The Concept of Readability from a Librarian's Perspective

Alif Hossain
alifhossain1994@gmail.com

Swapna Banerjee
swapna.banerjee98@gmail.com

University of Calcutta
Department of Library and Information Science, Calcutta, India

Abstract

In our present society, there is a need for information that will help individuals in decision-making and developing personal policies and will help them to transform into better information-literate citizens. An information-literate citizen can help generate new information by combining or analyzing all the information he or she has gained. It would develop new theories, which would lead to the generation of new knowledge. A reader needs to understand the document to process the information that is presented in it. This is where the feature of readability comes into play. In layman's language, readability can be described as the ability of the written content to make the reader easily understand the information. This paper aims to provide a review of the articles that are related to readability. The review is done by searching for relevant articles on Google Scholar and analyzing them. Also, the paper aims to support readability studies in libraries as they can play a huge role in selecting appropriate books for users.

Keywords: information literacy, readability, readability formula, readability score, readability test, libraries, collection development.

Introduction

In the book The Principles of Readability,[1] William H. DuBay says that "readability is what makes some texts easier to read than others". "It is often confused with legibility, which concerns typeface and layout". A high readability grade can help the user understand the text very easily with less effort. There are various groups of users, ranging from doctors, engineers, and service technicians to academicians, politicians, students, and even farmers and labourers. Everyone needs information, which would make their lives much easier and make them information literate. So, to be information literate, one has to understand the information. Readability tests and grade levels help us differentiate between books with lucid language and those with complex language.

Librarians have to manage libraries and provide the required documents to their intended users. For this purpose, librarians have to acquire various types of documents to keep in their libraries. A librarian has to keep several ideas in mind when selecting a proper document, which might be a book, journal, newsletter, etc., to be kept in the library. In this context, librarians prefer user studies. Through user studies, the librarian understands the needs of the users. These needs would justify the purpose, relevance and demand for the documents to be kept in the library. Besides, the librarian also has to keep in mind that the quality of the document is not

---

[1] William H. DuBay, *The Principles of Readability* (N.p.: ERIC Clearinghouse, 2004), 3.

compromised and that it is available in ample quantity to be served to the users. The book selection techniques of different libraries are different for selecting proper documents. But if the factor of analyzing readability is introduced for the selection of books, it can help users in various ways. Certain users can be provided with certain reading levels, increasing their understanding ability, saving them time to select appropriate resources, and also allowing them to set academic standards for education and work towards spreading information literacy.

There are various types of readability indexes and grades. The most important readability formulas that help in identifying the difficulty of language for a particular document are the Flesch Formulas (Flesch Reading Ease, Flesch-Kincaid Grade Level), the Dale Chall Readability Formula, the Gunning Fog Index, and the McLaughlin SMOG Formula. Other readability formulas are the Fry readability formula, the FORCAST formula, and the Golub syntactic density score. For advanced readability, there are the John Bormuth Formulae, the Lexile Framework, and the ATOS readability formula. More usage of these formulas in texts would make them much easier to understand, which would eventually help the individual become information literate. This would also help to promote readability literacy and encourage more individuals to use the information for their benefit. In the current scenario, the concept of automated readability assessment typically consists of:

I) A training corpus of each text,
II) A set of linguistic features,
III) A machine-learning model using the computed linguistic feature value.

In the year 2012, in an article titled "On Improving the Accuracy of Readability Classification Using Insights from Second Language Classification"[2] Vajjala and Meurers of the University of Tubingen introduced AI-based readability for investigating the problem of readability assessment using a range of lexical and syntactic features and studying their impact on predicting the grade level of texts.

Purpose of the Paper

The primary objective of reviewing the literature, which is related to readability, is to understand which readability formula is best suited for analyzing the vocabulary of a particular text so that modifications can be made to serve the user.

Methodology

The approach to reviewing the literature on the readability of the text is done by studying papers related to documents. Many studies have been published regarding the readability of brochures, journals, scientific texts, books, manuals, pamphlets, and even newspapers. The process was done using simple search techniques offered by Google Scholar and trying to find and understand the articles that had been analyzed using the most popular readability formulas.

A librarian's role in the selection of books

The selection of books is crucial in librarianship. A librarian must know about the selection of books and the criteria by which users can be provided with authentic types of information that are understandable to them. Librarians use various types of book selection techniques in every type of library. For a particular type of library, librarians have to look into various factors while

---

[2] Sowmya Vajjala and Detmar Meurers, "On improving the accuracy of readability classification using insights from second language acquisition", *The 7th Workshop on the Innovative Use of NLP for Building Educational Applications* (2012): 163–173, accessed January 25th 2019, https://aclanthology.org/W12-2019.pdf.

