Using Natural Language Processing to Improve Transparency by Enhancing the Understanding of Legal Decisions

Tiago **Lima**[1], Kellyton **Brito**[2], André **Nascimento**[1], George **Valença**[1] and Fábio **Pedrosa**[2]

[1]*Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n - Dois Irmãos, Recife - PE, 52171-900, Brazil*

[2]*Tribunal de Contas de Pernambuco, R. da Aurora, 885 - Boa Vista, Recife - PE, 50050-910, Brazil*

Abstract

The current technological advances and real-time worldwide communications transformed the government transparency scenario, with the popularization of open government data and transparency portals holding the premise of transforming government. However, much government data, particularly in the legal domain, is context-specific and is not understandable to ordinary people, thereby reducing its access and benefits related to its publication. Hence, the automatic generation of simpler translations and summaries of legal decisions, taking advantage of the advances in natural language processing (NLP), is a promising approach. This paper presents a design science study aimed to increase the transparency of a Brazilian court of accounts by enhancing the understanding of its legal decisions through NLP techniques such as text simplification and text summarization. We then discuss the approaches, challenges, and difficulties of developing artificial intelligence systems in this as yet unexplored domain, especially considering the Portuguese language.

Keywords

Natural Language Processing, Transparency, Legal Decisions, Portuguese, Brazil

1. Introduction

Since the 1950s [1], governments and society have agreed that transparency, "the right to know," and Open Government Data, may bring about many benefits. At the beginning of the 2010s, the movement resurfaced with the possibility of using Web 2.0 to publish and consume data, which would promote better public services, government efficiency, and effectiveness, increased accountability, citizen participation, engagement, and collaboration, and a decrease in corruption, among other benefits [2]. Additionally, the recent popularization and success of artificial intelligence (AI) solutions have revealed it as a new frontier technology for the public sector [3]. However, researchers have highlighted that the results of AI implementation in government are still unknown and unexpected [4].

One of the challenges is that it is not sufficient simply to publish data and information if ordinary people are not able to understand and use the data [5]. This challenge particularly applies to the legal context, which contains a large number of technical and legal-specific terms, thus, making it difficult for the general public to understand legal decisions. One way to deal with this challenge is to generate simpler translates or summarizations of these legal decisions. As the manual generation is both costly and time consuming, the use of AI models, methods, and techniques of Natural Language Processing (NLP) [6], such as automatic text simplification [7], text summarization [8], and named entity recognition [9] could possibly be promising solutions. In addition, providing this information under Visual Law [10] concepts can also increase the benefits.

Within this context, this paper describes a design science research aimed at increasing the transparency of the Pernambuco's Court of Accounts (TCE/PE), by enhancing the understanding of its legal decisions through NLP and Visual Law. The institution is a control entity responsible for public government accountability, and audits and judges public accounts of all the municipalities in a Brazilian state. Besides describing this contribution, we also discuss the challenges and difficulties involved in developing AI systems in the legal domain of the Portuguese language.

The remainder of this paper is organized as follows. Section 2 presents the research background regarding NLP and related work in Brazilian Portuguese. In Section 3, we present the methodology of the study. Section 4 presents additional details regarding NLP implementation and discusses the main results, challenges, and difficulties. Lastly, Section 5 presents the concluding remarks.

2. Research Background and Related Works

2.1. Research Background

NLP is an interdisciplinary field that employs computational techniques to learn, understand, and produce human language content [6]. It includes diverse approaches, from creating spoken dialogue systems and speech-to-speech translation engines, to identifying sentiment and emotion toward products and services. The present research focuses on three areas of NLP: text simplification, text summarization, and named entity recognition.

