Journal of King Saud University – Computer and Information Sciences 31 (2019) 528–540

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Connect the dots: Accessibility, readability and site ranking – An investigation with reference to top ranked websites of Government of India

Abid Ismail *, K.S. Kuppusamy *, Ajit Kumar, Pawan Kumar Ojha

Department of Computer Science, School Of Engineering and Technology, Pondicherry University, Pondicherry 605014, India

ARTICLE INFO

Article history:
Received 26 December 2016
Accepted 16 May 2017
Available online 22 May 2017

Keywords:
Readability
Language complexity
Readability indices
Accessibility
Ranking
Grading
Correlation
Classification
Measurement techniques

ABSTRACT

With the emergence of World Wide Web (WWW) as the primary reference channel of information, the need for making it barrier-free for all categories of users has evolved into a critical factor. Making the web resources barrier-free for users requires action across various dimensions such as accessibility and readability. This paper presents an analysis of accessibility, readability, and site-ranking of top ranked (N = 20) government websites of India. The accessibility analysis has been carried out using aChecker and WAVE tools. The readability of the contents of the website is measured with six different indices such as Flesch-Kincaid reading ease, Flesch-Kincaid grade level, Gunning fog, SMOG, Coleman-Liau index and Automated Readability Index. The ranking of sites by National Informatics Centre (NIC) has been utilized to select the top ranked websites and their corresponding rankings are compared with global site ranking services such as Alexa. The correlation among these three factors of accessibility, readability, and site-ranking has been carried out with Spearman's rank correlation method and the inferences derived from the results are presented.

1. Introduction

In the twenty-first century, the World Wide Web (WWW) has become an essential tool for information consumption. A wide array of tasks can be done using the web. Governments are also moving towards web-society or digital society to provide e-services to their citizens. At the same time, we need to make these resources accessible for all, in order to achieve universality. The web accessibility is the cornerstone of achieving this universality of access across a wider spectrum of users.

Currently, all types of users are utilizing the web resources for information and communication purposes so it is essential to find the accessibility status in terms of web accessibility standards. The main motive of web accessibility is to make the web accessible for all. According to *Tim Berners Lee*, *The power of the Web is in its Universality. Access by everyone regardless of disability is an essential aspect.* More specifically, web accessibility means that people with disabilities can *perceive, understand, navigate, and interact* with the Web.

To achieve web accessibility, every site should follow the guidelines which are proposed by World Wide Web Consortium (W3C).[1] There are two versions of Web Content Accessibility Guidelines (WCAG). They are called WCAG 1.0 and WCAG 2.0. For checking and evaluating accessibility status of websites, we have various evaluation tools available such as aChecker (AChecker, 2016), WAVE (WAVE, 2016), Evalacess 2.0, HERA, Cynthia Says, TAW, FAE, etc which are based on the aforementioned guidelines.[2]

After the accessibility evaluation process, it is important to check the readability and site ranking of the sites in order to find

* Corresponding authors.
E-mail addresses: abidpu2015@gmail.com (A. Ismail), kskuppu@gmail.com (K.S. Kuppusamy), ajitkumar.pu@gmail.com (A. Kumar), pawanojhacs2@yahoo.in (P.K. Ojha).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

[1] W3C W3C, 2010. Accessibility W3C. Retrieved June 12, 2012, from. http://www.w3.org/standards/webdesign/accessibility, 2010. [Online]. Available: W3C, 2010. Accessibility W3C. Retrieved June 12, http://www.w3.org/standards/webdesign/accessibility.

[2] https://www.w3.org/WAI/ER/tools/.

http://dx.doi.org/10.1016/j.jksuci.2017.05.007
1319-1578/

their relationships. Readability makes the quality of written text easy or difficult to read and understand. The processing of data gives information and the information is useful only when we can make sense of it. Because of this, readability checking for web pages becomes an important task to make the contents easily consumed by various categories of users.

Ranking of the sites is also an important factor considering the mammoth amount of web resources available today. There are two types of site ranking, one is *query specific*and another is *non-query specific*. The query specific site ranking is the rank of site in the search engine result list with respect to a particular query. The non-query specific ranking is based on the various attributes like contents of the page, number of visitors, etc. There are various non-query specific site ranking systems which are used for providing the ranking of the sites with respect to particular features like SortSite, Alexa, NIC ranking etc. We have used Alexa ranking (Alexa, 2016) for providing the ranking of top 20 government websites which are given by National Informatics Centre (NIC) so that we can find the association between the Alexa ranking and NIC ranking of the websites. In addition to this, we also generated ranking from accessibility and readability data for linking the dots, *accessibility, readability and site ranking*.

The assessment of top twenty government websites of India for persons with disabilities in terms of their accessibility and readability score is the main objective of this study. In order to connect the dots such as accessibility, readability and site ranking, first of all, we should measure the websites with respect to corresponding parameters and then find their correlations. The measurement and correlation of top 20 websites in terms of accessibility, readability, and ranking has been presented in this paper. The overall goals of this study are as listed below:

1. To find the accessibility score of top 20 government websites in terms of WCAG 2.0 guidelines by using accessibility evaluation tools.
2. To find the readability status of top 20 government websites by using different readability evaluation tools.
3. To find website ranking of these selected websites by using an on-line automatic tool called Alexa tool.
4. To classify these selected websites based on complexity score.
5. To find a correlation between accessibility, readability and ranking score by using the statistical approach.

2. Related Work

The *web accessibility* is an active domain of research. There exists many studies on the accessibility and readability analysis of web resources. In the paragraphs we have listed some of the studies relevant to this paper.

An evaluation of Government websites of Kerala on the basis of Indian Government Guidelines and the five point analysis of *accuracy, authority, objectivity, currency and coverage,* was conducted to determine the quality of information and evaluation of web resources (Rajani and Muralidhara, 2016). In 2016, an exploratory study on Indian university home pages was carried out by another study. This study carried out the analysis by using various accessibility analysis tools and readability checking tools. They have used Gunning Fog Index (Ismail and Kuppusamy, 2018) as a measure for readability. Another study on web accessibility analysis of 20 Malaysian universities was performed (Ahmi and Mohamad, 2015) by using aChecker and WAVE tool.

For computing readability, there are few standard formula that are widely used by experts in different fields such as business, health care, publishing, education, industry, military, etc. With these readability formulae, many articles were published by different researchers such as Manzo (1970), Bruce et al. (1981), Lange (1982), Connatser (1999) and Misra et al. (2013), etc.

Also, William H Dubay's book named *Smart Language* (DuBay, 2007) and its reviewed version (DuBay, 2008) presents the fundamental concepts about readability and the formulae used to check readability. The computational procedure and grading of texts on the basis of readability score are also described in the said book (DuBay, 2008). In June 2014, Raj et al. have carried out a cross-sectional study of health information websites in India. It was found that only thirty-two (32) health information websites out of fifty (50) websites were evaluated in terms of quality (LIDA Tool) and readability (Flesch Reading Ease Score, Flesch-Kincaid Grade Level, and SMOG). The result found that only three websites have high LIDA score and only five websites have recommended sixth-grade level readability score (Raj et al., 2016).

Researchers such as Poonam Misra et al. have performed the readability evaluation of 17 healthcare-oriented education resources by using Readability Studio Professional (version 2012.1) tool. It was found that the resources or websites have 12th grade as average level by using ten (10) different readability scales (Misra et al., 2013).

An analysis of readability study on 121 articles related to spine related patient websites was performed by researchers (Vives et al., 2009). It is recommended that readability of patient textual material be less than sixth grade level. The results were found that the mean Flesh-Kincaid grade levels greater than 10, so patients have difficulty in comprehension. A study was carried out by Eika (2016) for achieving universal design of text on the web by establishing a deeper readability criteria based on language anti-patterns. Another study based on machine learning approach to investigate the text readability of Bangla language was performed in 2014 (Sinha and Basu, 2014).

A cross-validation experiment was performed (Yamasaki and Tokiwa, 2014) to find the performance of readability method for web documents. For language-independent aspects, easy-to-read (E2R) web content checking and author support tool (LanguageTool) was presented (Nietzio et al., 2014) to check the web content in terms of understandability. The *OSMAN- Open Source Metric for Measuring Arabic Narratives,* a modified version of readability formulae called Flesh and Fog was used to calculate readability of Arabic text in the study conducted by El-Haj and Rayson (2016).

A survey on the Computational evaluation of text readability was performed by Collins-Thompson (2014) to predict the text difficulty, and a new challenges and opportunities for future investigation were also prepared by researchers Collins-Thompson (2014) for future exploration.