ordering and purchasing books as well as other documents. For example, if a public library's librarian needs to order books, he or she has to put forward the general interest of the people residing in the locality near the library. He or she has to understand the language of books that are suitable for users who speak and understand a certain language. Alongside this, an entire goal is set up by the library to educate the masses and make most of the users who are using the library's information literate. These factors are to be kept in mind by a librarian while selecting books and also to keep them covered under the budgetary funds that they receive from the organization. A librarian in an academic institution's library, which may be a school, college, or university, has a similar role. The librarian has to review various factors while selecting books and other documents. For school libraries, the selection criteria solely depend on the aims of the library, where the prime focus is on the overall development of students in reaching their curriculum. It also has to be kept in mind that students develop an interest in reading and observing information that is not only curriculum-oriented but also in the information and skills they develop, which would help in the short term. For college and university academic libraries, the librarians have to keep a few things in mind while selecting documents. These institutions provide specialization in various subjects, so books, journals, and other documents that are to be purchased and supplied to the students as well as the research scholars need to be appropriate, authentic, concise, and updated. In these situations, librarians at higher education institutions get the assistance of the library advisory committee. Other factors that the librarian should keep in mind while selecting documents for the library are the currency of the document, the appropriateness of the document, the resources that can be accessed both physically and virtually, the presentation of a few points on various subjects, as well as proper collection development in the libraries. Academic libraries are not only used by students but also by faculty and professionals who work in academic institutions. Thus, the standards of the libraries of higher education institutions should provide documents with diverse depths of specialization in specific field formats and are up-to-date to support not only the function of teaching and learning but also serve as a helping tool for further research.

Review of Literature: Readability of a Document

William H. DuBay, in his book *Classic Readability Studies*,[3] explains the background of readability theories. In 1880, Lucius Adelno Sherman, an English literature professor at the University of Nebraska, introduced a new method of literary criticism. He believed literature was a moral and spiritual edification tool and advocated an objective approach. Sherman's approach was effective in allowing students without a taste for reading to appreciate and enjoy poetry. This method, if tried intelligently and fairly, can reveal those who believe they have no taste for the best literature and make those who have never found poetry aware of its power.

He made the statistical decision to investigate this and started by counting the average sentence length per 100 sentence intervals. He demonstrated in his book how, on average, sentences have grown shorter through time:

Pre-Elizabethan periods had sentences with 50 words, Elizabethan times had sentences with 45 words, Victorian times had sentences with 29 words, and Sherman's time had sentences with 23 words.

Sherman's work influenced reading research by suggesting statistical analysis, shorter sentences, and efficient spoken language over time. Sherman discovered that average sentence lengths across writers were stable, and that influenced readability predictions using text

---

3 William H. DuBay, ed., "1893 – L.A. Sherman: The Analytics of Literature", in *Classic Readability Studies* (ERIC Clearinghouse, 2007), accessed March 3rd 2012, https://files.eric.ed.gov/fulltext/ED506404.pdf.

samples. The sentences become shorter, simpler, and less abstract over time, influenced by spoken language. He believed that literary English followed conventional spoken English conventions. Sherman emphasizes the importance of involving the reader in the universally best style, considering their expectations, and avoiding book words. The style should be correct, not stiff, and avoid vulgarity.

An article from 1968 in the *Encyclopaedia of Library and Information Science* named "Rubakin Nicholai Alexanderovic"[4] said that Rubakin published an article that discussed 10,000 texts written by common people. He sorted 15,000 words from those texts he thought most people would understand. He found that unfamiliar words and long sentences were the main reasons that obstructed a person's understanding. DuBay, in his book *The Classic Readability Studies*, also said that 1921 was the year when E. L. Thorndike published his book *Teacher's Word Book*,[5] which had an extensive listing of English word frequency. It provided teachers with an objective means for measuring the difficulty of words and texts. This book is also the foundation stone for the 1923 ability formulas. In 1923, L. Bertha, A. Lively, and Sidney L. Pressey were very concerned with the selection of scientific textbooks for junior high school. The textbooks for science were difficult to read and understand due to a huge number of technical terms, which led to the teachers teaching the technical vocabulary to the students all day. That's where they published their study, "A Method for Measuring the Vocabulary Burden of Textbooks",[6] in the *Journal of Educational Administration and Supervision*. Following in the footsteps of Lively and Pressey, Mabel Vogel and Carleton Washburne came up with "The Winnetka Formula"[7] in 1928, which is considered one of the most important studies of readability. Vogel and Washburne were the first to study the structural characteristics of the text and also use a criterion based on an empirical evaluation of the text. The Winnetka formula was set against 700 books that had been suggested by 25 out of 37,000 children as the ones they had read and liked. They also had the mean reading score for the books suggested by the children, which they used to scale the difficulty measure. The era of the 1930s gave new directions to readability. Douglas Waples and Ralph W. Tyler (1931) published *What People Want to Read About*,[8] which is a comprehensive study of adult reading interests that was formulated by interviewing people and dividing them into 107 groups. Ralph Ojemann (1934)[9] invented a method of assessing the difficulty of adult-parent education materials. Ralph Taylor and Edgar Dale (1934)[10] published their readability formula and their first study on adult readability formulas. William S. Gray and Bernice Leary (1935)[11] published *What Makes a Book Readable*. In 1931, W. W. Patty and W.I. Painter[12] discovered a formula for selecting the correct textbooks for high school. Irving Lodge (1939)[13] demonstrated that a new combination of vari-

---

[4] Marianna Tax Choldin, "Rubakin, Nikolai Aleksandrovič", in *Encyclopedia of Library and Information Science. Vol. 26, Role indicators to scientific literature* (New York: Dekker, 1979), 178–179.

[5] William H. DuBay, ed., "1921 – E. L. Thorndike: The Teachers' Word Book", in *Classic Readability Studies* (ERIC Clearinghouse, 2007).