Text simplification (TS) was studied by linguistics long before using AI [11]. It reduces the complexity of a text to improve its readability and understandability, while retaining its original informational content [7]. It also modifies syntax and lexicon to improve the understandability of language for users. Over time, TS has become an essential tool in helping those with low literacy levels, reading comprehension problems, and non-native learners. The automation of this process is a complex problem, and current related research is still far from reaching a satisfactory solution [7]. TS commonly focuses on lexical simplification, syntactic simplification, or machine translation. Lexical simplification [12] replaces complex words with simpler alternatives with equivalent meanings in a given sentence. Syntactic simplification [7] reduces grammatical complexity by replacing complicated syntactic structures with simpler ones. Additionally, Machine translation (MT) addresses the TS task as a mono-lingual translation problem, where complex sentences are made simpler. It adopts two main approaches: statistical machine translation (SMT), based on statistical and probability models, and neural machine

translation (NMT), using deep learning techniques that have achieved very satisfactory results [13]. Another NLP technique is **automatic text summarization** (ATS), which aims to produce a summary that includes the main ideas in the input document using less space and keeping repetition to a minimum [8]. Thus, it enables users to obtain the main points of the original document without the need to read it in its entirety. There are three main approaches for ATS: extractive, abstractive, or hybrid. The extractive selects the most important sentences in the input document and then concatenates them to form a summary. The abstractive represents the input document in an intermediate representation, which then generates the summary with sentences different from the original sentences. The hybrid approach combines these both. **The named entity recognition and classification** research focuses on finding the members of various predetermined classes, such as person, organization, location, date/time, quantities, numbers etc. [9]. There are usually three approaches: rule-based approaches, using syntactic-lexical patterns; machine learning approaches, automatically learning complex patterns; and hybrid approaches, combining the previous approaches.

2.2. Related Work

Most studies regarding NLP are focused on the English language. However, some studies may be found that adapt the strategies for texts in Portuguese. Aluísio et al. presented the PorSimples project [14], which aimed at simplifying Portuguese text for digital inclusion and accessibility. They developed several technologies, such as an authoring system that helps authors to produce simplified texts targeting people with low literacy levels and a web content adaptation tool for assisting low-literacy readers to perform detailed reading [14]. One of the project's main challenges is text summarization, since the simplification increases text length while enhancing text comprehensibility. In 2010, [15] addressed the problem of simplifying Portuguese texts at a sentence level by treating it as a "translation task", using SMT. Given a parallel corpus of original and simplified texts, a standard SMT system was trained and evaluated. In [16], Estivalet and Meunier presented the Brazilian Portuguese Lexicon (LexPorBr), a word-based corpus for psycholinguistic and computational linguistic research. The final corpus has more than 30 million word tokens, 215 thousand word types, and 25 categories of information on each word. A more recent example was presented by [17]. The work described an empirical study on the use of state-of-the-art ATS methods to simplify texts in Portuguese, by using different NMT techniques for ATS over two parallel corpus extracted from complex and simplified translations of the Bible, and achieved promising results.

In summary, despite some advances, performing NLP tasks in Brazilian Portuguese still remains a challenging task. Most studies are strongly based on building a parallel corpus for training and comparison, which is a costly task. There is also little evidence and assessment of these initiatives with regard to practical usage by the population.

3. Research Method

This research aimed at answering the following research question: *what are the main challenges for developing an AI system focused on enhancing the understanding of legal decisions*? To address this question, we employed a Design Science Research (DSR) [18], which involved the gradual

implementation of a minimum viable product (MVP) in different research cycles. This research method seeks to design and investigate artifacts in context (TCE/PE) by iterating over the activities of designing and investigating. Hence, we performed three design cycles formed by three activities: problem investigation (what phenomena must be improved?), treatment design (how to design an artifact that could treat the problem?), and treatment validation (would these designs treat the problem?), as proposed by Wieringa [18].

Problem Investigation: During this phase, in each design cycle, we aimed at defining the stakeholders of our project – from TCE/PE (president and two auditing directors) and from society in general (e.g. representatives of citizens, who were studied as personas). Then, we understood their levels of awareness of the problem and the treatments. For instance, in the first cycle, we noticed the need to deepen the understanding about the issue raised by the top management of Institution: the difficulty for citizens to comprehend the results of the legal decisions presented in the processes available for external access. The second design cycle aimed at providing stakeholders with rich examples of Legal Design and Visual Law (part of our conceptual problem framework) so that we could clarify these concepts and verify whether the goals for the project should be refined. Therefore, this phase allowed us to regularly evaluate the effects of the solution being created in terms of contribution to stakeholders' goals.

