant to this research, for readers with disabilities, most notably, dyslexia. Yusri et al. [11] proposed a speed reading tool for students with ADHD, dyslexia or short attention spans. The AI-based speed reading model was trained on a wide text corpus. To analyze and summarize the sentences so they could be read more quickly, researchers used a Multilayer Perceptron method, made possible by the Hugging Faces Text-To-Text Transformer. The tool allows for text to be customized, inspired by principles of Bionic Reading, such as adjusting the bolding of text and spacing between characters, words, and lines. Half-bolding, where the first half of the characters in the summarized text were bold, was employed using a selection algorithm and is proven to draw the reader's attention to the text. Customizing character spacing also allows readers to suit their reading preferences. The web framework Flask was used to test the speed reading tool, and the interface was made user-friendly using HTML, CSS and JavaScript. The tool has been concluded to improve the readability and comprehension of texts for its intended audience, including dyslexic readers.

3. PROPOSED METHODOLOGY

This research paper focuses on evaluating the readability of textual content and improving the content for dyslexic persons. The methodology employed in this research involves a systematic approach to defining a paragraph's

readability index on the basis of various parameters. To evaluate the readability of text for dyslexics several parameters were taken into consideration which are explained below along with their weights based on their role in making text complex. Weights were chosen based on the frequency of encountering these parameters in texts and their significance to readability for dyslexics. For example, words with higher syllable counts are more frequent than double negatives or block capitals. Similarly, long sentences impact readability more negatively than passive voice - hence these were given a higher weightage.

- A. **Word and Sentence Count:** The basic level of text analysis is word and sentence count analysis, which provides information on the length and complexity of a document. One can quickly learn important details about the type and complexity of the material by quantifying the linguistic parts.
- B. **Average Sentence Length:** Longer sentences generally imply a more complex composition. Essentially, it measures the average number of words in a statement. Higher averages frequently denote more intricate sentence structures (such as compound, complex or compound-complex), which may make the content harder for readers to understand. Longer sentences can also be a result of over use of semicolons or conjunctions to join multiple ideas together.
- C. **Syllable Count and Average Syllables per Word:** This aspect of the method explores word usage complexity. Words with more syllables frequently allude to more complex vocabulary or a higher reading level.
- D. **Passive Voice Detection:** The amount of passive voice in a text can have a significant effect on how easy it is to read and understand. Because they obscure the subject of the action, passive constructions make sentences less clear-cut and more complex. This may create ambiguity, which could make it difficult for readers to understand who is speaking and what the intended message is. Understanding the impact of passive voice usage aids in improving comprehension of the text's clarity and can direct efforts to improve its readability.
- E. **Unique Word Usage:** The metric evaluating vocabulary diversity serves as a gauge for the richness of language within a text. Unique words are words that are relatively uncommon, and indicate a more extensive vocabulary, which is frequently correlated with the increased complexity of a text.
- F. **Mirrored Letters:** The term "mirrored letters" describes characters that have the same symmetry when read backwards. For instance, the characters "b," "d," "p," and "q" are all reversed. Mirrored letters can be a sign of possible reader difficulty, especially for dyslexic readers.
- G. **Double Negatives:** When two negative elements are used in the same sentence, it's known as a double negative and can cause ambiguity or confusion. It's critical to recognize double negatives in order to ensure comprehension and clarity. The inclusion of double negatives is pivotal when evaluating the complexity and readability score of a sentence.
- H. **Block Letters:** When a word or string of words is written entirely in capital letters, it is referred to as a block letter. Because capital text is typically harder to read than mixed-case or lower-case writing, this may have an impact on readability.
- I. **Flesch Reading Ease Score:** This widely used readability assessment assigns scores based on sentence length and word length. Higher scores indicate that the material is easier to read.

A custom scoring logic was introduced for sample text paragraphs in this research in which the final score was calculated using an algorithm that takes into account multiple language aspects. Using the Textstat library the score is calculated out of 200 (where a score closer to 200 indicates higher readability, while a score farther from 200 suggests lower readability for the paragraph). Longer sentences, high syllable counts that are suggestive of complicated terms, and the frequent use of passive voice are penalized by the scoring system. It simultaneously favors text with less unique word usage and higher Flesch Reading Ease scores, suggesting that it is more readable. This grading system's design includes precisely calibrated penalties and weights to guarantee a thorough assessment. Crucially, the ability to freely modify these characteristics satisfies particular needs or preferences,

enabling a customized method of text evaluation based on distinct standards. Testing was done on the code to determine the readability index of paragraphs at various grade levels, including at a university level, eighth grade level, and third grade level. The algorithm shows promise in meeting all criteria, offering readability scores for lengthy paragraphs (between 500 and 700 words). After readability index evaluation, the GPT-3 model is used in an effort to make paragraphs easier for dyslexics to understand. It turns out that simplifying the text requires using a number of strategies, such as changing certain words with synonyms. Manipulating language, subject-verb relationships, and overall sentence structure are all part of complexity reduction. The GPT-3 model was programmed by providing prompts instructing the model to make the content suitable for a lower-grade comprehension level, limiting word variation, and taking into account other requirements of the project and tested with several prompts to get the desired and precise output so as to improve the readability of the output text for dyslexics.