In 2016, Kumar et al. calculated the readability metrics from patient education materials (PEM) by using the TextStat 0.1.4 textual analysis package for Python 2.7 (Kumar et al., 2016). They also compared the readability and content of online patient education materials (PEM) with institutions having fellowship and the institutions without fellowship (Kumar et al., 2016). It was found that mean Flesh Kincaid grade level for PEM for institutions having fellowship is 13.8 and for non-fellowship 10.8. Also, the readability score falls above the recommended grade level which is sixth-grade level or lesser.

Another study (Sheehan, 2016) was conducted to provide information about *TextEvaluator- a text analysis tool* by its measurement approaches such as traditional readability metrics, and classification by human experts (Inheritance and Exemplary approach) to help in understanding the text variations for teachers and other educators. Also, the test of user experience, with the inclusion of persons with disabilities was carried out by Devaner et al.

(2016). They provide an accessible system solution for distance education and the result of the testing process was found to be simple, effective, and productive (Devaner et al., 2016).

The method of readability used for Japanese texts to measure the performance of correlation coefficient based on textbook corpus was done by Satoshi Sato et al. (Sato et al., 2008). Also, an exploratory study was performed by Tomas Persson (Persson, 2016) to find the correlation between Linguistic features in TIMSS science and results from different groups of Swedish 8th-grade students. The study also measured readability and information load by analyzing four features (packing, precision, personification, and presentation of information) of scientific language (Persson, 2016). Another study was conducted to examine the existing readability formulas and help to design effective text simplification software related to health (Kauchak and Leroy, 2016). Also, researchers carried out the work to evaluate the output of ATS (Automatic Text Simplification) system by using automatic metric (Popovic and Štajner, 2016).

Using websites as a tool to share information with all categories of users in the society has become a very popular method by Governments. To make it success, there should be a better result of accessibility and readability of websites. For this success, an evaluation process is necessary to find the readability and accessibility of websites, and accordingly, we can improve or modify them towards a better result.

In this paper, we have selected the top 20 ranked Government websites of India for analyzing their readability and accessibility score with the help of various formulas and tools. Also, we have computed their average grade score level for readability of the content. Based on their estimated score, we find the associations between the accessibility, readability, and site ranking variables by using statistical procedure.

3. Readability

According to Edgar Dale and Jeanne Chall (1949), "Readability is the sum total of all those elements within a given piece of printed material that affect the success a group of readers has with it. The Success is the extent to which they understand it, read it at an optimal speed, and find it interesting" (DuBay, 2007).

Another definition is given by George Klare (1963) that "the ease of understanding or comprehension due to the style of Writing" (DuBay, 2007). Gretchen Hargis and her Colleagues at IBM (1998) state that Readability is defined as the "ease of reading words and sentences" (DuBay, 2007).

Harry McLaughlin (1969)- the creator of SMOG readability formula (DuBay, 2007) defines the readability as "the degree to which a given class of people find certain reading matter compelling and comprehensible." So, this definition focuses the interaction between the text and the readers of known levels of skill, knowledge, and interest. There are two contributors namely the reader and the text, too easy reading. The features include prior knowledge, reading skill, interest, motivation etc., of the reader that make reading easy; and the features like content, style, design, Organization etc., of the text that makes reading easy. So, readability is the ease of reading in terms of above features.

3.1. Guidelines

There are generally eight guidelines (Idler, 2012) for achieving the better readability on the web. So, developers and designers need to concentrate these desirous points so that readability can be achieved in a better way. For better web readability, the eight (8) guidelines are given in Table 1.

Table 1
Eight Guidelines.

S.No	Guidelines
01	Choose Fonts Wisely
02	Font size and Line Spacing are important.
03	Use of High Contrasts.
04	Keep the lines short.
05	Keep paragraphs also short.
06	Get straight to the point.
07	Don't use Jargon.
08	Use lists, images and highlights.

3.2. Reading level algorithms

There are many reading level algorithms which are used for determining how readable the content is by using different features for analysis. Like Gunning Fog, Flesh reading ease, Flesh-Kincaid, etc., helps in determining how readable the content is, and also give the useful indication whether you have created your content at the right level for your intended audience.

The following readability indices (ReadabilityFormulas, 2016) with their formulae (Webpagefx, 2016a) are used in this paper to calculate the readability estimation score of top 20 government ranked websites. Because, as per literature survey, these metric based reading level algorithms mostly used for evaluation of websites.

1. **Flesh Kincaid Readability Ease (FKRE):** It is a simple approach by which we can find the grade level of the reader in terms of text understand-ability. The formula for computing Flesch-Kincaid Readability Ease (FKRE) is in Eq. (1).

$$FKRE = 206.835 - 1.015 \times \left(\frac{words}{sentences}\right) - 84.6 \times \left(\frac{syllables}{words}\right) \quad (1)$$

According to this readability formula, the best text should contain shorter sentences and words. The acceptable score of this formula lies between 60 and 70 reading ease number. The reading ease number normally ranges from 0 to 100 *(Note: It gives two scores: one is a regular number on a scale 1–120 (Higher is better) and other is Grade that shows American school year need for understanding the text)*. For example, the score between 90.0 and 100.0 falls under an average 5th grade considered easily understandable. Similarly, the score between 60.0 and 70.0 are considered easily understood for the 8th and 9th graders and the Score between 0.0 to 30.0 are considered easily understood by college graduates.
The overall summary of Understanding status of text by Flesch Kincaid Readability Ease (FKRE) formula as shown in Table 2.

2. **Flesh Kincaid (FK) Grade Level:** The Flesh Kincaid (FK) Grade Level Formula is in Eq. (2).

$$\text{FKGrade} = 0.39 \times \left(\frac{words}{sentences}\right) + 11.8 \times \left(\frac{syllables}{words}\right) - 15.59 \quad (2)$$

This formula is updated version of Flesch Reading Ease formula and is mostly used in the field of education. The Defence department of US Government uses this formula as a standard test. The grade value above 12 in this formula will be treated as equivalent to grade value 12. Likely, a score of 5.0 indicates grade school level and 7.4 score indicates text can be understood by an average student in 7th grade.

3. **Gunning Fog Score (GF Score):** It is also called *Fog Index* and is similar to Flesch scale but based on a name 'Foggy' words means words having 3 or more syllables. The ideal score of this

Table 2
FKRE: Summary of Understanding status of text.

Readability Score	Understanding Status
90–100	Very Easy
80–89	Easy
70–79	Fairly Easy
60–69	Standard
50–59	Fairly Difficult
30–49	Difficult
0–29	Very Confusing

index is 7 or 8 and anything above 12 is too hard to read. That is, in general, score 5 is readable, 10 is hard, 15 is difficult and 20 is very difficult to understand the text. It estimates the years of formal education needed to understand the text on a first reading. The formula of this test is given in Eq. (3).

$$\text{GFScore} = 0.4 \times \left(\left(\frac{words}{sentences}\right) + 100 \times \left(\frac{\text{complexwords}}{words}\right)\right) \tag{3}$$

In Eq. (3), Complex words are those words with three or more syllables. The count do not include common suffixes such as -es, -ed, or -ing. The Fog Index Reading level score by grade as shown in Table 3.

4. **SMOG Index:** The SMOG is an acronym for Simple Measure Of Gobbledygook[3] and its formula is considered appropriate for secondary age readers that is, 4th grade to college level readers. The output is in US school grade level indicates that the average student who can read the text when falls in that grade level scale. For instance, average student in 7th grade can understand the text having 7.4 score. The formula of the SMOG index is represented as in Eq. (4):

$$\text{SMOGIndex} = 1.0430 \times \left(\sqrt{\left(30 \times \frac{\text{complexwords}}{sentences}\right)}\right) + 3.1291 \tag{4}$$

The description of the formula is as given: In a 30 selected sentences (10 in a row near the beginning, 10 in middle, and 10 in the end), count every word of three or more syllables called *complex words* or polysyllabic words. And next to estimate the square root of a number of polysyllabic words counted and take the root nearest perfect square. Finally, add 3 to the approximate square root. This gives the SMOG grade formula as in Eq. (5) mentioned.

$$\text{SMOGGrade} = 3 + \sqrt{(\text{Polysyllablecount})} \tag{5}$$

5. **Coleman Liau (CL) Index:** It is based on characters instead of syllables per word and sentence length. It also uses US grade based formula to understand the text. This character based formula given by *Meri Coleman and T. L. Liau* (Webpagefx, 2016b) believed that as compared to counting syllables and sentence length, the computerized evaluations of understanding characters are more accurate and easy. The Formula of Coleman Liau Index called CL Index as shown in Eq. (6).

$$\text{CLIndex} = 5.89 \times \left(\frac{characters}{words}\right) - 0.3 \times \left(\frac{sentences}{words}\right) - 15.8 \tag{6}$$

6. **Automated Readability Index (ARI):** It is derived from the ratios representing word difficultly and sentence difficultly. In this formula as shown in Eq. (7), characters are represented as the number of letters and numbers. ARI gives number as output that approximates the age needed to understand or comprehend the text and is also based on US grading level system as shown in Table 4.

$$ARI = 4.71 \times \left(\frac{characters}{words}\right) + 0.5 \times \left(\frac{words}{sentences}\right) - 21.43 \tag{7}$$

[3] SMOG stands S-Simple, M-Measure, O-Of, G-Gobbledygook means Jargon or meaningless language.