[6] Bertha A. Lively and Sidney L. Pressey, "A method for measuring the vocabulary burden of textbooks", *Educational administration and supervision* Vol. 9, no. 7 (1923): 389–398.

[7] Mabel Vogel and Carleton Washbmurne, "A Year of Winnetka Research", *The Journal of Educational Research* Vol. 17, no. 2 (1928): 90–101.

[8] Douglas Waples and Ralph W. Tayler, *What People Want to Read About: a study of group interests and a survey of problems in adults* (Chicago: American Library Association; The University of Chicago Press, 1931).

[9] William H. DuBay, ed., "Ralph Ojemann: The Difficulty of Adult Materials", in *Classic Readability Studies*, 27.

[10] William H. DuBay, ed., "Dale and Tyler: Adults of Limited Reading Ability", in *Classic Readability Studies*, 28.

[11] William S. Gray and Bernice E. Leary, *What makes a book readable, with special reference to adults of limited reading ability* (Chicago, Ill.: University of Chicago Press, 1935).

[12] Willard W. Patty and W. I. Painter, "A technique for measuring the vocabulary burden of textbooks", *The Journal of Educational Research* Vol. 24, no. 2 (1931): 127–134.

[13] William H. DuBay, ed., "1944 – The Lorge Readability Index", in *Classic Readability Studies*, 44.

ables gave predictions of higher accuracy than the Gary-Leary formula by publishing an article titled "Predicting Reading Difficulty of Selection for Children". The most popular readability formulas emerged in the 1940s. In 1943, Rudolf Flesch[14] came up with the formula for the Flesch Reading Ease (FRE) score after studying the vocabulary provided by the early researchers. Later on, the Reading Ease formula was modified in 1976 by a study ordered by the U.S. Navy to yield a grade-level score. This widely used formula is also referred to as the Flesch-Kincaid formula, the Flesch Grade-Scale formula, or the Kincaid formula.[15] In 1948, Edgar Dale and Jeanne S. Chall were inspired by the Flesch formula and came up with the Dale – Chall readability formula.[16] In 1952, a book called *A Technique of Clear Writing*[17] by Robert Gunning came up with the Gunning Fog Index formula. This formula can be considered the simplest formula to apply to a textual work. Edward Fry (1963, 1968)[18] was working as a Fullbright scholar in Uganda, where he was preparing teachers to teach English as a second language. He developed a graph that had an average number of sentences on the x-axis and syllables per hundred words on the y-axis. This came to be known as the Fry readability graph. This work was cited in the book *Smart Language: Readers, Readability, and Grading of Text* by William H. Dubay. G. Harry McLaughlin published a study called "SMOG Grading: A New Readability Formula",[19] which provided the formula for the SMOG grade readability index. The FORCAST formula[20] (1973) by the US Military and the Golub Syntactic Density Score (1974)[21] by Lester Golub was also part of the readability text analysis in this period for Army documents and syntactic units of texts, respectively. In the year 1975, an article by Meri Coleman and T.L. Liau titled "Computer Readability Formula Designed for Machine Scoring"[22] marked the formulation of the Coleman Liau Readability Index. The advanced readability formulas that came in the late '80s and early 2000s were the Automated Readability Index,[23] the John Bormuth Formula (1981),[24] the Lexile Framework (1988),[25] and the ATOS Readability Formula[26] for books.

Discussion

The study of articles' readability revealed that the Flesch formula, followed by the Gunning Fog Index, the Dale Chall Readability Index, and the SMOG grade formula, are the most popular among them, based on average sentence length and syllable count per 100 words. The formula for Flesch Reading Ease is defined as:

---

[14] Rudolf Flesch, *Marks of readable style: a study in adult education* (New York: Teachers College, Columbia University, 1943).

[15] William H. DuBay, ed., "1948 – The Flesch Formulas", in *Classic Readability Studies*, 97.

[16] Edgar Dale and Jeanne S. Chall, "A formula for predicting readability", *Educational Research Bulletin* Vol. 27, no 1 (1948): 37–54.

[17] Robert Gunning, *Technique of clear writing* (New York: McGraw-Hill, 1952).

[18] William H. DuBay, "The Fry Readability Graph", in *Smart Language: Readers, Readability, and the Grading of Text* (Costa Mesa, California: Impact Information, 2007), 84.

[19] G. Harry Mc Laughlin, "SMOG grading-a new readability formula", *Journal of Reading* Vol. 12, no. 8 (1969): 639–646, accessed April 12th 2021, https://ogg.osu.edu/media/documents/health_lit/WRRSMOG_Readability_Formula_G._Harry_McLaughlin__1969_.pdf.

[20] William H. DuBay, "The FORCAST Formula", in *Smart Language: Readers, Readability, and the Grading of Text*, 87.

[21] Lester S. Golub and Carole Kidder, "Syntactic density and the computer", *Elementary English* Vol. 51, no. 8 (1974): 1128–1131.

[22] Meri Coleman and Ta Lin Liau, "A computer readability formula designed for machine scoring", *Journal of Applied Psychology* Vol. 60, no. 2 (1975): 283.

[23] William H. DuBay, "The Army's Automated Readability Index (ARI)", in *Smart Language: Readers, Readability, and the Grading of Text*, 90.