4. RESULTS

The results of the study show that the readability index (which incorporated the Flesch Reading Ease Score in addition to specific textual parameters related to dyslexics) scores more readable text with a higher score, as it is designed to. Additionally, the GPT-3 model is able to effectively simplify texts to make them more readable, which was tested using the readability index and via a study done with dyslexic students. Table. 1 shows the readability score generated by the custom scoring model for different paragraphs. A rise in the readability score signifies a favorable trend towards increased accessibility of text for dyslexic individuals. The higher score indicates that attempts have been made to improve the text's readability, making it thereby easier to understand for dyslexic readers. This was tested using paragraphs with pre-determined complexity (at a grade school level and a university level). For the paragraphs at university-level reading difficulty there was a lower score as opposed to the paragraphs at a grade school level. This validated the algorithm for the index. Thereby, this readability index was also used to test the effectiveness of the GPT-3 model made to simplify text for dyslexic readers. Table. 1 shows the scores for the input paragraphs (unsimplified text) and the scores for the output paragraphs (simplified text):

Serial number	Input paragraph Readability Index	Output paragraph Readability Index
1	119.86399999999999	162.1055849403122
2	90.36700734618917	145.0653106497501
3	88.66666117969821	147.36847242512783
4	107.45577469135803	170.21784252960444
5	119.77145069393718	135.29660720486112
6	106.88775	160.5872677125272
7	113.69824691358025	113.55103537981269
8	122.44699999999999	144.9111405895692
9	94.7161720226843	152.971

10	92.18437953599047	131.72263118547525
11	67.0858888888889	107.03795770857363
12	112.59	139.6500896639883
13	113.7070826446281	140.50975
14	96.7364375	164.57164814395657
15	95.47580380499406	155.61928470298554

Table 1. Readability Index for different paragraphs

As the complexity level of the text decreases, the readability score increases after applying the GPT-3 model to simplify the language of the original text. This indicates that the GPT-3 model is effectively able to simplify texts according to the readability index. The text optimization approach significantly improved the reading abilities of dyslexic kids, according to the study's findings. In order to validate the GPT-3 model's effectiveness for dyslexic readers, a study was conducted where dyslexic participants were asked to read the input and output paragraphs so their reading performance could be measured and compared using Microsoft Team's Reading Progress feature. A depiction of the results based on a sample of twenty dyslexic adolescents who participated is shown in Fig. 1. Notably, compared to the input text, there was a significant improvement in overall reading ability when exposed to the simplified text created by the model i.e. the output text. The average number of accurate words read per minute in the unsimplified input text was 59.20. This number increased to 76.45 with the simplified output text, representing a gain of 17.25 words (around 29%). The accuracy rate of reading performance for the unsimplified input text was found to be 65%. After the input text was processed by the text optimization model, the accuracy rate increased by 10%, yielding a final rate of 75%. According to Microsoft Teams' Reading Progress, the accuracy rate takes into account a number of factors, such as mispronunciations, omissions, insertions, repeats, and self-corrections.

Participant No.	Correct words per minute (I/P)	Correct words per minute (O/P)	Accuracy Rate (I/P)	Accuracy Rate (O/P)	Difference in Accuracy Rate
1	50	48	73%	81%	8%
2	12	30	17%	37%	20%
3	46	58	69%	77%	8%
4	61	84	70%	81%	11%
5	51	69	51%	75%	24%
6	56	76	68%	78%	10%
7	41	45	61%	63%	2%
8	49	64	69%	84%	15%
9	14	75	32%	62%	30%
10	101	108	60%	81%	21%
11	75	69	62%	80%	18%
12	37	53	68%	74%	6%
13	43	49	61%	61%	0%
14	66	99	80%	79%	-1%
15	79	95	83%	85%	2%
16	54	49	71%	59%	-12%
17	104	159	68%	86%	18%
18	84	101	82%	82%	0%
19	70	105	69%	80%	11%
20	91	93	83%	85%	2%
AVERAGES	59.2	76.45	65%	75%	10%

Fig 1. Showing improvement in readability for Dyslexic children

Fig. 2 shows an example of the measured values for reading performance. To measure performance overall, five parameters were taken, which describe the complexity of the sentence for dyslexic children.



