Improving Accessibility in Public Web Pages

César Domínguez[1], Jónathan Heras[1], Félix Lanas[1], Gadea Mata[1], Julio Rubio[1] and Mirari San Martín[1]

[1]Department of Mathematics and Computer Science, University of La Rioja, Spain

Abstract

The accessibility of web pages from public institutions is important in ensuring equal access to information and services for all individuals, including those with disabilities. In this project, we aim to improve the accessibility of the web page from the Government of La Rioja. In particular, we are focused on three aspects that involve applying different natural language processing techniques. First, we will automatically caption all the images from the web page; second, we will provide transcriptions of all the videos from the Government of La Rioja; and, finally, we will improve the readability of the contents of the web page. In summary, this project is a first step towards making web pages from public institutions adaptable to the needs of each particular user that visits them.

Keywords

Web Accessibility, Image Captioning, Video Transcription, Readability, Text Simplification

1. Introduction

With the widespread use of the Internet, information is increasingly within the reach of the entire population in a more direct manner. This has made it easier for public administrations to share all kinds of information and documentation through their web sites, ranging from news about specific events to specific regulations. However, the immediate access to information by the public should not be confused with how accessible the information actually is.

In the context of this project, which concerns web pages, we define *accessibility* as making a web site usable by the entire population, enabling navigation and interaction with the web while ensuring that everyone can perceive and understand the information. Accessibility in this context should be understood in a broader sense than it is usually given, which is often associated with people who have some specific difficulty in accessing information. In reality, each person visiting a web page has particular interests, which may vary from one interaction to another for the same individual. For example, a person might seek a quick overview of the contents of a web page regarding an administrative procedure, but the same person might be interested in accessing the complete regulatory information later. The ideal situation we should strive for is one where a web portal adapts, at the beginning of each connection, to the accessibility needs of each user.

There are protocols aimed at improving the accessibility of web portals, such as the W3C web standards [1]. These standards or recommendations are designed to make a web page universal, accessible, user-friendly, and trustworthy for everyone. Some of the standards are related, for example, to the graphical representation of content on web pages, stating that images must be described through their "alt" attribute, or that videos should have subtitles.

Furthermore, according to the Spanish General Audiovisual Communication Law 13/2022[1], dated July 7, in Title VI, Chapter II on Accessibility, it is established that the accessibility of audiovisual communication services provided by any provider of this service must be improved. Specifically, within this scope, the Universal Accessibility Law of La Rioja (10L/PL-0017) has been recently approved (25/01/2023) in La Rioja. In this autonomous community, this project is developed[2]. This law establishes provisions regarding accessibility in communication for media outlets dependent on the Autonomous Community of La Rioja, but undoubtedly, its principles can be extended to other areas.

2. Goals

With the general objective of making more accessible the information of web portals with relevant information for citizens, for instance the web site of a public administration, this project is divided into three parts:

SEPLN-PD 2023: Annual Conference of the Spanish Association for Natural Language Processing 2023: Projects and System Demonstrations

cesar.dominguez@unirioja.es (C. Domínguez); jonathan.heras@unirioja.es (J. Heras); felix.lanas@unirioja.es (F. Lanas); gadea.mata@unirioja.es (G. Mata); julio.rubio@unirioja.es (J. Rubio); mimartla@unirioja.es (M. San Martín)

0000-0002-2081-7523 (C. Domínguez); 0000-0003-4775-1306 (J. Heras); 0000-0002-5567-8463 (G. Mata); 0000-0002-4282-3692 (J. Rubio)

[1]Ley 13/2022, 7 de julio, General de Comunicación Audiovisual

[2]Ley de Accesibilidad Universal de La Rioja (10L/PL-0017)

- `Part I`: Image accessibility.
- `Part II`: Video accessibility.
- `Part III`: Easy read and Plain Language.

This project aims to investigate how Deep Learning-based language models can help us to approach this situation of maximum information accessibility.

3. State of the art

In this section, we detail the framework and the previous work available in the literature for each part of our work.

