



Automated Readability Assessment for Spanish e-Government Information

Jorge Morato ^{1*}, Ana Iglesias ¹, Adrián Campillo ¹, Sonia Sanchez-Cuadrado ²

¹Computer Science Department, Universidad Carlos III de Madrid, Leganes, SPAIN

²Library and Information Sc. Dep., Universidad Complutense de Madrid, Madrid, SPAIN

*Corresponding Author: jmorato@inf.uc3m.es

Citation: Morato, J., Iglesias, A., Campillo, A. and Sanchez-Cuadrado, S. (2021). Automated Readability Assessment for Spanish e-Government Information. *Journal of Information Systems Engineering and Management*, 6(2), em0137. <https://doi.org/10.29333/jisem/9620>

ARTICLE INFO

Published: 21 Jan. 2021

ABSTRACT

This paper automatically evaluates the readability of Spanish e-government websites. Specifically, the websites collected explain e-government administrative procedures. The evaluation is carried out through the analysis of different linguistic characteristics that are presumably associated with a better understanding of these resources. To this end, texts from websites outside the government websites have been collected. These texts clarify the procedures published on the Spanish Government's websites. These websites constitute the part of the corpus considered as the set of easy documents. The rest of the corpus has been completed with counterpart documents from government websites. The text of the documents has been processed, and the difficulty is evaluated through different classic readability metrics. At a later stage, automatic learning methods are used to apply algorithms to predict the difficulty of the text. The results of the study show that government web pages show high values for comprehension difficulty. This work proposes a new Spanish-language corpus of official e-government websites. In addition, a large number of combined linguistic attributes are applied, which improve the identification of the level of comprehensibility of a text with respect to classic metrics.

Keywords: readability, e-government, information, assessment, web pages, accessibility, authoring tools

INTRODUCTION

One of the missions of today's governments is to facilitate access to administrative services. To achieve this objective, the European Union has encouraged and promoted this digital transformation. Since the beginning of the 21st century, Spain, as a member of the European Union, has developed several initiatives to make public service information available on government websites. Spain is slightly above the European Union average in the Digital Economy and Society Index, DESI, published by the European Commission (European Commission, 2020). This report draws attention to persistent deficiencies in human capital indicators and basic digital literacy skills. This fact has implications in terms of a low usage rate of digital services compared to the wide range of services provided by central and local governments (Morato et al., 2016).

Despite the high number of Internet users, low usage rates of digital services are detected, as seen in the DESI report aforementioned. The digital divide, and the low rates of use of digital services, may be due to different reasons. Some of these reasons are lack of digital literacy skills, poor infrastructure or reluctance about digital security and the validity of the procedures involved. Another possible explanation for the limited use of digital services is the lack of understanding of the information available to conduct administrative procedures (Morato et al., 2016). These administrative procedures and formalities can be very complex. According to the OECD report, Spanish population between 16 and 65 years old reaches low text comprehension rates (OECD, 2016). On the websites, this low readability is often combined with usability and accessibility issues. We think that it is due to the complexity of some procedures, combined with issues related with the scarce readability, accessibility and usability the reason for the proliferation of external help websites for explaining these tasks. These help sites explain, in a simple way, the most problematic and frequent administrative procedures carried out by citizens. The mere existence of external help websites clearly shows that citizens do not properly understand the information provided on e-government websites. This is a factor that shows the need to provide editing tools to avoid understanding issues on official websites.

Certainly, reading is a complex cognitive activity, in which personal factors might harm comprehension. In recent years, various organizations have been promoting guidelines that encourage clear, understandable, unambiguous language to provide universal access for all citizens. Not only in terms of classic web accessibility criteria, but also in terms of cognitive accessibility. These initiatives are included in different laws and regulations around the world. For instance, in Europe, the *European accessibility act*; in Spain, the *Information Society Services and Electronic Commerce Act* (Spanish Law 34/2002) and the Law 19/2013 on the Law

of Transparency, Access to Public Information and Good Governance (Spanish Law 19/2013); in the United States of America, *Section 508 Amendment* to the *Rehabilitation Act* of 1973, etc. All of them protect the rights of people with disabilities and promote the accessibility of the public information technology, electronic products (documents, webpages, etc.) and services based on guidelines and standards such as the *World Wide Web Consortium (W3C)* guidelines.

Easy-to-read guidelines is another example of these efforts to facilitate the understanding of people with cognitive, sensory and physical disabilities. In Spanish, the Easy-to-read, guidelines and recommendations for the elaboration of documents (UNE 153101:2018 EX) is a significant breakthrough to facilitate this adaptation to easy-to-read texts. Another relevant initiative in this regard is the WCAG 2.1. WCAG suggests a number of guidelines to different cognitive profiles on the Web.