Treatment Design: We initiated this phase with a proper understanding of the phenomenon (automatic simplification of legal decisions) as we could discuss its causes (e.g., jargon, terms, and difficult expressions adopted in the texts) and effects (e.g., reduced readability, lack of appeal of the documents for the general public). Such overview allowed us to specify the requirements for the solution through the different iterations. We briefly describe the main final requirements (focused on features – functional/FR – and constraints of the solution – non-functional/NFR) defined with the stakeholders in Table 1. After specifying the requirements, which reflected stakeholders' goals, we created varied versions of our treatments (i.e. the MVP, our potential solution). We not only developed but documented each designed artifact (e.g. code, software engineering outputs) as they represented our decisions as a group of researchers and practitioners. In each cycle, an improved version of the MVP was presented biweekly to the stakeholders so that they could verify the prototypes developed, check to what extent it addressed the project goals, and suggest improvement. Such guidance and shared decision-making enabled us to take plan the next iteration.

Treatment Validation: In this phase, we could validate a treatment (i.e. our MVPs/proto-types gradually evolved) to ensure that it contributed to stakeholder goals. For an objective assessment of the implementation, the project considered three approaches: initial validation by the stakeholders' perception, legibility metrics, and manual inspection by non-specialists. As legibility metrics, it was used the length of the summaries and the Flesch-Kincaid adapted for the Portuguese language [19]. For the manual inspection, a structured analysis was performed on 5% of the summaries (53 out of 1003), randomly selected. Three non-specialists inspected the original documents and the summaries, and some questions were answered: (a) Completeness: Is the summary complete, representing all the important points of the original? (b) Easiness: Is the summary easy to understand? (c) Does the system highlight the important terms? (d) Not important terms were highlighted? (e) Do difficult words present a dictionary? and (f) Not difficult words present dictionary? For questions (a) and (b), possible answers were presented on a Likert scale, from totally disagree to totally agree. For the others, possible answers were

Yes or No, and additional comments and suggestions for including/excluding words could be provided.

Table 1
Requirements description

Req. ID and title	Description
FR1 – search and filter options	The solution must have search and filter options, enabling the user to find the desired decisions easily in terms of municipality and mayor, time, and decision's status (approved or rejected).
FR2 - Automatic text simplification	The solution must read the original text and automatically translate it to a version that the general population may easily understand.
FR3 - Automatic text summarization	The solution must summarize the decisions by selecting the most important sentences from the text using an automatic algorithm.
FR4 - Named Entity Recognition - Highlight	The solution must recognize and highlight members of predetermined classes, such as financial values, dates, references to laws, percentages and other entities defined jointly with the stakeholders.
FR5 - Named Entity Recognition - Dictionary	The solution must recognize and highlight legal terms and entities contained in legal texts, presenting a dictionary with their meaning.
FR6 - Graphical Visualization of Spending Limits	The solution must provide graphical visualizations of compliance with the rules of spending limits.
NFR1 - User Experience based on Visual Law	The solution must consider Visual Law concepts, including visual elements that facilitate the visualization and improve user experience.

4. Results

The system is based on a client-server architecture. Its main components are presented in Figure 1.

The **data layer** is the origin of the data, collected from the open data repository of the Brazilian court, through its REST API [20]. The server layer contains the main components of the system. The extractor is responsible for collecting data from the open data repository and persist all data on the consolidated database. As a direct mapping from the requirements, the Dictionary component implements FR5; the Highlight component enhances decisions according to FR4; spending limits processes and enhances data with information regarding spending limits, according to FR6; and summarization processes the data and generates the summarization according to FR3. Lastly, the publisher is responsible for generating the interface with the client layer, making all data available, and allowing the options of search and filtering defined in FR1.