Table 3
Summary of Fog Index Reading level score (GF Score) by grade.

GF Score	Grade	GF Score	Grade
6	Sixth grade	12	High school senior
7	Seventh grade	13	College freshman
8	Eighth grade	14	College sophomore
9	High school freshman	15	College junior
10	High school sophomore	16	College senior
11	High school junior	17	College graduate

4. Data set selection

Indian Government has recently launched a campaign "Digital India" in order to progress the country towards the Digital world. The Government of India is taking efforts in all sectors to make this campaign fully successful. Also, Government websites have been tracked and analyzed by National Informatics Centre (*Web Analytics service, NIC- an initiative under* **india.gov.in** *National Portal of India*)[4] based on Web traffic to help them in enhancing and understanding the usage of their websites, and also helped in rank calculation of websites. The Table 5 list out the top 20 Indian Government websites along with Name of Website and ranking Status based on web traffic performed by National Informatics Centre (NIC), India and the said ranking report was collected in September 2016.

Due to their high ranking, these twenty websites of Government of India are chosen for analysis in our study. Our study has checked their accessibility and readability stature by using different accessibility and readability tools in order to connect the dots-accessibility, readability and site ranking.

The following Table 5 shows the top 20 Govt. websites of India along with Name and ranking status performed by NIC (Webanalytics, 2016), India.

5. Readability test tools and analysis

There are many online readability score calculation tools available. For our analysis, we use following methods which are based on the collection of readability score calculation tools.

1. *Online-Utility.org:* It is a collection of free readability score calculation tools. The tool covers four measures of readability in terms of US grade level to comprehend the text, these are included *Coleman Liau index, Flesh Kincaid Grade level, ARI (Automated Readability Index) and SMOG (Simple Measure of Gobbledygook).* This online-utility.org (OnlineUtility, 2016) also includes *Gunning Fog Index* that provides the indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading. It also displays the suggestions of complicated sentences in order to do improvement in readability.
2. *Readability Test Tool:* This tool (Webpagefx, 2016a) is available online to check the readability of web pages by three different ways, that is *Test by URL, Test by Direct Input or Test by Referer.* For our testing, we use test by URL method for top 20 Government websites of India to check the readability estimation score by different readability indices methods.

[4] http://www.webanalytics.gov.in/.

Table 4
The Grade Level V/S Age.

Age (Yrs)	Grade	Age (Yrs)	Grade
5–6 old	Kindergarten	12–13 old	Seventh
6–7 old	First	13–14 old	Eighth
7–8 old	Second	14–15 old	Ninth
8–9 old	Third	15–16 old	Tenth
9–10 old	Fourth	16–17 old	Eleventh
10–11 old	Fifth	17–18 old	Twelfth
11–12 old	Sixth	18–22 old	College

Table 5
List of Top 20 Ranked Indian Government Websites.

Rank	Name of Website	URLs
1	Madhya Pradesh Educational Portal	http://educationportal.mp.gov.in/
2	Ministry of Drinking Water and Sanitation Swachh Bharat Mission- Gramin	http://tsc.gov.in/
3	High Court of Bombay	http://bombayhighcourt.nic.in/
4	Press Information Bureau, GOI	http://pib.nic.in/
5	Indian army, GOI	http://indianarmy.nic.in/
6	Prasar bharati	http://newsonair.nic.in/
7	Commissioner Land Record and settlement	http://landrecords.mp.gov.in/
8	National Informatic Centre	http://www.nic.in/
9	Ministry of Drinking Water and Sanitation National Rural Drinking Water Programme	http://indiawater.gov.in/
10	The President of India	http://presidentofindia.nic.in/
11	Indian Govt. Tenders Information System	http://tenders.gov.in/
12	National Portal of India,GOI	http://india.gov.in/
13	National Petroleum Limited	http://petroleum.nic.in/
14	Data Portal India	http://data.gov.in/
15	My LPG	http://mylpg.in/
16	Ministry of finance,GOI	http://mof.gov.in/
17	Ministry of Environment, Forest and Climate Change, GOI	http://www.moef.nic.in/
18	Indira Gandhi National Center For Arts	http://ignca.nic.in/
19	National Fertilizers Limited, GOI	http://nationalfertilizers.nic.in/
20	Indian Air Force	http://indianairforce.nic.in/

The usage of readability test tool (Webpagefx, 2016a) involves the following steps:

1. Pass URL of the website in a given tool and then start the process of testing.
2. Testing counts a number of sentences, words, complex words, the percentage of complex words, average words per sentence, and average syllables per word.
3. Then, testing involves the following metrics to calculate the readability score individually.
   - Flesh Kincaid Reading Ease 1.
   - Flesh Kincaid Grade Level 2.
   - Gunning Fog Score 3.
   - SMOG Index 4 & 5.
   - Coleman Liau Index 6 and
   - Automated Readability Index 7.
4. Based on step 2 and 3, we get average grade level and a number of years old to understood the text of the web pages as per US Grade system.

The Section 5.1 provides the calculated readability score of top 20 Govt. websites in terms of different readability methods used. The Section 5.2 mentioned the Text Statistics report of websites which are selected for testing. The Section 5.3 represents the average grade level score and correspondingly the number of years needed to easily understand the language of the selected top 20 websites of Indian Government.

5.1. Readability indices

The estimation report of different readability indices methods as shown in Table 6 which are used for top 20 Government websites of India to check their readability estimation score. Also, we find their status of readability achievement in terms of complexity levels classification that is, very low, low, normal, high and very high levels, as shown in Table 7 and its diagrammatically representation as in Fig. 3. The complexity level indicators are explained as:

- VL = Very Low means text is too complicated to understand.
- L = Low means text is complicated to understand.
- N = Normal means text is easy to read.
- H = High means text is easier to read.
- VH = Very High means text is easiest to read.

The graphical representation of calculated score of Top 20 Government Websites by Flesh Kincaid Reading Ease formula is in Fig. 1 and by other indices such as Flesch Kincaid grade level, Gunning fog, SMOG, Coleman-Liau index, Automated Readability index is in Fig. 23.

The diagrammatic representation of data obtained from these readability indices tools are shown in Fig. 4.

5.2. Text statistics

This section provides detailed information about the text inside the websites which were analyzed by using different indices in terms of their readability to find a number of sentences, words, complex words, percent of complex words, average words per sentences and average syllables per word. Also, there are some indices like Coleman Liau and Automated Readability Index (ARI) based on counting the characters, words and sentences, and some other based on a number of syllables and complex words. The overall text statistics of Top 20 ranked government websites of India are shown in Table 8.

5.3. Test results

On the basis of readability score calculated by all readability tools, the overall average grade level per website with their number of years to understand the text is mentioned in Table 9. The result indicates that 10% websites are very easy to understand, 35% websites are easy to understand and rest of the websites falls in hard (35%) and very hard (20%) to understand the text. The overall test result summary of Top 20 Government websites of India is shown in Table 9.

6. Web accessibility analysis

For accessibility analysis of websites, various types of tools are available for the evaluation process. As per literature survey like (Alahmadi and Drew, 2016; Ahmi and Mohamad, 2015; Ismail and Kuppusamy, 2018), etc., *aChecker and WAVE* were observed to be a major tools used by accessibility analysis and are also open source software tools. So, we used *aChecker* (AChecker, 2016) and *WAVE*(WAVE, 2016) evaluation tools for our analysis of Top 20 Government websites.

Table 6
Readability Indices of Top 20 ranked government websites of India.