Fig. 2 shows the increased accuracy rate for sentences.

The data for the accuracy rate between the pre-model unsimplified input text and the post-model simplified output text can be visualized in Fig. 3.

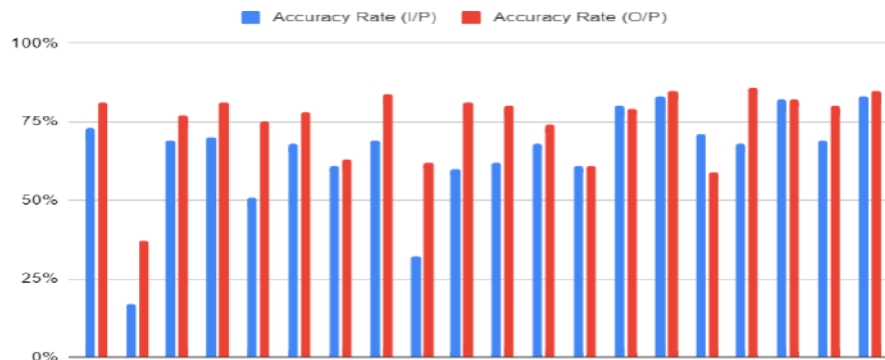


Fig 3. The difference in accuracy between previous and after model implementation

5. CONCLUSION

In conclusion, the two elements of this research, the readability index and the text optimization model, were created and implemented to gauge and improve the readability of text for a dyslexic audience. The former was a bespoke readability index curated based on specific factors that potentially evaluate the reading index of readers, particularly those with dyslexia; this includes the effect of the number of syllables, word and sentence count, sentence length, unique word usage, passive voice detection, mirrored letters, double negatives, and block letters. This, on the whole, successfully provided a readability score i.e. Flesch reading ease score. Additionally, a text optimization model using the GPT-3 architecture was created to make the text more readable for the dyslexic audience. This model, tested with dyslexic students, proved to be significantly effective - and provided a final accuracy rate in reading performance (measured using Microsoft Teams' Reading Progress) of 75%, a 10% increase from the initial rate.

ACKNOWLEDGMENT

I would like to express my gratitude to the mentor, Mr. Punit Kaushik of On My Own Technology Pvt. Ltd., for extending their help in carrying out this project.

REFERENCES

1. Rello Luz (2014) DysWebxia – A Text Accessibility Model for People with Dyslexia. <http://hdl.handle.net/10803/283113>
2. Rello L. , Baeza-Yates R. & Saggion H. (2013). The Impact of Lexical Simplification by Verbal Paraphrases for People with and Without Dyslexia. In: Gelbukh, A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2013. Lecture Notes in Computer Science, vol 7817. https://doi.org/10.1007/978-3-642-37256-8_41
3. Rello L., Ballesteros M. & Bigham J. (2016) A Spellchecker for Dyslexia. <https://doi.org/10.1145/2700648.2809850>
4. Fourney A., Morris M., Ali A. & Vonessen L. (2018). Assessing the Readability of Web Search Results for Searchers with Dyslexia. <https://doi.org/10.1145/3209978.3210072>
5. Zabkar J. , Urankar T. , Javornik K. & Babuder K. (2023). Identifying Reading Fluency in Pupils with and without Dyslexia Using a Machine Learning Model on Texts Assessed with a Readability Application. <https://doi.org/10.26529/cepsj.1367>
6. Chakraborty S., Nayeem M. & Ahmad W. (2021) Simple or Complex? Learning to Predict Readability of Bengali Texts. <https://doi.org/10.1609/aaai.v35i14.17495>
7. Baeza-Yates R., Dempere-Marco L., Saggion H & Rello L.. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. https://doi.org/10.1007/978-3-642-40498-6_15
8. Erlikhman G., Strother L., Barzakov I. & Caplovitz G. (2017). On the Legibility of Mirror-Reflected and Rotated Text. <https://doi.org/10.3390/sym9030028>
9. Ownby RL. (2005) Influence of Vocabulary and Sentence Complexity and Passive Voice on the Readability of Consumer-Oriented Mental Health Information on the Internet. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560876/>
10. Collins-Thompson K. (2014). Computational assessment of text readability: A survey of current and future research. <https://doi.org/10.1075/itl.165.2.01col>
11. Yusri ANB., Manap MHBA., Ismail MIZB. & Kamarozaman MMB. (2023) Speed Reading Tool Powered by Artificial Intelligence for Students with ADHD, Dyslexia, or Short Attention Span. <https://doi.org/10.48550/arXiv.2307.14544>