[24] William H. DuBay, "The Bormuth Studies", in *Smart Language: Readers, Readability, and the Grading of Text*, 81–83.

[25] William H. DuBay, "Lexile Framework", in *Smart Language: Readers, Readability, and the Grading of Text*, 95.

[26] William H. DuBay, "The Bormuth Studies", in *Smart Language: Readers, Readability, and the Grading of Text*, 81.

FRE = 206.835 -1.015[total words/ total sentences]- 84.6 [total syllabus/total words]

and the formula for the Flesch-Kincaid grade level is stated as:

F_KGL = 0.39* [total words/ total sentences] + 11.8 [total syllables/ total words] -15.59

Two important studies related to the usage of the Flesch reading formula were found in the Shodhganga repository. One by N. Murlimohan (2014)[27] and the other by Bidyarani Asem (2016),[28] respectively. Both the research theses represent the Department of Journalism and Mass Communication in two of the universities in India. The research thesis, which was published in 2016, had Flesch Reading Ease as a preliminarily determined formula, and it was compared alongside the Gunning Fog Index and the SMOG Index, and the readability was calculated. The study found that the readability of hard news between the Hindu and Times of India does not differ significantly, while the readability of soft news does. Flesch formulas are the primary readability formula used in papers, but other formulas are also used to represent comparative and distinctive differences in a single text, with most articles comparing and presenting multiple formulas.

The next most commonly used formula is the Dale Chall readability formula, which was formulated by Edgar Dale and Jeanne S. Chall. This formula depends on the number of difficult words. The formula stands as follows:

DC Score = (difficult words/words) + 0.0496 (words/ sentences)

where 4.9 is lower with an understandability of a 4th grade and 9.9 is the highest with an understandability of a college graduate grade. Marie Burkhead and Greg Ulferts[29] (1977) took a text and sampled 100 words from every 10 pages, compared it with Dale's list of 3,000 difficult words, and tried to find the difference in every 10-page interval. This work was done on 48 textbooks. The study found no significant difference in mean scores between samples taken at different intervals in the book, despite careful examination of the specific material. Another article by Pontus Plave et al.[30] (2017) reports on the readability of scientific research documents. The study analyzed 709,577 article abstracts from PubMed and 123 highly cited journals in 12 biomedical and life sciences fields to identify trends in the readability of scientific documents. The journals from which the articles were taken were Nature, Science, NEJM, The Lancet, PNAS, and JAMA, ranging from 1881 to 2015. They used two methods of readability, i.e., the Flesch Reading Ease and the new Dale Chall Readability Formula. The results of the tests came out as follows: The Flesch Reading Ease formula significantly decreased the readability of documents, while the New Dale Chall readability formula significantly increased their readability. The next thing they tried was to establish a relationship between the components of readability metrics and the year of publication. The component of the Flesch formula, i.e., the average number of syllables in each word, and the component of the Dale Chall formula, i.e. the number of difficult words showed an increasing graph through the years. Sentence length, a key component of

[27] N. Murali Mohana, *Measuring the readability of newspapers a comparative study of two national English dailies*, 2016., accessed April 12th 2021, http://hdl.handle.net/10603/145898.

[28] Asem Bidyarani, *Measuring readability a message analysis of newspaper editorials*, accessed December 31st 2014, http://hdl.handle.net/10603/82520.

[29] Marie Burkhead and Greg Ulferts, "Sample frequency in the application of Dale-Chall readability formula", *Journal of Reading Behavior* Vol. 9, no. 3 (1977): 287–290.

[30] Pontus Plavén-Sigray et al., "The readability of scientific texts is decreasing over time", *Elife* 6 (2017): e27725, https://doi.org/10.7554/eLife.27725, Sep 5, 2017.

112 Читалиште 44 (мај 2024)

ОДРАЗ

Hossain A. at al. "The concept of readability from a librarian's perspective", 106–118

both the Flesch formula and the Dale Chall formula, experienced a gradual increase post-1960 when most abstracts were published. The study found that the complexity of scientific writing increases over time, with a decreasing readability trend not specific to any particular field, despite differences in magnitude among the 12 selected fields. Furthermore, 2 of the 123 journals showed a clear increase in Flesch readability across time. There could be several explanations for the observed trend. The researchers found that the increasing number of co-authors over time can lead to an increase in scientific jargon, with 2,949 most common words from 12,000 abstracts not included in the NDC (New Dale-Chall) common word list, similar to the science-specific common word list. These words were removed, and 2,138 common words that matched the New Dale Chall list of difficult words were listed. Here it was shown that New Dale Chall's common words usage decreased over the year, while science-specific words, or technical terms, increased, partially affecting readability.

Next in the row is the Gunning Fog Index. The articles with the Gunning Fog Index are mostly used alongside the Flesch formulas and the SMOG index readability formula. The Gunning Fog Index comes with the term 'hard words' which means words with more than two syllables. The formula stands as:

GFI = 0.4 [(words/sentence)+ 100 (hard words/words)]

An article by L. Timana, D. Lozano, and J. Garcia (2020)[31] showed the usage of the Gunning Fog Index not only for English texts but also for Spanish texts. The researchers converted the corpus into modified texts to ensure its readability. Finally, after the application of the formula, they calculated the result and later concluded that the texts were appropriate for understanding individuals with that particular grade of readability.