3.1. Image accessibility

Usually, the description of images is a task conducted by a human annotator responsible for describing the image's content and obtaining the appropriate metadata to facilitate accessibility to users. However, in public portals of the administration, it is common to find images that are already published but, due to their high number, it is unfeasible to process them individually by a human annotator.

The generation of the description of an image is a problem that can be carried out by Artificial Intelligence methods from two fields: on the one hand, computer vision methods to acquire the image content; and on the other hand, language models to convert the image content into meaningful words in a correct order [2]. Deep Learning methods have proven to be useful and provide good results in generating descriptions for images in specific cases [3]. This type of method, such as Deep Learning techniques applied to images, has already been put into practice, in particular, by this work team and, in general, by the research group within the University of La Rioja [4] and [5].

In order to be able to apply this type of method, it is necessary to carry out a first step of analysis of the initial data. For web portals under investigation, the set of images is usually large and exceeds 200,000 records. This large volume of images requires a preprocessing step to filter the images and, for example, remove those that are repeated or those that are thumbnails of larger images.

Some of the techniques to tackle this problem are based on using the model known as VirTex [6]. This model is accessible at [7] and is trained with images obtained from the web site https://redcaps.xyz/, which collects images with descriptions. This model was created to add captions to images automatically; see Figure 1.

Other known methods that may be useful in the development of this activity are those based on the use of Convolutional Neural Network (CNN) [8] and Transformers, a method commonly used for natural language processing [9]. For this purpose, the *Keras* library is used, accessible through the reference [10]. More recently, models such as BLIP-2 [11] have emerged, which allow information to be extracted from images.

Figure 1: A herd of sheep (description of the model) - Sheep on a cattle track (description given).

In previous works of the group, that are in a proof of concept stage, we have explored Deep Learning mechanisms to label authentic images extracted from an active portal, obtaining encouraging results and partial successes but also showing certain shortcomings. In addition, this type of model is mostly trained for English texts, and therefore their success is reflected in this language, but we work with Spanish texts.

3.2. Video accessibility

The action of extracting text from audio is known as Automatic Speech Recognition (ASR), the process by which a computer system identifies spoken words [12]. We use this process in our daily lives, as it is used in many areas, such as voice assistants or dictation systems. For this purpose, we work with Recurrent Neural Networks (RNN) [13], which are neural networks with memory that work with sequential information. Each neuron that forms it represents a temporal moment and can pass the collected information to the next one.

There are different ASR classes depending on the spoken text, like single words, connected words, continuous speech, and spontaneous speech. The class that is of most interest for this activity is continuous speech. In addition, there are two types of models: acoustic models, which extract information such as the speaker's gender and dialect, and language models, which deal with what constitutes a possible word, which words may go together, and in which sequence.

The activity of generating subtitles of a video can be split into four parts:

- Separating the audio from the image.
- Transcribing the audio to text.
- Aligning the transcribed text with the audio.
- Reassembling the video (image + audio + text).

The second step requires studying known models for creating transcribed text. Some of these models, such as Wav2Vec or PocketSphinx, are pre-trained models for both English and Spanish, but there are others that have only been pre-trained for English. The latter requires dealing with the transition between English and Spanish since these models do not have a direct application in videos with Spanish audio. The study and subsequent evaluation of the different models will result in selecting a model to be used with the different videos to obtain the text of their subtitles.

Once we have the text transcribed correctly, we must proceed to assemble it with the video — in this phase the problem of synchronizing the audio with the text appears. For this step, we must take into account that each person speaks at a different speed, so we cannot standardize the alignment. To solve this problem, we can use a neural network known as Connectionist Temporal Classification (CTC) [14], which is used to train Recurrent Neuronal Networks (RNNs), which, as mentioned above, are a very useful tool in problems such as speech recognition or handwriting. This kind of network allows us to take into account the context and not only the last step.

The technique described here are in constant evolution, and, in addition to the aforementioned methods, in September 2022, a new ASR system was published known as Whisper [15]. Using this model, in a previous work by the group, we have explored the feasibility of captioning all videos present on a live web portal. From this first proof of concept, it has been inferred that this is a computationally expensive endeavor, but is feasible once a sufficiently efficient model has been generated.