The European Commission Clear Writing for Europe Conference (European Commission, 2019) and the Commission Style Guide (European Commission, 2020b) seek to establish a basis for improving the readability of government texts at European level. In general, these are guidelines or recommendations based on classic comprehensibility features, such as recommending short sentences and words. These worldwide efforts promote legislative initiatives to encourage governments to improve the readability of legal documents and e-government procedures. One relevant law that exemplifies these efforts is the US 2010 Plain Writing Act. This law establishes that federal agencies must use a language that the public can understand and use.

RELATED WORKS

One of the first attempts to provide guidelines to improve readability was provided by Dale and Chall (1949). In their work, these authors proposed a definition of readability that includes the different interactions that favour understanding and reading at optimal speed. Klare (1963) defined "Readability" as the ease with which a reader can understand a document due to the style of writing. In these first approaches the focus was in the vocabulary and the grammar. With the popularization of web pages, other aspects related to the visualization and organization of texts took prominence in the studies, to the detriment of readability. Legibility, text organization or design were emphasized over the readability, in accessibility and usability studies. In recent years, these areas are showing that its improvement is not possible without considering readability.

The readability of the text does not depend only on the grammar or vocabulary, but there are other parameters that must be considered. For example, quantum physics is inherently more complex than other topics. Thus authors, such as DuBay (2006), take into account the perspective of the users. That is, they consider aspects such as the user's previous knowledge of the subject, reading ability, interest and motivation in the subject, or physical and cognitive characteristics.

Since the mid-twentieth century, a need has prevailed in this area; this is how readability can be measured. Different formulas for automated evaluation have been developed (Dale and Chall, 1949; Fernandez-Huertas, 1959; Kinkeid, 1975). These studies focused on establishing thresholds corresponding to different educational levels. Therefore, the corpus used to tune these thresholds were the texts used in education, mainly schoolchildren texts (Lijun, 2011).

Classical metrics were developed mainly in English. Then, those were adapted into different languages, including Spanish (i.e. Inflesz, μ legibility (μ), Fernández-Huerta (FH) or Szigriszt-Pazos (SP)) (Morato et al., 2018). The metrics were basically based on the length of sentences, paragraphs and words. This approach is a good approximation, but as can be seen it is not easy to apply the same metrics to different domains and population groups. As an example, it is clear that molecular biology has longer words than other fields. That is, different fields need different thresholds. Dale and Chall (1949), stress the importance of vocabulary, and in fact the words most often used are usually shorter. But an approach based only on these features has the risk that if the phrases are in a random order they can be judged to be as difficult as a passage that makes sense (Benjamin, 2012).

Readability assessment is not a trivial task, there are many interrelated variables, for instance, the number of words that a person knows from a certain text has been established as one of the most determinant factors in readability (Schmitt et al., 2011). Accordingly, the difficulty of words established by their occurrence in texts is used to assess the readability of documents. It depends on the language and the topic of the text, so for the authors, the usage of simple formulas may differ from one person to another. tools. The use of statistical language modeling supports that different word usage could give accurate predictions with language models (Kincaid et al., 1975).

Other proposals use methods inspired by cognitive science like semantic analysis. For example, Latent Semantic Analysis (LSA) (Lauduer et al., 1998), or modern artificial intelligence is being used to assess text readability building models based on different parameters, which provide higher accuracy rates (Mohammadi and Khasteh, 2019).

Automated readability assessment has many different scopes, approaches and study areas. Recently, an important number of those works are related to medical research, specifically into informed consents (Kauchank and Hogue, 2017; Leroy and Endicott, 2012; Venturi et al., 2015). In terms of public information, regarding cultural dissemination, readability has been studied from a user perspective focusing on information panels (Morato et al., 2018; Serna et al., 2018).

The number of projects to facilitate readability is increasing. Two projects are particularly noteworthy. On one hand, the Capito project, the goal of this project is to simplify German texts. The simplification levels are from A1 to B1 of the Common European Framework of Reference for Languages. On the other hand, there is the FALC project, a work applied to French has been proposed, which simplifies the text according to easy reading guidelines. As can be seen, projects are usually language-dependent.