Details and discussion on the implementation of all these components are presented in Section 4. The final architecture lacks a component for the implementation of FR2, the automatic text simplification, because it was not approved in the validation. Thus, despite the non existence of the automatic simplification, the other features, mainly the summarization, performed well and were able to deliver the main objectives. Details regarding this implementation and results are presented in Section 4. Lastly, the **Web interface** component is responsible for interacting with the Publisher component and dynamically generating the web pages related to the project.
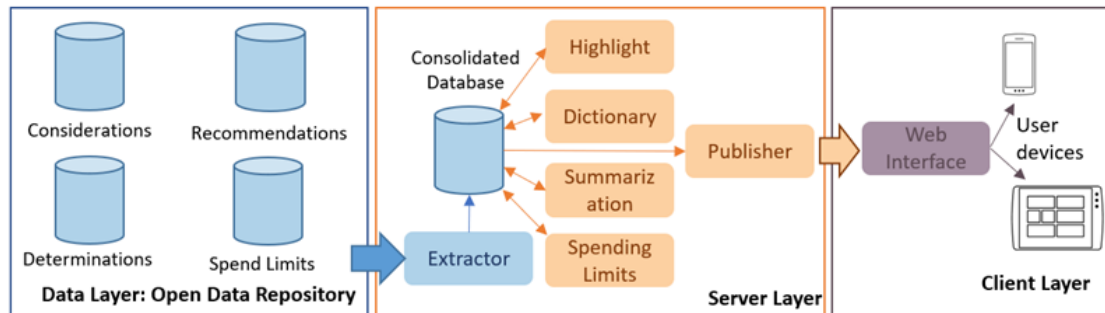
Figure 1: The Architectural Design of the System

The **implementation** of the system was made using python and java technologies and is publicly available at *decisoestce.innovagovlab.org*. As this is a live project, the page may be different at the time the reader accesses this paper. The most important page is the process details. In addition to a header containing more information on the decision, including details of the process and a link to the original document, the system also presents a short and an expanded summary with visual information about main indicators. The expanded summary is presented in Figure 2a. The figure presents the three-paragraph summary (a), including highlights, links to the dictionary, and the main decision's aspects. For comparison Figure 2b presents part of the original three-page PDF file.

Figure 2: (a) the summarized text and indicators. At the right (b), part of the original three-page PDF

5. Discussion

This section discusses the main challenges and lessons learned from implementing the NLP features: automatic text simplification, automatic text summarization, and named entity recognition. In addition, we present a user evaluation study based on an online questionnaire, which enhanced our post-mortem analysis of the entire project.

The **automatic text simplification** was initially the primary objective of the project. Two approaches were employed: lexical simplification and neural simplification. For the **lexical simplification**, three different corpuses were used: a corpus from the Bible in Portuguese, containing 60,357 sentences; a corpus published by [14] (NILC), containing 1,521 sentences; and the project corpus (PC), containing 2,143 sentences. The approach consisted of four phases: (i) pre-processing; (ii) identifying difficult words; (iii) replacement with a better synonym; and (iv) final adjustments. The final adjustments used Cogroo [21]. The evaluation was performed according to the treatment validation. Despite presented promising results for the Bible corpus [17], results were not approved by the stakeholders' perception, because in some cases the simplification changed the meaning of the sentence. For neural simplification, we used two Neural Networks models: Recurrent Neural Networks (RNN) [22] and RNN with Attention (Attention) [23]. The approach combines the use of these neural networks with the addition of pre-trained embeddings and pre-trained bidirectional encoder representations from transformers (BERT). As in the lexical simplification, the model obtained good results with the Bible and NILC corpus, but poorer results with the PC corpus. In particular, the manual inspections detected that results presented longer texts with excessive, repeated words. As both approaches were not approved by the stakeholders' perception, this requirement was suspended and the project focused on the other requirements, particularly the summarization.

The **automatic text summarization** became the main feature of the system. Thus, two summaries were produced. A very short summary is generated automatically from the data gathered from the spend limits API, and no NLP methods were applied. It is on the form (translated from Portuguese): "*There were [not] fount irregularities regarding <items> on the accounts of <municipality> under the management of <manager> for the year of <year>, and the accounts were [Approved|Rejected|Approved with reservations]*." The full summary is produced by applying summarization methods over the main information presented in the text. It was generated by an extractive approach, using the traditional steps: pre-processing of the original sentences; processing, where a representation of the text is created and high-scoring sentences are extracted; and post-processing. For implementation, the pysummarization [1] library was used, which is based on an Encoder/Decoder centered on LSTM, thereby improving the accuracy of summarization by sequence-to-sequence learning. After approval by the stakeholders', legibility metrics, and manual inspection by non-specialists were performed. The evaluation by non-specialists for the summaries presented the following results: (i) the average of answers for the completeness was 3.9, and the median was 4; and (ii) the average of answers for the easiness was 4.2, and the median was 4. This data indicates that, for them, despite not being perfect, the summaries present most of the important points of the text and are also easy to understand. The legibility metrics confirm this result: the full text presented an average readability score of

---

[1]https://pypi.org/project/pysummarization/

51 (median = 103) and 505 words (median = 456), whilst the summaries presented a readability score of 177 (median = 186) and 96 words (median = 91). Considering that higher scores signify easily of understanding, we can conclude that the summaries that presented shorter texts were easier to understand.