3.3. Easy Read and Plain Language

Plain Language and Easy read are terms that should not be confused; so we define them as follows.

Easy read is a way of adapting information to make it easier to read and understand text for people with reading difficulties. It is a method of adaptation with simple and clear language. Easy read is aimed at a number of groups with certain reading comprehension difficulties. Some of them are the following:

- People with learning difficulties (such as dyslexia).
- People with low literacy or little schooling.
- Foreigners or immigrants who do not have a good understanding of the Spanish language.
- Children who need reading reinforcement.
- Deaf people with comprehension difficulties.
- Elderly people with mental disorders.
- People with hyperactivity and attention deficit disorders.
- People with intellectual or developmental disabilities (such as autism, aphasia, and so on).

In order to make Easy-to-Read adaptations of any text, a number of guidelines must be followed. One of the first documents on how to produce Easy-to-Read text was published by IFLA (International Federation of Library Associations and Institutions) [16]. There is a second document, which was produced by several Europe organizations, under the title "Information for All" [17]. In addition, there are some adaptations classified by *Plena Inclusión* [18].

Easy read does not have a fixed standard, but there are proposed different levels since it is impossible to adapt a text in the same way for all people with difficulties, as these are very diverse. This adaptation refers to both text and images or any other element that can be incorporated into a document (graphics, diagrams, and so on). IFLA establishes three levels, similar for both original Easy Read and adapted Easy Read works:

- `First level`. In this simplest level, there are many images and little text. The text has a low syntactic difficulty.
- `Second level`. In this intermediate level, less simple than the previous one, text is written using a vocabulary and expressions that are known to everyone, and it is easy to follow and understand. Images are also used at this level.
- `Third level`. In the most complex levels, text are longer, unusual words are used, and jumps in time and space appear. At this level, there are few images.

A level of this classification is chosen according to the user to whom the adapted text is addressed.

On the other hand, we will use the definition given by the International Federation of Plain Language in relation to the concept of clear text or plain language: "A communication is written in plain language if its wording, structure, and design are so transparent that the readers to whom it is addressed can find what they need, understand what they find and use that information".

Plain language benefits the general public, but it is not intended for people with comprehension difficulties, such as Easy read. Although there is no current regulation, an international ISO standard on plain text is being worked on, and it is supported by different organizations

that promote it (Clarity International[3], Plain Language[4] or the International Plain Language Federation[5]). Plain language has become widespread, especially in the administrative and legal fields, which handle texts that are aimed to the general public, but with highly technical language that is sometimes difficult to understand.

Many features of any text can be modified or transformed to make it more readable and understandable. The goal of adaptations to Easy read and/or Plain language is to transform complex sentences into simpler ones [19]. Adaptations also include the way the text is displayed.

Text simplification is usually based on the following four tasks:

- `Lexical simplification` aims to replace difficult words with easier words (synonyms) that are considered better to understand or read, as long as the meaning is not altered.
- `Syntactic simplification` aims to transform long phrases or sentences containing syntactic figures that are unreadable or incomprehensible into simpler ones and in active form (passive form should be avoided whenever possible).
- `Eliminating information` consists in reducing phrases or sentences, keeping the essential information, and eliminating unnecessary details that do not add anything new to the idea to be transmitted.
- `Adding information` provides extra knowledge that may help the reader to understand and learn the meaning of one or more unknown terms.

These four simplification modes are related, and sometimes a mixture of them is needed to maintain coherence and obtain the final text.

There is a wide variety of software tools for Natural Language Processing (NLP) in different languages, such as NLTK, spaCy, or Scikit-learn. In this project, we will study these tools to apply them to Easy read and Plain language.

These libraries provide objects that help to represent text elements, such as sentences and words. They are used to extract information, as natural language understanding systems or for pre-processing text before applying Deep Learning techniques.

The main tasks to be performed are as follows:

- `Tokenization`.
- `Part-of-speech (POS) Tagging`.
- `Dependency Parsing`.
- `Lemmatization`.
- `Sentence Boundary Detection`.
- `Named Entity Recognition`.
- `Entity Linking`.
- `Similarity`.
- `Text Classification`.
- `Rule-based Matching`.