Table 1. Corpus statistics and readability indexes

	easy-to-read document	difficult-to-read document
Number of documents	133	115
Words average	1023.2	466.6
#Sentences avg	30.7	11.5
Fernandez-Huerta index	61.6	50.1
μ index	50.7	47.9

For a long time, one of the solutions proposed to improve readability has been to paraphrase the text to a simpler version, a process called simplification. This line is followed by the Spanish, in this case the application tries to reduce the length of sentences, avoiding unfamiliar words. An unresolved challenge for these systems is to avoid automated change of meaning. This problem is particularly relevant when the subject matter of the documents is regulations or e-government documents. In this case the change of terms to a synonym, in the use of common language, is not possible due to the restrictions of meaning in legal documents.

METHODOLOGY

This work aims at the automated evaluation of readability applied to Spanish e-government information from the websites. An analytical methodology is applied to identify the indicators that are most decisive for measuring the level of comprehensibility of a text.

The following sections detail the steps followed for this investigation:

1. **Corpus Collection:** one of the main challenges of this research is the lack of corpus of significant size in Spanish language. from e-Government information and annotated according to its readability level. As it has been stated previously the source and language of this corpus is relevant. Therefore, a new corpus was collected for this research work. Section 2.1 details the corpus compilation.
2. **Readability assessment based on traditional metrics.** Traditional metrics have been applied to predict the readability level of the Spanish e-Government web-pages.
3. **Identification and Annotation of linguistic features.** Different linguistic features have been implemented and annotated in the corpus' documents using language processing tools.
4. **Readability Prediction.** Finally, automatic learning methods are used to determine the most influential features which influence the readability of e-Government documents.

Corpus

For the study, in order to collect a corpus related to the Spanish e-Government public information from scratch, and annotate the documents with readability levels, the next hypothesis was defined: if a web-page needs to be clarified, it is because its readability level is not good enough.

Therefore, on the one hand, help websites were analysed in depth. The main goal of these websites is to clarify the administrative procedures of the public administration. Finally, two websites were chosen due to the way they write, clarify, and complete the administrative procedures. On these sites the information is clear and they seem to follow some of the easy-reading techniques or smart reading of Dubai (2007), such as content curation, organization, and completeness. The websites chosen were: 1) <https://www.adminfacil.es> and 2) <https://loentiendo.com>.

A total of 133 documents related to public administrative procedures were collected from these websites and they were annotated as easy-reading documents, with a good readability level.

On the other hand, we look for the same administrative procedures (the counterpart), but in the Spanish Government websites. These documents are annotated as difficult to read and understand (bad readability level). A total of 115 documents were found, where one or more *easy-to-read* documents usually correspond to one *difficult-to-read* counterpart (sometimes the same public administrative procedure was explained in both websites).

The corpus was compiled in June 2019 and consists of 258 documents classified as *easy-to-read* or *difficult-to-read* according to their origin (official website or procedures explanation webpage). **Table 1** shows the learning corpus statistics. Easy documents have significantly higher word count and more sentences compared to documents classified as difficult. Moreover, administrative procedures are split in more than one page in the easy-to-read set. The two factors, more text and a better organization seems to help to the understanding of these procedures.

However, the readability level according to the classical indexes do not totally agree with our classification (and our hypothesis): Fernandez-Huerta index classifies as 'normal' (61.6) our *easy-to-read* set of documents, and μ index classifies them as 'difficult' (50.7). The *difficult-to-read* documents are classified as 'somewhat difficult' (50.1) by Fernandez-Huerta index and 'difficult' (47.9) by μ index.

During the study, other linguistic features are going to be considered to learn which are the factors which most influence this classification.

Tools and Resources

In this work, a text analyzer based in natural language processing (NLP) was used for the accurate identification of terms and sentences, as well as other linguistic features that could be relevant in the readability assessment of the documents.

For corpus processing, an open-source package from Language Analyzer (Padró, 2011) has allowed a linguistic analysis of the whole text with which the characteristics of the texts have been extracted (number of sentences, POS tagging, punctuation mark, number identification, classification of named entities, multiple word detection, etc.).

Current Spanish for Reference Corpora (CREA by its acronym in Spanish) from the Royal Spanish Academy (Real Academia Española, 2020) has been consulted for establishing a term frequency level.

When all data was gathered, supervised machine learning methods (Witten et al., 2011) have been applied for the analysis of the indicators that affect over text readability. Classification algorithms provide mechanisms to detect the better linguistic features for the best classification of e-government texts, according to their difficulty level.