The most sought-after Gunning Fog readability index is the SMOG grade readability index, developed by G. Harry McLaughlin[32] in 1969. The SMOG grade readability formula was mostly used in safety manuals, especially health-related manuals. SMOG is the abbreviation for "Simple Measure of Gobbledygook". The vocabulary of polysyllables comes into play in this account. The formula of SMOG grade readability stands as:

SMOG Grading = 3+√(Polly syllable count * 30/number of sentences)+ 3.1291

Margaret Comerford Freda[33] (2005) showed the purpose of evaluating the readability of patient education brochures. 74 brochures were analyzed using the SMOG grade formula and the Flesch-Kincaid formula. The SMOG grade readability formula demonstrated readability superior to the Flesch formula, indicating its perfect suitability for users. Dolnicar and Chappel[34] (2015) applied the Flesch reading formula and the SMOG formula in tourism journals (*Annals of Tourism Research*, *Journal of Tourism Research* and *Tourism Management*). The result showed that readability had decreased from 21 in the year 1993 to 20.6 in the year 2003 to 15.5 in the year 2013. Here, the score obtained by the SMOG grade formula was higher than the score obtained using the Flesch formula.

---

[31] Timaná Rodriguez et al., "Software to Determine the Readability of Written Documents by Implementing a Variation of the Gunning Fog Index Using the Google Linguistic Corpus", in *International Conference on Applied Technologies* (Springer: Cham, 2019), 409–420.

[32] Mc Laughlin, "SMOG grading-a new readability formula": 639–646.

[33] Margaret Comerford Freda, "The readability of American Academy of Pediatrics patient education brochures", *Journal of Pediatric Health Care* Vol. 19, no. 3 (2005): 151–156.

[34] Sara Dolnicar and Alexander Chapple, "The readability of articles in tourism journals", *Annals of Tourism Research* Vol. 52 (2015): 161–166.

ОДРАЗ   Читалиште 44 (мај 2024)   **113**

Hossain A. at al. "The concept of readability from a librarian's perspective", 106–118

The formula of the Coleman Liau Index is also used in many of the documents, mostly along with other formulas. Coleman and Liau (1975) formulated it and published it in the paper "A Computer Readability Formula Designed for Machine Scoring".[35] The formula is represented as:

CLI= 0.0588L-0.296S-15.8

where L is the average number of letters per 100 words and S represents the average number of sentences per 100 words. Hosseinzadeh et al. (2021),[36] where the objective of the study was to evaluate the readability of current online patient information regarding Dupuytren's contracture using three search engines: Yahoo, Google, and Bing. The study analyzed 30 websites from three engines using Coleman-Liau Index, Flesch Reading Ease, Flesch Kincaid Grade Level, and SMOG Grade Readability Index, revealing average Flesch reading ease of 46.4 Flesch Kincaid Grade Level: 10.2, Coleman-Liau Index: 13.1, and SMOG Grade Index: 14.4. The study revealed lower readability levels than previous publications, except for the Coleman Liau Index, suggesting a larger internet body on Dupuytren's Contractures.

In the modern scenario, we can see that the Automated Readability Index,[37] developed by E.A. Smith and R. J. Senter in 1967, is a widely used readability formula alongside other popular tools. The Automated Readability Index, developed after analyzing the challenges of US Air Force technical manuals, offers a quick and cost-effective method for assessing the readability of Air Force documents. The formula for the Automated Readability Index stands as follows:

GL = 0.50 (w/s) + 4.71 (s/w) -21.43

where "s" is defined as the number of characters, in a word. This formula differs from others by counting characters instead of syllables, indicating that an increase in characters in a word makes it harder.

Automated readability index usage has grown in non-English textbook readability studies, selecting and testing various texts through a corpus of texts. Jorge Morato et al. (2021),[38] discuss their proposal of an automated readability model that is inspired by the converted model of the Flesch Formula for the Spanish texts, i.e., the Fernandez-Huerta Model (1959). From the Spanish e-government website, they gathered 133 documents from two websites and categorized them into easy and difficult reading groups. Readability, as well as precision, were calculated using two formulas: the Fernandez-Huerta Model and the new proposed model "μ". The results show that the proposed model "μ" gives much more accurate readability as well as precision to difficult documents. Grose (2021)[39] discussed the readability of the Internet-based manuals found on the websites for patients to know about nasal septoplasty. Six readability tests were conducted on 400 documents, removing 249 duplicate articles and 66 web pages, based on Flesch Kincaid grade level, reading ease, SMOG index, Coleman Liau, Gunning Fog, and automated readability test. The study analyzed 85 patient education materials from six sources: academic institutions, private medical clinics, professional organizations, government websites, medical

[35] Coleman and Liau, "A computer readability formula designed for machine scoring": 283.

[36] Shayan Hosseinzadeh, et al., "Dupuytren's Contracture: The Readability of Online Information", *Journal of Patient Experience* no. 8 (2021, doi.org/10.1177/23743735211056431.

[37] Edgar A. Smith and R. J. Senter, "Automated readability index", *Aerospace Medical Research Laboratories* Vol. 66, no. 220 (1967): 8.

[38] Jorge Morato et al., "Automated readability assessment for Spanish e-government information", *Journal of Information Systems Engineering and Management* Vol. 6, no. 2 (2021): em0137, https://doi.org/10.29333/jisem/9620.