Rank	Name of Website	Flesch Kincaid Readability Ease	Flesch Kincaid Grade Level	Gunning Fog Score	SMOG Index	Coleman Liau Index	Automated Readability Index
1	Madhya Pradesh Educational Portal	62.4	5.8	4.1	5.5	11.2	2.8
2	Ministry of Drinking Water and Sanitation Swachh Bharat Mission -Gramin	49	08	5.7	6.9	14.3	5.9
3	High Court of Bombay	52.4	6.6	4.8	4.9	15.3	4.9
4	Press Information Bureau, GOI	121.2	−3.4	0.4	1.8	−16.1	−20.9
5	Indian army, GOI	39	09	4.3	6.8	16.5	6.8
6	Prasar bharati	45.9	10.1	8.5	9.8	13.1	8.3
7	Commissioner Land Record and Settlement	116	−1.9	1.9	02	−14.2	−18
8	National Informatic Centre	40.9	09	7.1	7.5	16.9	7.8
9	Ministry of Drinking Water and Sanitation National Rural Drinking water Programme	41.1	9.3	4.7	7.4	15.7	7.4
10	The President of India	−19.4	17.8	5.1	8.9	25.6	15.4
11	Indian Govt. Tenders Information System	34.8	12.4	09	10.8	15.9	11.9
12	National Portal of India,GOI	26.1	11.2	4.9	8.1	19.7	10.1
13	National Petroleum Limited	40.8	9.2	6.1	08	15.1	6.6
14	Data Portal India	46.9	7.7	5.7	5.9	14.8	05
15	My LPG	55	7.8	07	7.4	13.3	6.4
16	Ministry of finance,GOI	69.5	5.1	6.4	4.8	11.3	3.4
17	Ministry of Environment, Forest and Climate Change, GOI	30.6	10.1	5.8	6.9	18.4	8.3
18	Indira Gandhi National Center For Arts	52.9	8.4	12	8.3	17.2	9.9
19	National Fertilizers Limited, GOI	42.2	8.6	6.8	6.8	15.5	6.1
20	Indian Air Force	45.6	8.4	2.4	8.3	14.6	5.9

Table 7
Complexity Level score in terms of Percentage for Readability Indices estimation score.

Complexity	Percentage Readability Indices Value (Top 20 Websites)					
	Flesch Kincaid Readability Ease	Flesch Kincaid Grade Level	Gunning Fog Score	SMOG Index	Coleman Liau Index	Automated Readability Index
Very Low (VL)	10	5	0	0	25	0
Low (L)	55	0	0	0	55	5
Normal(N)	20	35	15	10	10	25
High (H)	5	50	70	80	0	50
Very High (VH)	10	10	15	10	10	20

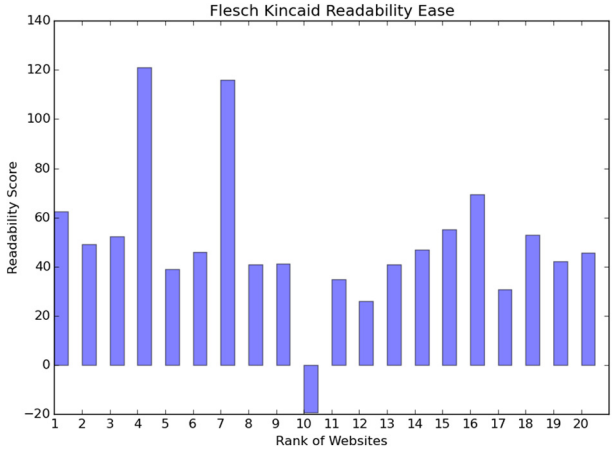

Fig. 1. Graphical Representation of Top 20 Govt. Websites by Using Flesh Kincaid Reading Ease formula.

6.1. aChecker tool

The aChecker tool (also called Web Accessibility Checker) is available online to check the accessibility of websites under the three Levels of Conformance of WCAG 1.0 as well as WCAG 2.0. In addition to this, we have also many options like Section 508, Stanca Act, etc. for the evaluation process. The aChecker tool gives results in three different categories namely *Known problems, Likely problems, and potential problems*. Also, it presents result classification in a well-ordered manner. There are three levels of conformances used to meet the needs of different groups in different situations namely Level A(lowest), Level AA (medium) and Level AAA (highest). We have adopted AA, as it is the common proactice adopted by accessibility studies (Calvo et al., 2016).

The process of website accessibility calculation by the aChecker tool is as follows:

Collect URL of the website and passed it into aChecker tool address box as shown in Fig. 5. Then, go to option button select WCAG 2.0 (Level AA) level of conformance and then start to check it. After processing, we get three types of results with proper descriptions namely Known Problems, Likely Problems, and Potential Problems. We then sum up all three problems and then make final web accessibility violation score. Based on this score, accordingly we provide the status of the website that is, high score violation website have lower rank by using SPSS.

The summarized evaluation result report of Top 20 Govt. websites by aChecker is given in Table 10 and the diagrammatic representation is shown in Fig. 6.

6.2. WAVE tool

A WAVE tool is also called Web Accessibility Evaluation Tool, developed and maintained by WebAIM called Web Accessibility In Mind. It is originally launched in 2001 and millions of web pages are evaluated for accessibility with the help of this tool. It also helps to determine the accessibility of web content. Here, we also passed URL of the website into WAVE tool and then click or enter

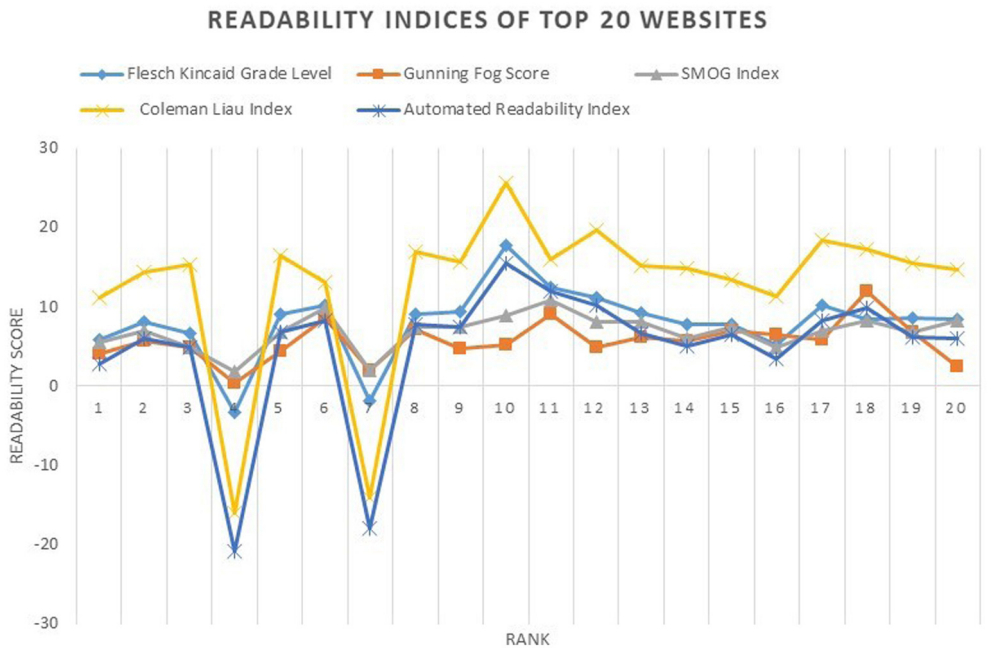

Fig. 2. Graphical Representation of Top 20 Govt. Websites by using Flesch Kincaid grade level, Gunning fog, SMOG, Coleman-Liau index, Automated Readability index.