Indicators

In this study, we are going to focus on the content, although other aspects play a role in readability like the aforementioned related to users' perspective or other related with the document design and organization. A set of 19 different indicators has been studied to analyze the readability level: number of sentences, μ index, Fernandez-Huerta index, percentage of verbs in infinitive, participle or gerund forms, nouns, proper nouns, determinants, prepositions, number of commas per sentence, verbs per sentence, total number of words, word frequency range, long sentence number per document (where a sentence is considered long if it has more than 20 words) and long words number (where a long word is one that is longer than four syllables), both metrics according to (Freyhoff, et al., 1998), and TF-IDF (Ramos, 2003).

Word frequency range indicator comes from CREA. The list of frequencies of terms has been divided into three ranges: Very low frequent terms; medium frequent terms; and very frequent terms. The number of words in each document that are present in each range has been calculated.

Also, two traditional metrics have been used to assess the level of difficulty: Fernandez-Huerta Index and μ index.

Fernandez-Huerta Index (1) is the formula adapted to Spanish from the Flesch-Kincaid Index Fernandez-Huerta (1959):

$$206.84 - 60 * \text{syllables/words} - 102 * \text{sentences/words} \quad (1)$$

And μ index, whose formula is (2):

$$\mu = (n/n - 1) * (X/6^2) * 100 \quad (2)$$

Where X is mean, σ^2 is variance, and n is the number of words in the text (Muñoz, 2006).

Process

First, the corpus was processed (Padro, 2011) to identify and extract the linguistic features explained in the previous section. Then with machine learning methods and using classification, a model was learned using the relevant linguistic features for the readability assessment.

The machine learning algorithm selected for the model detection has been the supervised classification algorithm J48, which generates a decision tree and provides the researchers a simple and visual path based on heuristics to check the algorithm effectiveness from a linguistic perspective. The reason to select this algorithm is that best results have been obtained with the C4.5 algorithm. This algorithm was proposed by Quinlan, and implemented as J48 in Weka (Witten et al, 2011). Machine learning has been analyzed according to difficulty classification due to documents source, assuming that the existence of a counterpart which tries to explain and clarify the public procedure means that those are more difficult to read and understand. A representative set of documents has been considered for the results validation.

RESULTS

From the set of 19 attributes tested, a total of 10 attributes were selected by the model learned with J48 algorithm:

1. Number of sentences
2. μ index
3. Fernández-Huerta index
4. Percentage of infinitive verb
5. Number of words in the document
6. Percentage of prepositions
7. Verbs/phrase ratio
8. Percentage of terms in the middle range of CREA list
9. Long sentence number per document
10. TF-IDF

Table 2. Readability indexes comparison

Metrics	Measures		
	Precision	Recall	F-measure
μ index	0.746	0.722	0.709
Fernandez-Huerta	0.699	0.698	0.698
Proposed model	0.795	0.794	0.793