[39] Elysia M. Grose, Connor P. Holmes and John M. Lee, "Readability and quality assessment of internet-based patient education materials related to nasal septoplasty", *Journal of Otolaryngology-Head & Neck Surgery* Vol. 50, no. 1 (2021): 1–8, doi.org/10.1186/s40463-021-00507-z.

information websites like WebMD, and other miscellaneous sources (i.e., *Wikipedia*). The results of readability show that all except the Automated Readability Index had higher grade levels. Gayatri Venugopal (2021)[40] focused readability tests on one of the most spoken vernacular languages in India, i.e., Hindi where a corpus has been taken to test for readability and analyze the complexity of words. The study analyzes complex words in Hindi sentences using Human Intelligence Task (HIT) variables and widely adopted readability measures, converting the Flesch readability formula according to Hindi literature using AI.

A few articles were subjected to other various readability tests, like the Automated Readability Index (J. Peter Kincaid, 1975)[41] and the AI-based readability tests (Vajjala, Meurs, 2008).[42] Luo Si and Jamie Callan (2001)[43] introduced readability tests of web pages through a statistical language model called the EM algorithm combined with the Flesch formula.

Findings

The study aimed to examine readability from a librarian's perspective, resulting in various findings. The first thing that was noticed was that the readability papers were mostly on documents which are basically for daily instructions like manuals, brochures, pamphlets, notices, etc. Secondly, it was seen that no single readability formula was used in many papers. A single text had been run through various popular readability formulas, and the difference was identified and the distinctive information concluded. Thirdly, it was observed that out of the documents that had been studied to find the literature of the work, most of them were concerned with the medical health sector's information manuals for patients. Fourthly, it was recognized that most of the documents had an unsatisfactory level of readability, and it was suggested that organizations ought to work on the documents' vocabulary to increase readability.

If librarians reviewed readability as a factor while selecting books

An important aspect to consider while evaluating the readability of a book is the intended audience. A librarian must have enough understanding of the language level of the book. In a book that is written at a level that is too difficult for the intended audience, the readers might be discouraged from seeking information from it. On the other hand, if the readability of a book suits the users, they will be engaged in the book or document and will also have an urge and interest to see further knowledge. Other factors that can affect the readability of a book include the length and complexity of the language as well as the layout and formatting of the text and the subject matter being covered. For example, a fiction book with a difficult plot will have much less engagement than a book with a much more straightforward plot. Librarians consider readability when selecting books for their library to ensure that the books they choose will be accessible and enjoyable for their patrons. There are several factors that librarians consider when evaluating the readability of a book. They are:

1. Lexile Measure: The text's complexity is a standardized measure determined by the number of words, sentence length, and word complexity.

---

[40] Gayatri Venugopal, Dhanya Pramod, and Jatinderkumar R. Saini, "Analyzing Complex Words in Hindi using Parameters of Classical Readability Formulae (Part 1)", *Computer Science* Vol. 29, no. 3 (87) (2021): 366–387, https://doaj.org/article/0574527ac7f14b13a4aacb5da4106a04.

[41] J. Peter Kincaid et al., *Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for navy enlisted personnel* (Springfield, Virginia: National Technical Information Service, 1975).

[42] Vajjala and Meurers, "On improving the accuracy of readability classification using insights from second language acquisition": 163–173.

[43] Luo Si and Jamie Callan, "A statistical model for scientific readability" in *Proceedings of the tenth International Conference on Information and Knowledge Management* (2001): 574–576, doi/10.1145/502585.502695.

ОДРАЗ   Читалиште 44 (мај 2024)   **115**

Hossain A. at al. "The concept of readability from a librarian's perspective", 106–118

2. Grade Level: Librarians may consider the recommended grade level for a book as it is indicated by the publisher or an educational organization.
3. Interest Level: Librarians consider patrons' ages and interests when selecting books for specific age groups or interests.
4. Style and tone: Librarians may consider the writing style and tone of a book to ensure that it is engaging and appropriate for the intended audience.
5. Layout and design: The readability of a book is influenced by its layout and design, which librarians assess through factors like font size, spacing, and image use.

How can readability be used for the books in the library?

While studying various articles related to the readability of books that can benefit libraries, we came across an article by C. Metoyer-Duran (1993)[44] regarding the readability of papers that are accepted or rejected, and also published for the years 1990 and 1991 for college and research libraries. Here also, the Gunning Fog Index and Flesch Kincaid readability formulas were used. The study selected 271 papers for a supplement, revealing that published papers have better readability than rejected or accepted ones, potentially improving library browsing and searchability.

Another article by Elaine Robinson and David McMenemy (2020)[45], came across the readability of Acceptable Use Policies for public libraries. 200 out of 206 AUPs were analyzed using four readability tests: the Flesch Reading Ease, Coleman-Liau Index, Gunning Fog Index, and SMOG Grade. On the Flesch Reading Ease Score, only 5.5% of AUPs achieved a standard readability level or above (60+), and 8% achieved a very high level of difficulty, similar to academic writing. Similarly, in SMOG, only 7.5% of 200 AUP reached the recommended level of 10. Similarly, few AUPs achieved levels recommended for general audiences on either the Gunning Fog Index (11.5%) or the Coleman-Liau Index (2%). This could help create more user-friendly and readable public library policies.