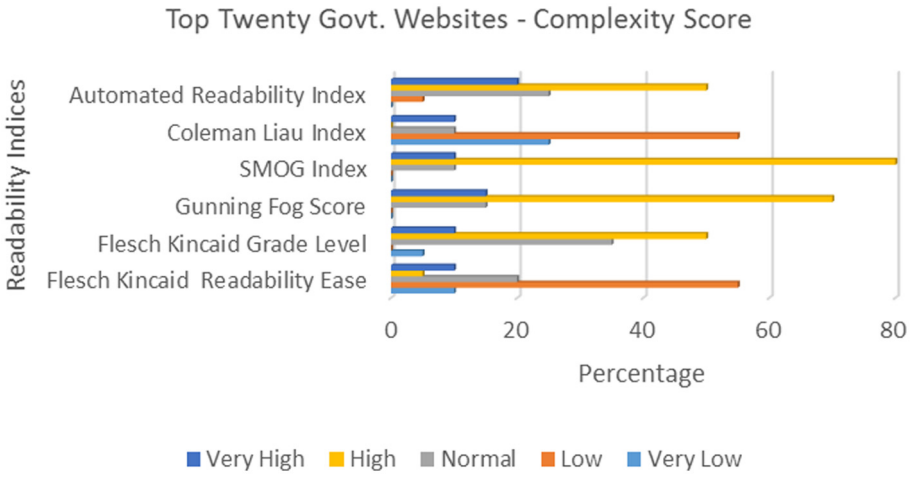

Fig. 3. Readability indices with complexity based classification of Top 20 Government websites.

the arrow button to start the processing. The WAVE tool detects the following features of the particular website in number namely Errors, Alerts, Features, Structural Elements, HTML5 & ARIA and Contrast Errors. Also, we will get details of WCAG 2.0 violations about styles, no styles, and contrasts by using this tool. Based on these violation scores, we added them and make a final report. The websites having high WAVE score, provides less rank regarding accessibility guidelines by using SPSS.

The WAVE report of top 20 government websites which were used for our accessibility analysis and their mean and standard deviation are given in Table 11. Also, their corresponding graphical representation is shown in Fig. 7.

7. Site ranking

We have used an online automatic tool called Alexa (2016) for site ranking to make the comparison between the ranking provided by NIC and Alexa. The Alexa tool provides a ranking of the sites in two ways namely Alexa National Ranking and Alexa Global Ranking. The overall ranking of websites with their comparison is shown in Fig. 8. Also, the ranking of dots such as readability, accessibility, and site ranking (NIC and Alexa ranking) generated by Test cases in SPSS and finding their associations are mentioned in Section 8.

8. Association between variables- A statistical inference

There are different statistical methods which are used for finding the associations between the variables. Due to the selection of ranked websites for our study, we used Spearman's Rank correlation to find the correlations between the variables in terms of positive, negative or zero correlations. The variables which are used for Spearman's Rank correlation are *NIC Ranking variable, Alexa National Ranking variable, Alexa Global Ranking variable, aChecker Ranking variable, WAVE Ranking Variable, and FK EaseReadability Ranking variable*. Hence, we have selected these above variables

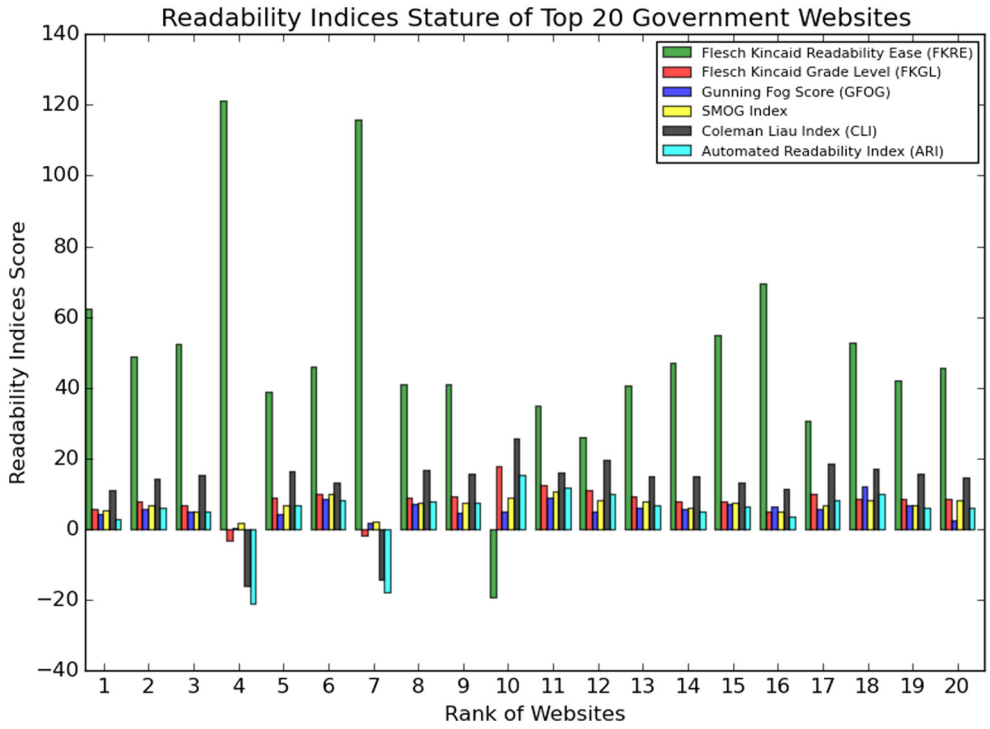

Fig. 4. Representation of Readability indices with Calculated data of Top 20 Government Websites.

Table 8
Text Statistics of Top 20 ranked government websites of India.

Rank	Name of Website	No. of Sentences	No.of Words	No. of Complex Words	Percent of Complex Words	Average Words per Sentence	Average Syllables per Word
1	Madhya Pradesh Educational Portal	303	1560	249	15.96	5.15	1.65
2	Ministry of Drinking Water and Sanitation Swachh Bharat Mission-Gramin	231	1504	309	20.35	6.51	1.79
3	High Court of Bombay	166	460	98	21.3	2.83	1.78
4	Press Information Bureau, GOI	01	01	00	00	01	01
5	Indian army, GOI	194	896	249	27.79	4.65	1.93
6	Prasar bharati	50	634	137	21.61	13.1	1.75
7	Commissioner Land Record and Settlement	61	228	01	0.44	4.25	1.02
8	National Informatic Centre	107	483	113	23.4	5.95	1.84
9	Ministry of Drinking Water and Sanitation National Rural Drinking water Programme	77	544	117	21.51	7.13	1.87
10	The President of India	116	856	269	31.43	7.38	2.59
11	Indian Govt. Tenders Information System	18	286	62	21.68	15.89	1.84
12	National Portal of India,GOI	222	1288	394	30.59	6.17	2.06
13	National Petroleum Limited	204	1148	301	26.22	6.54	1.85
14	Data Portal India	174	592	129	21.79	3.86	1.8
15	My LPG	35	249	35	14.06	8.96	1.62
16	Ministry of finance,GOI	06	32	03	9.38	6.2	1.52
17	Ministry of Environment, Forest and Climate Change, GOI	440	1445	374	25.88	4.55	02
18	Indira Gandhi National Center For Arts	01	10	02	20	10	1.7
19	National Fertilizers Limited, GOI	189	870	225	25.86	4.89	1.87
20	Indian Air Force	01	06	02	33.33	06	1.83

for finding the associations between them to connect the dots namely *Accessibility, Readability, and Ranking*.

We used different readability indices formulas, accessibility tools, Alexa tool and SPSS Statistics tool for the calculation of correlation between the said variables. First of all, we estimated the score and status of these top 20 ranked selected websites with the help of readability testing involving different readability indices tools. Next to collect web content accessibility guidelines (WCAG) report of these top 20 websites by using the online web accessibility tool called aChecker (AChecker, 2016) and WAVE (WAVE, 2016). Also, for finding website traffic, statistics and analytics of these top 20 selected government websites in terms of global as well as national ranking, another online tool called Alexa (2016), was used. After this, we used SPSS statistics technique to *transform this calculated data into rank cases* and then find correlations between corresponding combination of variables such as cor-

Table 9
Average Grade Level Test Results of Top 20 websites along with No. of years to understand the text.