For the **named entity recognition**, the first strategy was to train a model using a legal NER dataset [24]. However, the obtained results were not approved by stakeholders, so we developed a rule-based approach using syntactic-lexical patterns. The solution focuses on the use of regular expressions to highlight dates, percentages, laws and similar, and monetary values. An additional component was developed for creating a dictionary of legal terms, highlighting and showing their meaning. The creation of the dictionary was based on a term frequency approach, comparing the frequency of each word in the dataset with the most present words on the LexPorBR Brazilian Portuguese corpus [16]. The words with a high frequency in the documents and which were either not present or had a very low frequency in the LexPorBR corpus were indicated as candidates for domain words. Lastly, these words and their meanings were presented to the domain specialists at the court for validation and inclusion on the dictionary. This process generated a dictionary containing 814 words. After the manual inspection by non-specialists, only 14 new words were suggested for inclusion, indicating the good results of the process.

A **preliminary system evaluation** was also performed considering the user viewpoint. A questionnaire was presented to a small group of users who could evaluate the legal summary, the dictionary, the graphics, and general design. For each feature, possible answers were presented in a Likert scale, from not useful at all to very useful. In total, 11 participants, from different backgrounds, answered the questions. Results are presented in Table 2. This is a small experiment, and a further better assessment of user perception is needed. However, this preliminary result demonstrates a high-value perception of the features by the potential users.

Table 2
User evaluation results

Evaluation	Dictionary	Summary	Graphics	General Design
Very useful	2	2	2	1
Useful	7	7	5	6
Neutral	2	2	3	3
Less useful	0	0	1	1
Not useful at all	0	0	0	0

By summarizing the results and lessons learned, the following points may be highlighted: Considering text simplification evaluation is a complex and costly process. Neither lexical nor machine translation simplification presented acceptable results according to a manual evaluation, and further research and/or different implementations are needed in this regard. Considering text summarization, the results of the employed strategies were promising both considering the metrics evaluation and considering the manual inspection performed by non-specialists. Regarding named entity recognition, despite the existence of a specific dataset for named entity recognition in Brazilian legal texts, the results in our study were not approved by the stakeholders. Thus, a rule-based approach using syntactic-lexical patterns was employed.

6. Concluding Remarks

This paper presented a study on the application of NLP and Visual Law to increase the transparency of a Brazilian court of accounts (TCE/PE). A description of the system architecture, as well as the software engineering phases, from the requirements to implementation, were given. Its main functional requirements include search and filter options; automatic text simplification; automatic text summarization; named entity recognition, highlight and dictionary; and graphical visualization of spend limits. Its main non-functional requirement is regarding the user experience based on Visual Law concepts. Such software engineering perspective in NLP solutions is rare in the literature, and this paper may lead to new practical implementations and reports in the area. Additionally, this work highlights that text simplification in Portuguese is still a challenge, as both lexical and machine translation simplification did not achieve the expected results. However, the text summarization results achieved a good performance and was approved both by the stakeholders and non-specialists. Moreover, dictionary based simplification, based on the difference of the word frequency of analyzed documents and of a traditional Portuguese corpus also proved to be promising. The assessment of the results was done through manual validation, by stakeholders and young adults. In this regard, as the solution must suit the needs of citizens as a whole, a broader assessment including diverse niches of society will be performed in a future study. Indeed, this has already gone into a planning stage in order to prepare the system to move from the MVP status and for being incorporated into the public solutions of the court.

References

[1] W. Parks, Open government principle: Applying the right to know under the constitution, Georg. Wawhingt. Law Rev 26 (1957).