Readability in libraries can play a huge role in selecting books that are more appropriate for the users. Users need books they can relatively easily read and understand. An article presented by Godwin Shoki (2007)[46] at the 73rd IFLA conference said that readability needs to be considered while selecting a book. The study investigates book selection methods among 100 librarians from 12 academic libraries in South Nigeria, finding that only 10% consider readability a criterion.

Conclusion

The study of articles on readability of a document has led to the conclusion that readability tests are essential for ease of reading and understanding a vocabulary, as low readability can lead to improper use of information.

Libraries can benefit from readability in various ways: it could/would help professionals in libraries make a better selection of books for library collections; it can be considered in the collection development in libraries; it can help not only libraries but the whole academic institution

---

[44] Cheryl Metoyer-Duran, "The readability of published, accepted, and rejected papers appearing in College & Research Libraries", *College & Research Libraries* Vol. 54, no. 6 (1993): 517–526.

[45] Elaine Robinson and David McMenemy, "To be understood as to understand': a readability analysis of public library acceptable use policies", *Journal of Librarianship and Information Science* Vol. 52, issue 3 (2020): 713–725, doi. org/10.1177/0961000619871598.

[46] Godwin Shoki, "Readability as consideration for book selection criterion in book selection practices in some academic libraries in Nigeria", paper presented at World Library and Information Congress: 73rd Ifla General Conference and Council, Durban, South Africa, 2007, accessed June 28th 2007, https://archive.ifla.org/IV/ifla73/papers/132-Shoki-en.pdf.

116 Читалиште 44 (мај 2024)   ОДРАЗ

Hossain A. at al. "The concept of readability from a librarian's perspective", 106–118

if they could design the academic curriculum based on readability which would help students in accomplishing their objectives in the journey of their academic pursuit; and, a better selection of appropriate books would finally lead to a higher level of information literacy.

References:

1. Bidyarani, Asem. *Measuring readability a message analysis of newspaper editorials*. Accessed December 31st 2014. http://hdl.handle.net/10603/82520.
2. Burkhead, Marie and Greg Ulferts. "Sample frequency in the application of Dale-Chall readability formula". *Journal of Reading Behavior* Vol. 9, no. 3 (1977): 287–290.
3. Choldin, Marianna Tax. "Rubakin, Nikolai Aleksandrovič". In *Encyclopedia of Library and Information Science. Vol. 26, Role indicators to scientific literature*, 178–179. New York: Dekker, 1979.
4. Coleman, Meri and Ta Lin Liau. "A computer readability formula designed for machine scoring". *Journal of Applied Psychology* Vol. 60, no. 2 (1975): 283–284.
5. Comerford, Freda Margaret. "The readability of American Academy of Pediatrics patient education brochures". *Journal of Pediatric Health Care* Vol. 19, no. 3 (2005): 151–156.
6. Dale, Edgar and Jeanne S. Chall. "A formula for predicting readability". *Educational Research Bulletin* Vol. 27, no 1 (1948): 37–54.
7. Dolnicar, Sara and Alexander Chapple. "The readability of articles in tourism journals". *Annals of Tourism Research* Vol. 52 (2015): 161–166.
8. DuBay, William H. ed. *Classic Readability Studies*. ERIC Clearinghouse, 2007. Accessed March 3rd 2021. https://files.eric.ed.gov/fulltext/ED506404.pdf.
9. DuBay, William H. *Smart Language: Readers, Readability, and the Grading of Text*. Costa Mesa, California: Impact Information, 2007.
10. DuBay, William H. *The Principles of Readability*. N.p.: ERIC Clearinghouse, 2004. Accessed March 3rd 2021. https://files.eric.ed.gov/fulltext/ED490073.pdf.
11. Flesch, Rudolf. *Marks of readable style: a study in adult education*. New York: Teachers College, Columbia University, 1943.
12. Golub, Lester S. and Carole Kidder. "Syntactic density and the computer". *Elementary English* Vol. 51, no. 8 (1974): 1128–1131.
13. Gray, William S. and Bernice E. Leary. *What makes a book readable, with special reference to adults of limited reading ability*. Chicago, Ill.: University of Chicago Press, 1935.
14. Grose, Elysia M., Connor P. Holmes and John M. Lee. "Readability and quality assessment of internet-based patient education materials related to nasal septoplasty". *Journal of Otolaryngology – Head & Neck Surgery* Vol. 50, no. 1 (2021): 1–8. doi.org/10.1186/s40463-021-00507-z.
15. Gunning, Robert. *Technique of clear writing*. New York: McGraw-Hill, 1952.
16. Hosseinzadeh, Shayan, Phillp Blazar, Brandon E. Earp and Dafang Zhang. "Dupuytren's Contracture: The Readability of Online Information". *Journal of Patient Experience* no. 8 (2021). doi.org/10.1177/23743735211056431.
17. Kincaid, J. Peter et al. *Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for navy enlisted personnel*. Springfield, Virginia: National Technical Information Service, 1975.
18. Lively, Bertha A. and Sidney L. Pressey. "A method for measuring the vocabulary burden of textbooks". *Educational Administration and Supervision* Vol. 9, no. 7 (1923): 389–398.
19. Mc Laughlin, G. Harry. "SMOG grading-a new readability formula". *Journal of Reading* Vol. 12, no. 8 (1969): 639–646. Accessed April 12th 2021. https://ogg.osu.edu/media/documents/health_lit/WRRSMOG_Readability_Formula_G._Harry_McLaughlin__1969_.pdf.