Rank	Name of Website	URLs	Average Grade Level	No. of Years Old to Easily Understood
1	Madhya Pradesh Educational Portal	http://educationportal.mp.gov.in/	6	11 to 12
2	Ministry of Drinking Water and Sanitation Swachh Bharat Mission-Gramin	http://tsc.gov.in/	8	13 to 14
3	High Court of Bombay	http://bombayhighcourt.nic.in/	7	12 to 13
4	Press Information Bureau, GOI	http://pib.nic.in/	−8	−3 to −2
5	Indian army, GOI	http://indianarmy.nic.in/	9	14 to 15
6	Prasar bharati	http://newsonair.nic.in/	10	15 to 16
7	Commissioner Land Record and Settlement	http://landrecords.mp.gov.in/	−6	0 to −1
8	National Informatic Centre	http://www.nic.in/	10	15 to 16
9	Ministry of Drinking Water & Sanitation National Rural Drinking Water Programme	http://indiawater.gov.in/	9	14 to 15
10	The President of India	http://presidentofindia.nic.in/	15	20 to to 21
11	Indian Govt. Tenders Information System	http://tenders.gov.in/	12	17 to 18
12	National Portal of India,GOI	http://india.gov.in/	11	16 to 17
13	National Petroleum Limited	http://petroleum.nic.in/	9	14 to 15
14	Data Portal India	http://data.gov.in/	8	13 to 14
15	My LPG	http://mylpg.in/	8	13 to 14
16	Ministry of finance,GOI	http://mof.gov.in/	6	11 to 12
17	Ministry of Environment, Forest and Climate Change, GOI	http://www.moef.nic.in/	10	15 to 16
18	Indira Gandhi National Center For Arts	http://ignca.nic.in/	11	16 to 17
19	National Fertilizers Limited, GOI	http://nationalfertilizers.nic.in/	9	14 to 15
20	Indian Air Force	http://indianairforce.nic.in/	8	13 to 14

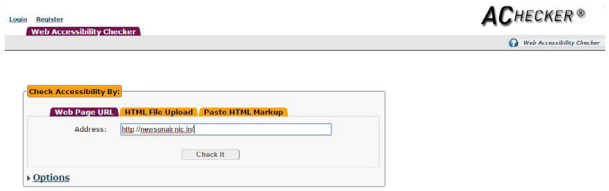

Fig. 5. Snapshot of aChecker Tool.

Table 10
Accessibility Report of Top 20 Governmnet websites by aChecker Tool.

WCAG 2.0	Level AA		
	Known	Likely	Potential
Total Errors	2437	100	9152
Mean	121.85	5	457.6
Std. Deviation	220.43	15.33	336.41

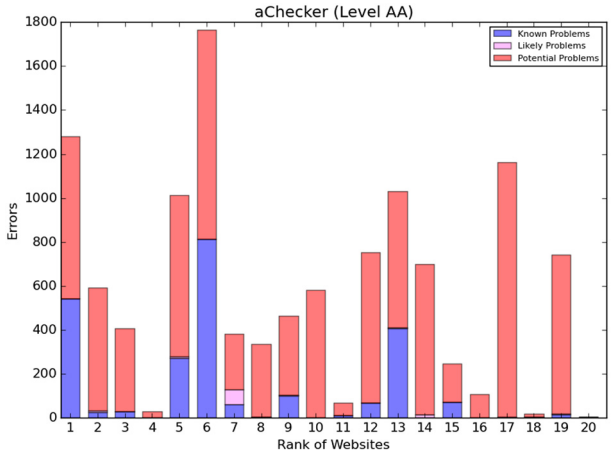

Fig. 6. Representation of aChecker Result of Top 20 Websites under WCAG 2.0 (Level AA).

relation between NIC Ranking and Alexa National Ranking, between aChecker Ranking and Alexa Global Ranking, between aChecker Ranking and NIC Ranking, between WAVE Ranking and NIC Ranking, between WAVE Ranking and Alexa Global Ranking, between FK EaseReadability Ranking and NIC Ranking, between FK EaseReadability Ranking and aChecker Ranking, between FK EaseReadability Ranking and WAVE Ranking under *bi-variate type*.

To estimate the *Spearman's rank correlation*, we have two equations (Spearman, 2016) to follow. One is used on data with no tied ranks as in Eq. (8) and other having tied ranks as in Eq. (9). The equations for data having no tied ranks 8,

$$\rho = 1 - \frac{6 \times \sum d_i^2}{n(n^2-1)} \tag{8}$$

where, ρ= Rank correlation coefficient, d_i=difference in paired ranks, and n = total number of cases in the data series.

and for data having equal or tied ranks 9,

$$\begin{aligned} \rho &= 1 - \frac{6\left(\sum d_i^2 + \frac{1}{12}CF\right)}{n\left(n^2-1\right)} \\ CF &= \sum m(m^2-1) \end{aligned} \tag{9}$$

where, ρ= Rank correlation coefficient, d_i=difference in paired ranks, CF=Correction factor, m= number of times the data repeats in the data series, and n = total number of cases in the data series.

Therefore, after running the Spearman's rank correlation procedure, we got the following correlation Table 12 with a different combination of variables to find the positive as well as the negative association between the selected variables. So, the Table 12 presents the different variables correlation in terms of Spearman's correlation Coefficient and the sample size that the calculation was based on.

The spearman's rank order correlation was runned in order to determine the relationship between the selected combination of

Table 11
WAVE Tool Report: Top 20 Government Websites.

Tools/ Statistical Inferences	WAVE					
	Errors	Alerts	Features	Structural Elements	HTML 5 & ARIA	Contrast Errors
Total Errors	498	1340	659	660	167	708
Mean	24.9	67	32.95	33	8.35	35.4
Std. Deviation	26.29	72.02	47.45	22.90	22.83	60.94

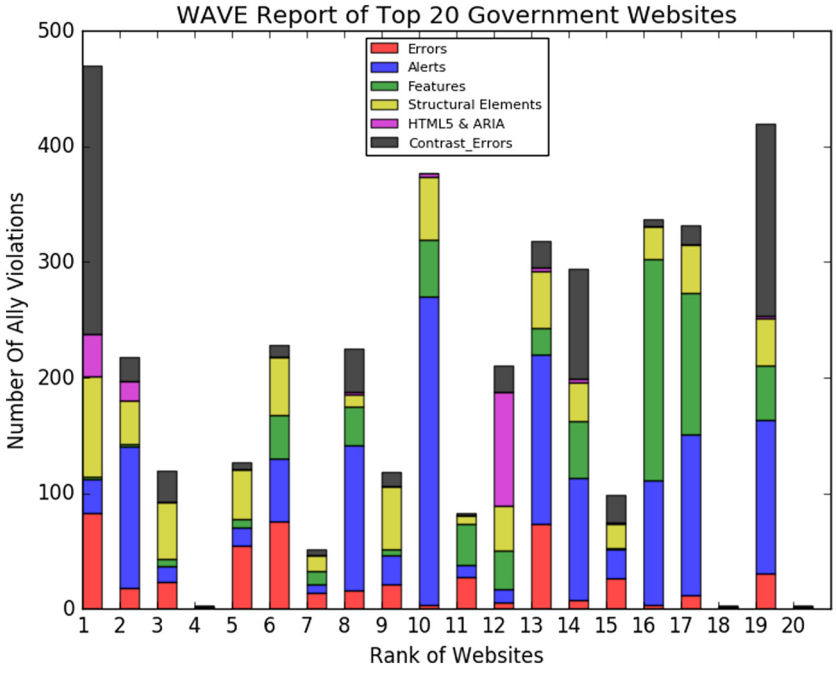

Fig. 7. Representation of WAVE Tool Result of Top 20 Government Websites.

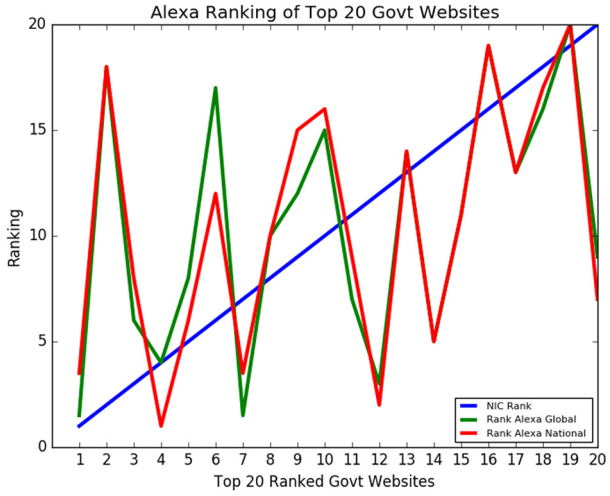

Fig. 8. Ranking Comparison of Top 20 Government Websites.

variables. For connecting the dots, the following correlations as mentioned in Table 12 are useful in general:

1. **Correlation between NIC Ranking and Alexa National Ranking Variables:** There is positive correlation between them having Spearman's correlation coefficient, $r = 0.384$ and coefficient of determination, $r^2 = 14.7\%$. Therefore, the r-value is positive means the variables namely NIC Ranking and Alexa National Ranking are good correlated than others but as compared with scale range $-1\text{to}+1$, their association falls in weak positive correlation category.
2. **Correlation between aChecker Ranking and Alexa Global Ranking Variables:** There is positive correlation between them having Spearman's correlation coefficient, $r = 0.049$ and coefficient of determination, $r^2 = 0.24\%$. Therefore, the r-value is too weak means the variables are weakly positive correlated.
3. **Correlation between aChecker Ranking and NIC Ranking Variables:** There is negative correlation between them having Spearman's correlation coefficient, $r = -0.280$ and coefficient of determination, $r^2 = 7.84\%$. Therefore, the r-value is weak means the variables are weakly negative correlated and falls in a negative region of the range scale towards nearer to zero.
4. **Correlation between WAVE Ranking and NIC Ranking Variables:** There is negative correlation between them having Spearman's correlation coefficient, $r = -0.006$ and coefficient of determination, $r^2 = 0.004\%$. Therefore, the r-value is negative means the variables are weakly correlated and falls in a negative region of the range scale towards nearer to zero.
5. **Correlation between WAVE Ranking and Alexa Global Ranking Variables:** There is positive correlation between them having Spearman's correlation coefficient, $r = 0.337$ and coefficient of determination, $r^2 = 11.36\%$. Therefore, the r-value is positive means the variables namely WAVE Ranking and Alexa Global Ranking are good correlated but as compared with scale range $-1\text{to}+1$, their association falls in weak positive correlation category.