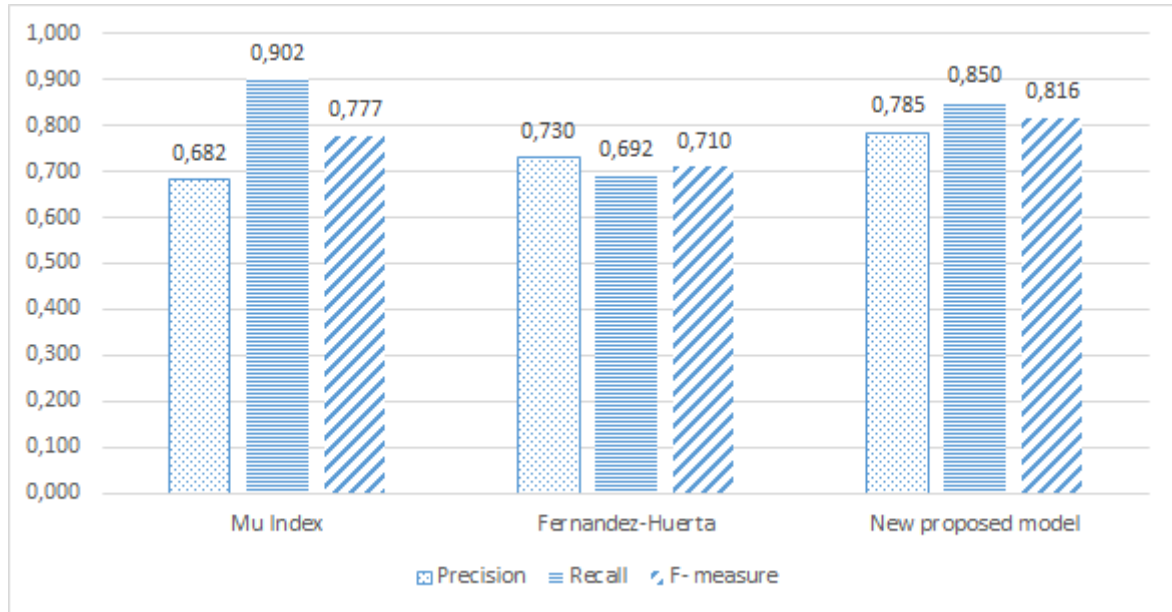
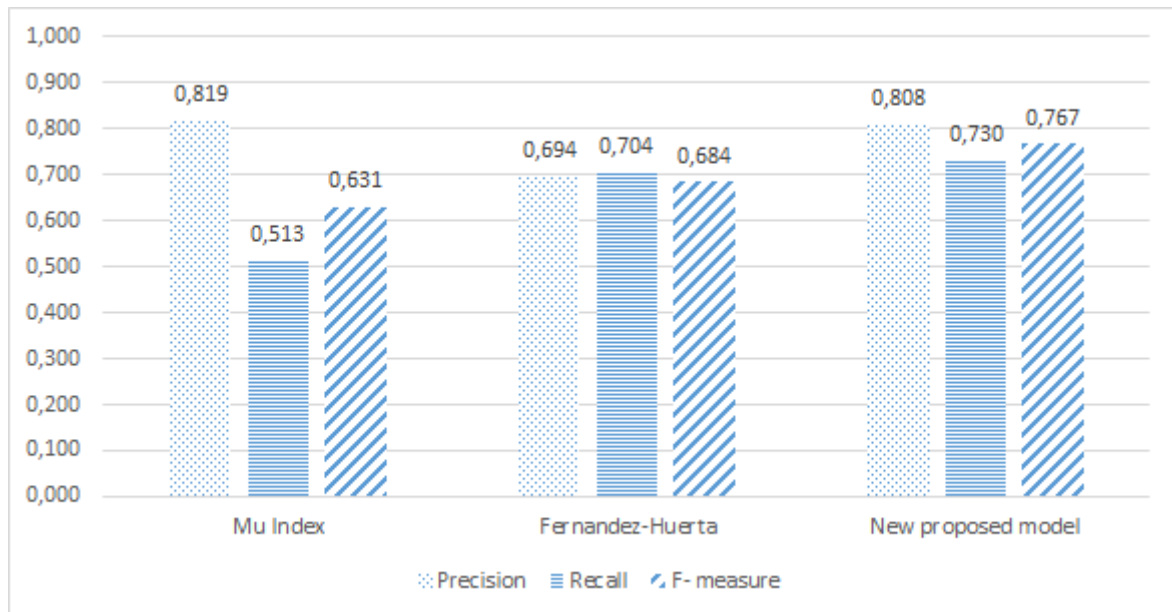
**Figure 1.** Precision for easy documents**Figure 2.** Precision for difficult documents

Table 2 collects obtained data for each of different traditional indexes in precision-recall terms. It is observed how indexes based on traditional metrics (μ and Fernandez-Huerta readability indexes) obtain moderate values, while the index proposed in this paper (called “*proposed model*”) and funded on linguistic attributes reach a higher rate in precision-recall terms (79.43% of success versus the 20.56% of error).

We conclude that those new metrics contribute to a more accurate establishment of which variables affect positively over text readability of government documents.

In **Figure 1** and **Figure 2** it is shown how different are the model performance when only traditional readability indexes are considered instead of using a higher set of linguistic attributes, for both the easy and difficult documents.

Figure 1 shows precision, recall and F-measure of the three readability models, for the easy-document set. Two classic models are included for Spanish language, as they are the μ (μ) index and Fernandez Huerta index, and are compared with the *proposed*

Table 3. Experiments using analogous features

Feature	Research
#sentences	This feature has shown its relevance in different languages: Swedish (Larsson, 2006); French (François and Fairon, 2012), or English (Zeng-Treitler et al., 2007)
Legibility Index Mu and Fernandez-Huerta	These classic formulas and related parameters are often found relevant: English (Leroy and Endicott, 2012; Zeng-Treitler et al., 2007); Italian (Venturi et al., 2015) or Spanish (Serna, 2018).
% Infinitive verbs	A survey is discussed in English (Kauchak et al., 2017) or Swedish (Larsson, 2006)
% Participles	Reported in Portuguese (Curto et al., 2015) or English (Lijun, 2010)
% Determinants	For instance, in (Zeng-Treitler et al., 2007)
Comma/sentence ratio	This criterion is related to others to analyze the syntactic simplicity, in English (Leroy and Endicott, 2012) or French (François and Fairon, 2012)
Ordinary words frequency	This is one of the most important factors in literature. Spanish (Morato, et al., 2018; Serna et al., 2018); English (Leroy and Endicott, 2012)
TF-IDF	tf-idf to determine word relevance in document queries (Ramos, 2003).
Number of sentences longer than 20 words	Easy-reading information (Freyhoff et al, 1998)
Words longer than 4 syllables	Easy-reading information (Freyhoff et al, 1998)

model. Overall values of F measure show that the model proposed is more accurate on precision-recall rates for the easy documents set. The lowest accuracy level is related to the Fernandez-Huerta model.