[2] J. Bertot, P. Jaeger, J. Grimes, Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies, Gov. Inf. Q 27 (2010) 264–271. URL: http://dx.doi.org/10.1016/j.giq.2010.03.001. doi:10.1016/j.giq.2010.03.001.

[3] M. Ahn, Y.-C. Chen, Artificial intelligence in government: Potentials, challenges, and the future, in: The 21st Annual International Conference on Digital Government Research, 2020, p. 243–252. doi:10.1145/3396956.3398260.

[4] D. Valle-Cruz, E. Ruvalcaba-Gomez, R. Sandoval-Almazan, J. Criado, A review of artificial intelligence in government and its potential from a public policy perspective, in: Proceedings of the 20th Annual International Conference on Digital Government Research, 2019, p. 91–99. doi:10.1145/3325112.3325242.

[5] K. S. Brito, M. S. Costa, V. Garcia, S. L. Meira, Brazilian government open data: implementation, challenges, and potential opportunities, in: Proceedings of the 15th Annual International Conference on Digital Government Research, 2014, p. 11–16. doi:10.1145/2612733.2612770.

[6] J. Hirschberg, C. Manning, Advances in natural language processing, Science 349 (2015) 261–266. doi:10.1126/science.aaa8685.

[7] S. Al-Thanyyan, A. Azmi, Automated text simplification: A survey, ACM Comput. Surv 54 (2022) 1–36. doi:10.1145/3442695.

[8] W. El-Kassas, C. Salama, A. Rafea, H. Mohamed, Automatic text summarization: A comprehensive survey, Expert Syst. Appl 165 (2021) 113679.

[9] A. Goyal, V. Gupta, M. Kumar, Recent named entity recognition and classification techniques: A systematic review, Comput. Sci. Rev 29 (2018) 21–43.

[10] C. Brunschwig, On visual law: Visual legal communication practices and their scholarly exploration, in: E. Schweihofer (Ed.), Zeichen und Zauber des Rechts: Festschrift für Friedrich Lachmayer, Editions Weblaw, Bern, 2014, p. 899–933.

[11] S. Blum, E. Levenston, Universals of lexical simplification, Lang. Learn 28 (1978-12) 399–415. doi:10.1111/j.1467-1770.1978.tb00143.x.

[12] G. Paetzold, L. Specia, A survey on lexical simplification, J. Artif. Intell. Res 60 (2017-11) 549–593. doi:10.1613/jair.5526.

[13] F. Stahlberg, Neural machine translation: A review, J. Artif. Intell. Res 69 (2020) 343–418. doi:10.1613/jair.1.12007.

[14] S. Aluísio, C. Gasperin, Fostering digital inclusion and accessibility: the porsimples project for simplification of portuguese texts, in: Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas, 2010, p. 46–53.

[15] L. Specia, Translating from complex to simplified sentences, in: International Conference on Computational Processing of the Portuguese Language, 2010, p. 30–39.

[16] G. Estivalet, F. Meunier, The brazilian portuguese lexicon: An instrument for psycholinguistic research, PLoS One 10 (2015). doi:10.1371/journal.pone.0144016.

[17] T. Lima, A. Nascimento, G. Valença, P. Miranda, R. Mello, T. Si, Portuguese neural text simplification using machine translation, in: Intelligent Systems. BRACIS 2021, 2021, p. 542–556.

[18] R. Wieringa, Design science methodology for information systems and software engineering, Springer, 2014.

[19] T. Martins, C. Ghiraldelo, M. G. V. Nunes, O. Junior, Readability formulas applied to textbooks in brazilian portuguese, 1996.

[20] Tribunal de contas do estado de pernambuco, "open data api tce/pe, 2021. URL: https://sistemas.tce.pe.gov.br/DadosAbertos/Exemplo!listar.

[21] W. Silva, M. Finger, Improving cogroo: the brazilian portuguese grammar checker, in: Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, 2013, p. 21–29.

[22] L. Medsker, L. Jain, Recurrent neural networks: Design and applications, Book (2001). doi:10.1201/9781420049176.

[23] A. Vaswani, Attention is all you need, in: 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017, p. 11.

[24] P. Araujo, T. Campos, R. Oliveira, M. Stauffer, S. Couto, P. Bermejo, Lener-br: A dataset for named entity recognition in brazilian legal text, 2018.