Table 12
Spearman's rho correlation result of twenty Websites

S. No.	Spearman's Rank Correlation	No. of Cases (n)	r- value
1	Between NIC Ranking and Alexa National Ranking	20	0.384
2	Between aChecker Ranking and Alexa Global Ranking	20	0.049
3	Between aChecker Ranking and NIC Ranking	20	−0.280
4	Between WAVE Ranking and NIC Ranking	20	−0.006
5	Between WAVE Ranking and Alexa Global Ranking	20	0.337
6	Between WAVE Ranking and FK EaseReadability Ranking	20	−0.252
7	Between FK EaseReadability Ranking and NIC Ranking	20	−0.224
8	Between FK EaseReadability Ranking and Alexa National Ranking	20	−0.330
9	Between FK EaseReadability Ranking and aChecker Ranking	20	0.174

6. **Correlation between WAVE Ranking and FK EaseReadability Ranking Variables:** There is negative correlation between them having Spearman's correlation coefficient, $r = -0.252$ and coefficient of determination, $r^2 = 6.35\%$. Therefore, the r-value is too weak means the variables are weakly negative correlated.
7. **Correlation between FK EaseReadability Ranking and NIC ranking Variables:** There is negative correlation between them having Spearman's correlation coefficient, $r = -0.224$ and coefficient of determination, $r^2 = 5.02\%$. Therefore, the r-value is weak means the variables are weakly negative correlated.
8. **Correlation between FK EaseReadability Ranking and Alexa National Ranking Variables:** There is a negative correlation between them having Spearman's correlation coefficient, $r = -0.330$ and coefficient of determination, $r^2 = 10.89\%$. Therefore, the r-value is weak means the variables are weakly negative correlated.
9. **Correlation between FK EaseReadability Ranking and aChecker Ranking Variables:** There is a positive correlation between them having Spearman's correlation coefficient, $r = 0.174$ and coefficient of determination, $r^2 = 3.02\%$. Therefore, the r-value is too weak means the variables are weakly positive correlated.

Thus, the overall result generated by SPSS during analysis of data indicates that there were weak positive correlation between the NIC ranking, readability and accessibility rank cases, but also showed the negative correlation between some pair of variables. So, we need to improve their performance in order to get the strong correlation between them. It is also proved that there was the strong correlation between two group of variables (such as NIC ranking and Alexa National ranking variables, and WAVE ranking and Alexa Global ranking variables) as compared with other seven groups of correlation coefficients. In general, it was found that accessibility and readability with the NIC ranking association are weaker than with the Alexa ranking association. But, the association between NIC and Alexa ranking are stronger than the accessibility and readability.

9. Discussions and interpretation of data

In Section 5, we used various readability formulas to calculate the readability score of top 20 selected government websites. The result indicates that 7% of websites are in very low category, 19% of websites are in low category, 18% of websites are in normal category, 43% of websites are in high category and 13% of websites are in very high category score of complexity in terms of their readability status. The average grade level of these websites was observed as 8.4 which indicates that age requirement to understand these text is around 14 years. The overall readability score of these websites are good but there is still scope for further improvements.

In Section 6, we have used two web accessibility evaluation tools to check the accessibility score of these websites in terms of WCAG guidelines. The result calculated by AChecker tool showed that the websites have more potential errors than known errors. Also, the average mean error of these websites was observed as 584.45 and standard deviation error is 480.69. This indicates that a detailed manual evaluation of accessibility shall be carried out to identify specific problems and fix them with context specific solutions.

As per WAVE tool report, HTML5 & ARIA violations are least in these websites and alerts are larger in number. The mean values for *errors* was observed as 24.9, *alerts* was 67, *features* was 32.95, *structural elements* was 33, *HTML5* & *ARIA* was 8.35 and *contrast errors* was 35.4. The standard deviations of errors, alerts, features, structural elements, HTML5 & ARIA, and Contrast errors are 26.29, 72.02, 47.45, 29.90, 22.83 and 60.94, respectively. Hence, the result indicates that a number of alerts and contrast errors are high. For achieving better results in terms of accessibility, we need to minimize these alerts and contrast errors.

In Section 7, we used *Alexa* tool to calculate the ranking of these top 20 government websites in terms of National (country based websites) and Global ranking. The result indicates that the ranking of Alexa national versus Alexa global is highly correlated (0.947) and have a strong connection. With respect to NIC ranking, Alexa national and Alexa global ranking are positively correlated with values of 0.384 and 0.391, respectively. After readability, accessibility and site ranking calculations, the Section 8 mentioned the strength of connection and association among the dots- *Accessibility, Readability and Site Ranking*.

First of all, we have converted all data collected from readability, accessibility and site ranking tools into rank cases by using SPSS transform technique. After the generation of rank cases, we have computed the *Spearman's* correlation under *bivariate type*. The overall Spearman's Correlations among the selected six variables namely NIC, Alexa National, Alexa Global, aChecker, WAVE, and FK Readability ranking variables are shown in Table 13.

The highlighted values in Table 13 indicates the following associations between the variables:

1. The NIC ranking variable with Alexa ranking variables (National and Global) are positively correlated with each other.
2. The Alexa ranking variables are positively correlated with accessibility ranking variables (aChecker and WAVE).
3. And the correlation of accessibility ranking variable (aChecker) with FK Readability ranking variable also shows a positive result.

The diagrammatic representation of associations between variables is illustrated in Fig. 9. In this Fig. 9, the red lines indicate negative correlation and the green lines indicate positive correlation between the variables. Dotted lines indicate how the dots (accessibility, readability and site ranking) are correlated. The differential width of these lines indicate the strength of correlation between the dots.

The overall status of correlation between the dots are *weakly positive*. It was also found that NIC ranking variable association with FK readability, aChecker, and WAVE ranking variables are negative but with Alexa ranking variable there exist a positive association. So, the connection between FK readability ranking, aChecker ranking, WAVE ranking and Alexa ranking variables are comparatively stronger than the connection with the FK readabil-

Table 13
Spearman's Correlations between selected (six) variables

Spearman Correlation Ranking Variables	NIC	Alexa National	Alexa Global	aChecker	WAVE	FK Readability
NIC	1	0.384	0.391	−0.280	−0.006	−0.224
Alexa National	0.384	1	0.947	−0.004	0.356	−0.330
Alexa Global	0.391	0.947	1	0.049	0.337	−0.354
aChecker	−0.280	−0.004	0.049	1	0.694	0.174
WAVE	−0.006	0.356	0.337	0.694	1	−0.252
FK Readability	−0.224	−0.330	−0.354	0.174	−0.252	1

Fig. 9. A Graphical Representation of Connection of Dots involving different variables.

ity ranking, aChecker ranking, WAVE ranking via Alexa ranking variables. Also, there is a negative association of NIC ranking variable with FK readability, aChecker, and WAVE ranking variables.

With the constant emphasis on *Universal design* that enables all categories of users to access the web resources without any barrier, it becomes mandatory that both the content writers and designers of the web pages need to incorporate readability and accessibility standards during the early stage itself. Thus, we should make the websites more accessible and readable in terms of accessibility standards and readability principles which would definitely make it possible to enable a large spectrum of users in accessing the web contents.

10. Limitations

Based on experimental results and analysis we found that there were some limitations which are explored in this section.