In **Figure 2**, the results of precision-recall for the difficult documents set are shown. As can be observed slightly lower rates of accuracy in every case, but especially in the μ index. This index shows a tendency to assign the documents as easy. The *proposed model* presents the highest accuracy rate for determining difficult documents.

A representative set of documents has been considered for the results validation, out of the corpus which created the model, applying cross-evaluation.

But, are the linguistic features selected by the machine learning algorithm relevant from a linguistic point of view? Could we explain this result from a linguistic perspective? **Table 3** shows examples of readability studies that agree on the relevance of the features selected by our model. These studies cover different corpus and languages.

Note that not all the features identified as relevant in the literature are present in our model. An enumeration of these other features can be found in two comprehensive surveys (Morato et al., 2018; Ojha et al., 2018). In Ojha et al. (2018), the research is focused on the differences due to different audiences and corpus, showing how features and thresholds vary in different studies. For instance, a research conducted in Spanish (Morato et al., 2018) found that the quantity of proper nouns is relevant to understand informative panels in a museum, but in e-government documents these proper nouns occur on rare occasions, therefore this criterion is absent in our linguistic model.

About the accuracy of the model, mention that there are not specific studies carried out in Spanish over this type of documents. Different studies show an F-measure value from 80% to 98%, although they are hardly comparable. For instance, in Portuguese a research tested 12 classifiers for linguistic features reporting an accuracy of 81.4% with second language learning materials (Curto et al., 2015). For informative panels in Spanish (Morato et al., 2018) is reported an 89%, the same value that Larsson (2006) obtained with Swedish news.

New experiments performed adding three new attributes show us a very interesting way to continue researching looking for better attributes to predict readability, from the previous article (Campillo et al., 2020).

One of the variables included in the model is the TF*IDF (Ramos, 2003) value for every text in our corpus, in terms of percentage to limit the values and make more accurate this attribute no matter how different our corpus is between one document to another.

The other two attributes included in this new model, are more focused on our future work. Our work is mainly focusing on readability and accessibility so we have been checking different rules to make documentation not only easier to understand, but also more accessible for everyone. This is why we included; number of long sentences (more than 20 words) and number of long words per document (more than 4 syllables) (Freyhoff et al., 1998). Advantages offered by including easy-to-read guidelines are twofold: on the one hand, there is a clear improvement in readability; on the other hand, there are benefits in terms of accessibility. As a result, the audience is larger, due to the inclusion of more profiles.

CONCLUSIONS AND FUTURE WORK

Nowadays, different initiatives promoted for e-government increase digital information all over the world. Unfortunately, not all the people can access the information published because accessibility barriers are found. That is why many governments encourage public administration to test and pay attention to the accessibility of web pages.

This research is focused on one of the accessibility factors: the readability. The paper aims to fill a gap in the literature by proposing a new automated readability assessment model of Spanish government electronic websites. In the Spanish language, traditional readability measures have been based on naïve models that only consider counting sentences per paragraph or words per sentence. However, vocabulary and word length factors among other linguistic factors are definitely important. They are not robust enough to measure the degree of readability by themselves, but its implication in the readability assessment could be useful. This paper studies the relevance of 19 linguistic features in the readability assessment, learning with machine learning algorithms that only 10 of them are necessary for the model: number of sentences; Mu index; Fernández-Huerta index; percentage

of infinitive verb; number of words in the document; percentage of prepositions; verbs/phrase ratio; percentage of terms in the middle range of CREA list; long sentence number per document and TF-IDF.

The results show that this new model proposed for automated readability assessments improves the results obtained by the classical readability measures, proving that the inclusion of new linguistics features is beneficial for the automated readability assessments.

Another contribution of the papers is the compilation of a corpus for readability assessment focused on Spanish e-Government texts and generalized to the whole population.