ОДРАЗ Читалиште 44 (мај 2024) **117**

Hossain A. at al. "The concept of readability from a librarian's perspective", 106–118

20. Metoyer-Duran, Cheryl. "The readability of published, accepted, and rejected papers appearing in College & Research Libraries". *College & Research Libraries* Vol. 54, no. 6 (1993): 517–526.

21. Morato, Jorge, Ana Iglesias, Adrián Campillo, and Sonia Sanchez-Cuadrado. "Automated readability assessment for Spanish e-government information". *Journal of Information Systems Engineering and Management* Vol. 6, no. 2 (2021): em0137. doi.org/10.29333/jisem/9620.

22. Murali, Mohana N. *Measuring the readability of newspapers a comparative study of two national English dailies*. 2016. Accessed April 12th 2021. http://hdl.handle.net/10603/145898.

23. Patty, Willard W. and W. I. Painter. "A technique for measuring the vocabulary burden of textbooks". *The Journal of Educational Research* Vol. 24, no. 2 (1931): 127–134.

24. Plavén-Sigray, Pontus, Granville James Matheson, Björn Christian Schiffler and William Hedley Thompson. "The readability of scientific texts is decreasing over time". *Elife 6 (2017):* e27725. doi.org/10.7554/eLife.27725.

25. Robinson, Elaine and David McMenemy. "To be understood as to understand': a readability analysis of public library acceptable use policies". *Journal of Librarianship and Information Science* Vol. 52, issue 3 (2020): 713–725. doi.org/10.1177/0961000619871598.

26. Rodriguez, Timaná et al. "Software to Determine the Readability of Written Documents by Implementing a Variation of the Gunning Fog Index Using the Google Linguistic Corpus". In *International Conference on Applied Technologies*, 409–420. Springer: Cham, 2019.

27. Shoki, Godwin. "Readability as consideration for book selection criterion in book selection practices in some academic libraries in Nigeria". Paper presented at World Library and Information Congress: 73rd Ifla General Conference and Council, Durban, South Africa, 2007. Accessed June 28th 2007. https://archive.ifla.org/IV/ifla73/papers/132-Shoki-en.pdf.

28. Si, Luo and Jamie Callan. "A statistical model for scientific readability". In *Proceedings of the tenth International Conference on Information and Knowledge Management*, 2001, 574–576. doi/10.1145/502585.502695.

29. Smith, Edgar A. and R. J. Senter. "Automated readability index". *Aerospace Medical Research Laboratories* Vol. 66, no. 220 (1967): 5–12.

30. Vajjala, Sowmya and Detmar Meurers, „On improving the accuracy of readability classification using insights from second language acquisition". *The 7th Workshop on the Innovative Use of NLP for Building Educational Applications*, 2012, 163–173. Accessed January 25th 2019. https://aclanthology.org/W12-2019.pdf.

31. Venugopal, Gayatri, Dhanya Pramod and Jatinderkumar R. Saini, „Analyzing Complex Words in Hindi using Parameters of Classical Readability Formulae (Part 1)". *Computer Science* Vol. 29, no. 3 (87) (2021): 366–387. https://doaj.org/article/0574527ac7f14b13a4aacb5da4106a04.

32. Vogel, Mabel and Carleton Washburne. "A Year of Winnetka Research." *The Journal of Educational Research* Vol. 17, no. 2 (1928): 90–101.

33. Waples, Douglas and Ralp W. Tayler. *What People Want to Read About*: *a study of group interests and a survey of problems in adult*. Chicago: American library Association; The University of Chicago Press, 1931.

Концепт читљивости текста из угла библиотекара

Резиме

Тестови и оцене читљивости помажу нам да разликујемо књиге писане разумљивим језиком од књига писаних сложеним језиком. У овом раду полази се од претпоставке да у данашњем друштву постоји потреба за информацијама које ће помоћи појединцу у доношењу одлука тако што ће му, кроз виши ниво информационе писмености, обезбедити успешније генерисање нових знања. Примарни циљ прегледа литературе везане за читљивост текстова јесте истраживање формула читљивости (Flesch Reading Ease, Flesch-Kincaid Grade Level, Dale Chall Readability Formula, Gunning Fog Index, McLaughlin SMOG Formula, Fry readability formula, FORCAST formula, Golub syntactic density score, John Bormuth Formulae, Lexile Framework, ATOS readability formula), како би се пронашле најпогодније формуле за анализу вокабулара одређеног текста и оптимизовало корисничко искуство. Једнако је важан и други циљ рада: примена методологије везане за испитивање читљивости докумената на библиотеке, односно на онај сегмент њиховог рада који се тиче набавке књига. Аутори закључују да би: а) испитивање читљивости помогло професионалцима да направе бољи избор књига за библиотечке збирке; б) бољи избор одговарајућих књига коначно довео до вишег нивоа информационе писмености корисника.

Кључне речи: информациона писменост, формула читљивости текста, оцена читљивости, тест читљивости, библиотеке, набавка књига