Firstly, the readability is measured with metrics proposed for English Language content. Hence, the contents which are in non-English may exhibit sub-optimal results.

Also, the readability was limited to provide the grade of websites based on US grading system. The other country specific grading systems are not included. One of the potential future directions is to evolve country specific metrics which would be based on country specific educational methodologies for various grades.

Moreover, the readability testing techniques were based on the plain text. In readability testing, the other web elements such as hyperlinks, tables etc were not considered. All calculated results were based on fully automated testing tools and no manual evaluation was performed. These tools are based on different parameters and the result was also generated differently. For comparison, we have converted all calculated results into rank cases by using SPSS. This makes us capable of finding correlations among the said variables. Also, it was found that there was a negative weak association between NIC ranking with accessibility variables (AChecker and WAVE variables), and NIC ranking with FK Readability variables. But, all other variables are positively correlated with each other via Alexa ranking.

11. Suggestions

The formulas or methods used for testing the readability of websites were built for English. Hence, there is a strong need to develop readability measures or tests that should be language independent. If the language independent scenario is not possible then there should be efforts to build metrics which would harness the specific features of languages.

Grading for readability has adopted readability testing indices which were tailor-made for US grading system. So, we should focus on making this grading system choice based like country based grading system.

The readability testing measures were primarily built for plain text document and hence they doesn't associate any weight to formating elements. As the web content includes many other elements apart from the plain text such as hyperlinks, tables etc, the readability measurements for web pages shall have provisions which would associate weight for formating elements in addition to the plain text characteristics. One of the important suggestions that we want to put forward is the development of webpage specific readability assessment mechanism.

12. Conclusions

This paper has presented an analysis of Top 20 websites of Government of India, carried out with dimensions such as accessibility, readability and site-ranking. The overall objective is to measure the association between the aforementioned dimensions.

The analysis of accessibility was carried out with tools such as *AChecker and WAVE*. It was observed that the accessibility status of these sites need to be enhanced further so that the contents shall be accessed by persons with disabilities without any barrier.

With respect to readability, the score was computed using six well established techniques. It was observed that the readability score of these sites were in the acceptable range. However, as the scores are based on US grade system, there is a strong need to develop country specific grading mechanism with respect to readability.

The correlations among the various scores were computed with *Spearman rank correlation* method, after porting them into ranks, in order to reduce the intra-tool (accessibility, readability) differences. The correlations between these variables were illustrated with a graphical representation.

The overall conclusion derived from this paper is as follows: The accessibility, readability and site-ranking are three major factors in enabling the web content to reach a wide group of users. With the mammoth amount of web resources available today, if a site didn't find a place in the top of the rank list then it may become invisible to a large number of users. Even though, the page is in the top of the list, if the accessibility status of a web page is not good then

it would loose a major chunk of users in the form persons with disabilities and elderly. With respect to readability, if the content's understandability level doesn't match with the target group (i.e. web designed for children should be at the readability level suitable for them) then the effectiveness of its reach would not be optimal. All these points emphasize the fact that the accessibility, readability and site-ranking should be given adequate priority in making the web resources truly *Universally Accessible.*

References

AChecker, 2016. URL http://achecker.ca/checker/index.php. (Accessed October 2 2016).

Ahmi, A., Mohamad, R., 2015. Web accessibility of the malaysian public university websites. In: Proceedings of International Conference on E-Commerce, pp. 171–177.

Alahmadi, T., Drew, S., 2016. An evaluation of the accessibility of top-ranking university websites: accessibility rates from 2005 to 2015, DEANZ Biennial Conference, pp. 224–233.

Alexa, 2016. URL http://www.alexa.com/siteinfo. (Accessed October 3 2016).

Bruce, B., Rubin, A., Starr, K., 1981. Why Readability Formulas Fail (Reading Education Report no. 28). University of illinois at urbana-cham-paign, Urbana. Center for the Study of Reading.

Calvo, R., Seyedarabi, F., Savva, A., 2016. Beyond web content accessibility guidelines: expert accessibility reviews, Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. ACM, pp. 77–84.

Collins-Thompson, K., 2014. Computational assessment of text readability: a survey of current and future research. ITL-Int. J. Appl. Ling. 165, 97–135.

Connatser, B.R., 1999. Last rites for readability formulas in technical communication. J. Tech. Writing Commun. 29, 271–287.

Devaner, M., Alves, M., Braga, C., Alves, F., Bezerra, W., 2016. Experience report about the inclusion of people with disabilities in the process of testing an accessible system for learning management. World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng. 10, 1174–1177.

DuBay, W.H., 2007. Smart Language: Readers, Readability, and the Grading of Text. ERIC.

DuBay, W.H., 2008. Unlocking language: the classic readability studies. IEEE Trans. Prof. Commun. 4, 416–417.

Eika, E., 2016. Universally designed text on the web: towards readability criteria based on anti-patterns. Stud. Health Technol. Inf. 229, 461.

El-Haj, M., Rayson, PE. 2016. Osman–a novel arabic readability metric, pp. 250–255.

Idler, S. (2012). URL http://blog.usabilla.com/8-guidelines-for-better-readability-on-the-web/

Ismail, A., Kuppusamy, K.S., 2018. Accessibility of indian universities' homepages: an exploratory study. J. King Saud Univ. Comp. Inf. Sci. 30, 268–278.

Kauchak, D., Leroy, G., 2016. Moving beyond readability metrics for health-related text simplification. IT Prof. 18, 45–51.

Kumar, G., Howard, S.K., Kou, A., Kim, T.E., Butwick, A.J., Mariano, E.R., 2016. Availability and readability of online patient education materials regarding regional anesthesia techniques for perioperative pain management. Pain Med. (p pnw179).

Lange, B., 1982. Eric/rcs: readability formulas: second looks, second thoughts. Reading Teacher 35, 858–861.

Manzo, A.V., 1970. Readability: a postscript. Elementary English 47, 962–965.

Misra, P., Agarwal, N., Kasabwala, K., Hansberry, D.R., Setzen, M., Eloy, J.A., 2013. Readability analysis of healthcare-oriented education resources from the american academy of facial plastic and reconstructive surgery. Laryngoscope 123, 90–96.

Nietzio, A., Naber, D., Bühler, C., 2014. Towards techniques for easy-to-read web content. Procedia Comp. Sci. 27, 343–349.

OnlineUtility, 2016. URL https://www.online-utility.org/english/readabilitytestandimprove.jsp. (Accessed September 13 2016).

Persson, T., 2016. The language of science and readability. correlations between linguistic features in timss science items and the performance of different groups of swedish 8th grade students. Nordic J. Literacy Res., 2

Popovic, M., Štajner, S., 2016. Machine translation evaluation metrics for quality assessment of automatically simplified sentences, Quality Assessment for Text Simplification (QATS) Workshop Programme, p. 32.

Rajani, S., Muralidhara, B., 2016. Government websites of kerala: an evaluation using government of India guidelines. Int. J. Comp. Appl., 140

Raj, S., Sharma, V., Singh, A., Goel, S., 2016. Evaluation of quality and readability of health information websites identified through india's major search engines. Adv. Preventive Med.

ReadabilityFormulas, 2016. URL http://www.readabilityformulas.com/free-readability-formula-tests.php. (Accessed September 24 2016).

Sato, S., Matsuyoshi, S., Kondoh, Y., 2008. Automatic assessment of japanese text readability based on a textbook corpus. In: LREC.

Sheehan, KM., 2016. A review of evidence presented in support of three key claims in the validity argument for the textevaluator text analysis tool. ETS Research Report Series.

Sinha, M., Basu, A., 2014. A study of readability of texts in bangla through machine learning approaches. Educ. Inf. Technol., 1–24

Spearman, 2016. URL https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php. (Accessed October 14 2016).

Vives, M., Young, L., Sabharwal, S., 2009. Readability of spine-related patient education materials from subspecialty organization and spine practitioner websites. Spine 34, 2826–2831.

WAVE, 2016. URL http://wave.webaim.org/. (Accessed December 15 2016).

Webanalytics, 2016. URL http://www.webanalytics.gov.in/. (Accessed September 13 2016).

Webpagefx, 2016a. URL http://www.webpagefx.com/tools/read-able/. (Accessed September 15 2016).

Webpagefx, 2016b. URL http://www.webpagefx.com/tools/read-able/coleman-liau-index.html. (Accessed December 13 2016).

Yamasaki, T., Tokiwa, K.-I., 2014. A method of readability assessment for web documents using text features and html structures. Electron. Commun. Jpn. 97, 1–10.