Currently, we are working on validating the model learned by enrolling readability experts and final users in the experimentation. Moreover, we are working on improving the creation of authoring tools to help the webpage developers and writers to improve the readability of their texts, by detecting the main problems and making recommendations based on the easy-to-read and other guidelines. Furthermore, we are combining linguistics metrics with semantic analysis and web-based readability metrics to improve readability on the Internet. In future works we plan to increase the corpus and test more features, as well as adapting the texts to distinct audiences and domains.

ACKNOWLEDGEMENT

Work supported by the Spanish Ministry of Economy, Industry and Competitiveness, (CSO2017-86747-R). We also want to thank Plena Inclusion Association for their advice.

REFERENCES

- Benjamin, R. G. (2012). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 24(1), 63-88. <https://doi.org/10.1007/s10648-011-9181-8>
- Campillo, A., Morato, J., Maqueda, A. I. and Sanchez-Cuadrado, S. (2020). Readability of Spanish e-government information. *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, Seville, 25-27 June, IEEE, 1-4.
- Capito project. (n.d.). Available at: <https://www.capito.eu/> (Accessed: 1 January 2021).
- Curto, P., Mamede, N. and Baptista, J. (2015). Automatic Text Difficulty Classifier - Assisting the Selection of Adequate Reading Materials for European Portuguese Teaching, in Markus Helfert, Maria Teresa Restivo, Susan Zvacek, James Uhomoihi (eds.) *Proceedings of the 7th International Conference on Computer Supported Education, INSTICC*, 23 - 25 May, 2015. Setubal: Scitepress, 1, 36-44.
- Dale, E. and Chall, J. S. (1948). A Formula for Predicting Readability. *Educational Research Bulletin*, 27(1), 11-28.
- DuBay, W. H. (2007). *Smart Language Readers, Readability, and the Grading of Text*. Costa Mesa: Impact Information.
- European Commission. (2019). *Clear writing for Europe Conference*. Available at: https://ec.europa.eu/info/sites/info/files/clear_writing_conference_notes_for_website.pdf (Accessed: 1 January 2021).
- European Commission. (2020a). *The Digital Economy and Society Index (DESI): Shaping Europe's digital future*. Available at: <https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi> (Accessed: 1 January 2021).
- European Commission. (2020b). *Commission Style Guide*. Available at: https://wikis.ec.europa.eu/download/attachments/6824833/commission_style_guide.pdf?version=1&modificationDate=1594633342434&api=v2 (Accessed: 1 January 2021).
- FALC project. (n.d.). Available at: <https://www.ideographik.org/communication/> (Accessed: 1 January 2021).
- Fernández-Huerta, J. (1959). Medidas sencillas de lecturabilidad [Simple readability measures]. *Consigna*, 214, 29-32.
- François, T. and Fairon, C. (2012). An "AI readability" formula for French as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL '12)*, Jeju (Korea), July 12-14. Stroudsburg, PA: Association for Computational Linguistics, 466 -477.
- Freyhoff, G., Hess, G., Kerr, L., Menzel, E., Tronbacke, B. and Van Der Veken, K. (1998). *Make It Simple, European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability*. Brussels: ILSMH European Association.
- Kauchak, D., Leroy, G. and Hogue, A. (2017). Measuring text difficulty using parse-tree frequency. *Journal of the Association for Information Science and Technology*, 68(9), 2088-2100. <https://doi.org/10.1002/asi.23855>
- Kincaid, J., Fishburne, R., Rogers, R. and Chissom, B. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) For Navy Enlisted Personnel. *Institute for Simulation and Training*, 56. Available at: <https://stars.library.ucf.edu/istlibrary/56> (Accessed: 1 January 2021).
- Klare, G. R. (2000). The measurement of readability: useful information for communicators. *ACM Journal of Computer Documentation*, 24(3), 107-121. <https://doi.org/10.1145/344599.344630>
- Landauer, T. K., Foltz, P. W. and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259-284. <https://doi.org/10.1080/01638539809545028>
- Larsson, P. (2006). *Classification into Readability Levels. Implementation and Evaluation* (Master's Thesis), Uppsala University. Available at: <http://www.diva-portal.org/smash/get/diva2:131028/FULLTEXT01.pdf> (Accessed: 1 January 2021).