These tasks will allow us to pre-process the text in order to generate a model using Deep Learning methods with the objective of automating the creation of Easy read and Plain language.

[3]https://www.clarity-international.org/
[4]https://plainlanguagenetwork.org
[5]https://www.iplfederation.org

4. Conclusions

The presented project aims to study existing tools and propose new ones to address the improvement of accessibility in web portals that are of interest to the general public, such as the web sites of public administrations.

This project consists of three parts, which can be considered separately, but also can be seen as interconnected. For instance, both image descriptions and video subtitles should be understandable to anyone who reads them, and this implies the use of plain and easy to read text.

The ultimate goal is to have a web page that adapts the accessibility of the provided information based on the users visiting it, allowing them to choose the level of accessibility thatthey prefer when accessing the information.

Working team

This project is conducted by members of the Computer Science group of University of La Rioja. In particular, several members of this team provide their experience in the areas of Computer Vision and Natural Language Processing that are necessary to conduct this project.

Acknowledgments

The work was partially supported by the project PID2020-115225RB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and PID2020-116641GB-I00 funded by MCIN/AEI/ 10.13039/501100011033 as well as by transfer projects OTCA211018 and OTCA221110.

References

[1] W3C - World Wide Web Consortium, Accessibility, 2023. https://www.w3.org/.

[2] K. Doshi, Image captions with deep learning: State-of-the art architectures, 2001. URL: https://ketanhdoshi.github.io/Image-Caption/.

[3] A. Roy, A guide to image captioning., 2020. URL: https://towardsdatascience.com/a-guide-to-image-captioning-e9fd5517f350.
[4] D. Lacalle, H. A. Castro-Abril, T. Randelovic, C. Domínguez, J. Heras, E. Mata, G. Mata, Y. Méndez, V. Pascual, I. Ochoa, SpheroidJ: An open-source set of tools for spheroid segmentation, Computer Methods and Programs in Biomedicine 200 (2021) 105837.
[5] A. Inés, C. Domínguez, J. Heras, E. Mata, V. Pascual, Biomedical image classification made easier thanks to transfer and semi-supervised learning, Computer Methods and Programs in Biomedicine 198 (2021) 105782.
[6] K. Desai, G. Kaul, Z. Aysola, J. Johnson, Redcaps: web-curated image-text data created by the people, for the people, 2021.
[7] U. Vision, Image captioning with virtex model trained on redcaps, 2022. URL: https://huggingface.co/spaces/umichVision/virtex-redcaps.
[8] K. O'Shea, R. Nash, An introduction to convolutional neural networks, 2015.
[9] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45.
[10] M. Hodosh, P. Young, J. Hockenmaier", Framing image description as a ranking task: Data, models and evaluation metrics, Journal of Artificial Intelligence Research 47 (2013) 853–899.
[11] J. Li, D. Li, S. Savarese, S. Hoi, Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models, arXiv preprint arXiv:2301.12597 (2023).
[12] K. Doshi, Audio deep learning made simple: Automatic Speech Recognition (ASR), How It Works., 2021. URL: https://towardsdatascience.com/audio-deep-learning-made-simple-automatic-speech-recognition-asr-how-it-works-716cfce4c706.
[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.
[14] A. Hannun, Sequence modeling with CTC., 2017. URL: https://distill.pub/2017/ctc/.
[15] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, Robust speech recognition via large-scale weak supervision, 2022.
[16] International Federation of Library Associations and Institutions, Directrices para materiales de lectura fácil, 2012. URL: https://www.ifla.org/files/assets/hq/publications/professional-report/120-es.pdf.
[17] Inclusión Europa, Información para todos., 2016. URL: https://www.plenainclusion.org/publicaciones/buscador/informacion-para-todos-pautas-europeas-de-la-lectura-facil/.
[18] O. Garcia-Muñoz, Lectura Fácil - Métodos de redacción y evaluación, 2012. URL: https://www.plenainclusion.org/publicaciones/buscador/lectura-facil-metodos-de-redaccion-y-evaluacion/.
[19] X. Wan, Automatic Text Simplification, Computational Linguistics 44 (2018) 659–661.