- Leroy, G. and Endicott, J. E. (2012). Combining NLP with Evidence-based Methods to Find Text Metrics Related to Perceived and Actual Text Difficulty. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 749-754.
- LEY 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno [LAW 19/2013, of December 9, on transparency, access to public information and good governance]. BOE, N. 295, 10th December. Available at: <https://www.boe.es/buscar/doc.php?id=BOE-A-2013-12887> (Accessed: 1 January 2021).
- LEY 34/2002, de 11 de julio, de Servicios de la Sociedad de la información del comercio electrónico [LAW 34/2002, of July 11, on Services of the Electronic Commerce Information Society]. BOE, n.166, 12th July. Available at: <https://www.boe.es/buscar/act.php?id=BOE-a-2002-13758> (Accessed: 1 January 2021).
- Lijun, F. (2010). *Automatic Readability Assessment* (Dissertation Ph.D.). City University of New York (CUNY). Available at: https://academicworks.cuny.edu/gc_etds/1934/ (Accessed: 1 January 2021).
- Mohammadi, H. and Khasteh, S. H. (2019). Text as Environment: A Deep Reinforcement Learning Text Readability Assessment Model. *ArXiv: 1912.05957* [Cs]. Available at: <http://arxiv.org/abs/1912.05957> (Accessed: 1 January 2021).
- Morato, J., Ruiz-Robles, A., Sanchez-Cuadrado, S. and Marzal García-Quismondo, M. A. (2016). Technologies for Digital Inclusion: Good Practices Dealing with Diversity. In B. Passarelli, J. Straubhaar and A. Cuevas-Cerveró (eds), *Handbook of Research on Comparative Approaches to the Digital Age Revolution in Europe and the Americas* (pp. 332-351). Hershey, PA: IGI Global <https://doi.org/10.4018/978-1-4666-8740-0>
- Morato, J., Sánchez-Cuadrado, S. and Gimmelli, P. (2018). Estimación de la comprensibilidad en paneles de museos [Measuring the readability of exhibit panels in museums]. *El Profesional de la Información*, 27(3), 570-581. <https://doi.org/10.3145/epi.2018.may.10>
- Muñoz Baquedano, M. (2006). *Legibilidad y variabilidad de los textos [Legibility and variability of texts]*. Universidad Playa Ancha de Ciencias de la Educación. Available at: <https://legibilidadmu.cl/1.pdf> (Accessed: 1 January 2021).
- OECD. (2016). *Skills matter. Further results from the survey of adult skills*. Paris: OECD Publishing. <http://doi.org/10.1787/9789264258051-en>
- Ojha, P. K., Ismail, A. and Kuppusamy, K. S. (2018). Perusal of readability with focus on web content understandability. *Journal of King Saud University - Computer and Information Sciences*, 32(10), 1221. <https://doi.org/10.1016/j.jksuci.2018.03.007>
- Padró, L. (2011). Analizadores multilingües en Freeling. *Linguamática*, 33(1), 13-20. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Public Law 111 - 274 - Plain Writing Act of 2010. Available at: <https://www.govinfo.gov/app/details/PLAW-111publ274> (Accessed: 1 January 2021).
- Ramos, J., Fawcett, T. and Mishra, N. (2003). Using TF-IDF to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, 2003 December. Piscataway, NJ, Vol. 242, pp. 133-142.
- Real Academia Española. (2020). *Real Academia Española*. Available at: <http://www.rae.es> (Accessed: 1 January 2021).
- Schmitt, N., Jiang, X. and Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1), 26-43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Serna, Y., Morato, J. and Sánchez-Cuadrado, S. (2018). Evaluación de la comprensión de los paneles interpretativos en parajes naturales [Assessment of understanding of interpretive panels in natural settings]. *Scire: Representación y organización del conocimiento-Scire: Representation and organization of knowledge*, 24(2), 53-62.
- Simplext project. (n.d.). Available at: <http://simplext.taln.upf.edu/> (Accessed: 1 January 2021).
- Venturi, G., Bellandi, T., Dell'Orletta, F. and Montemagni, S. (2015). NLP-Based Readability Assessment of Health-Related Texts: a Case Study on Italian Informed Consent Forms. In Cyril Grouin, Thierry Hamon, Aurélie Névéol, Pierre Zweigenbaum (eds) *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, LIMSI-CNRS, 17 September 2015, New York: Association for Computational Linguistics, pp. 131-141. <https://doi.org/10.18653/v1/W15-2618>
- W3C (2018). *Web Content Accessibility Guidelines 2.1*. W3C World Wide Web Consortium. Available at: <https://www.w3.org/TR/WCAG21/> (Accessed: 1 January 2021).
- Witten, I. H., Frank, E. and Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques* (3rd ed). Amsterdam: Elsevier, Morgan Kaufmann.
- Zeng-Treitler, Q., Kim, H., Goryachev, S., Keselman, A., Slaughter, L. and Smith, C. A. (2007). Text characteristics of clinical reports and their implications for the readability of personal health records. *Studies in Health Technology and Informatics*, 129(Pt 2), 1117-1121.