Open-Source Collaboration to Assess the Text Complexity, Helping to Read and Write in Spanish-Speaking Schools

Diego Palma[1], *Christian Soto*[1]

[1] *Universidad de Concepción*

Concepción, 1290 Víctor Lamas Street, Chile

ABSTRACT

In this paper, we describe the role of open-source tools to develop educational systems. In particular, we describe TRUNAJOD, an open-source library to extract readability indices from texts, and how it was used to develop a tool to analyze text complexity of school texts to ease the workload on teachers when selecting a text for a particular reading comprehension task. This educational tool helps the user to calibrate or select texts based on the target school level and desired complexity for the task, offering additionally important development possibilities for writing assessment.

Keywords: Natural Language Processing, Readability Assessment, Text Complexity, Lexical Diversity, Open-Source system, Intelligent Tutoring System, Reading Comprehension Tool

INTRODUCTION

In this paper we describe a collaborative model based on open-source software to develop educational systems. In particular, we describe how we built an educational system to ease the workload of teachers in the school system when selecting texts for a particular text comprehension task.

To achieve this, the tool uses TRUNAJOD open-source library to extract multiple proxies for text complexity such as lexical diversity, coherence, and cohesion-based metrics. To provide feedback to the end-user, we also rely on a data-driven approach, in which we try to summarize the set of text measurements into five textual dimensions: lexical similarity, referential cohesion, concreteness, connectivity, and narrativity. This set of predictors is obtained via a factor analysis on textbooks from the target school system and is also validated using human experts. The tool also relies on statistical techniques and machine learning techniques to create a model that can classify if a given text is adequate to a particular school level, given its linguistic content. Results showed that the accuracy of the tool to classify the adequacy of a text given a target level is close to 80%. Moreover, results also showed that the latent factors correlate with the school level and that different school levels have a different set of linguistic features that impacts the adequacy of the text to a particular school level.

The contribution of this work is threefold: Firstly, we present TRUNAJOD open-source library as a utility for researchers working on reading comprehension, text complexity, quality of writing, second-language acquisition, and any other task involving analysis of language/text. Secondly, we provide an architectural design based on open-source tools to create an educational system that eases the workload on school teachers for specific reading comprehension tasks. Thirdly the tool could be used to evaluate various aspects of writing, generating appropriate evaluation parameters for students of a given school level. This would allow undertaking new developments in educational technology, such as intelligent tutor systems, and since it collaborates with other existing open-source libraries, the architectural design could be extended to any other language.

RELATED WORK

Readability formulas have been the classical approach to estimate text difficulty. The most used ones are Flesch Reading Ease Readability Formula, Flesch-Kincaid Grade Level, New Dale-Chall Readability Formula, Gunning's Fog Index, Fry Graph Readability Formula.

Traditional formulas measure comprehension complexity via factors related to word and sentence complexity. Word complexity is measured through several word properties such as imaginability, concreteness, familiarity, frequency of usage, number of syllables, among others. When measuring sentence complexity, the common approach is using averages of word complexity among sentences. When these formulas are compared with independent measures of text complexity - e.g. text complexity assessment scores, cloze test, human expert assessments - results show that derived formulas are valid and good predictors to estimate text complexity.

These formulas have been used widely, but nonetheless they have been criticized. It has been stated that these formulas only measure some factors of the textual surface, but not the actual sources of difficulty in a text. The formulas would not be sensitive to complex cognitive processes that discourse processing implies, and would also not be sensitive to text genre, text rhetorical structure or social factors, pragmatic and communicational that are latent to comprehension situations (McNamara et al., 1996). A frequent argument is that even if the formulas have predictive validity and a practical value, the formulas are atheoric, meaning that the text difficulty is not explained from a theory of reading or comprehension, and thus, the formulas have no validity of construct. The reason that the formulas are useful is that the factors that they measure are generally correlated with complex syntactic structures and that infrequent words are correlated with complex concepts (Miller And Kintsch, 1981). Nevertheless, there would be no causal relationship between text readability and the proposed formulas. The validity limitations of traditional formulas are evident when they are contrasted with cognitive and psycholinguistic models that explain the processes that readers follow when they construct meaning from a text (Crossley, Greenfield And McNamara, 2008).

As a consequence of the mentioned drawbacks, the state of the art changed directions in deriving measurements that take into account cognitive processes that are responsible for comprehension, the previous knowledge and the text structure. A plethora of empirical research that happened in the 70s show findings around these cognitive processes and their measurements (Thorndyke, 1977), (Kieras, 1978), (Haviland and Clark, 1974), (Kintsch and Keenan, 1973), (Kintsch et al., 1975).

In the last decades, research around readability has progressed at a fast pace, and this is a consequence in the development of other related fields such as computational linguistics, corpus linguistics, and the cognitive psychology related to discourse comprehension. This has made possible the development of computational tools that are able to measure readability based on indicators that consider several levels of discourse (e.g. surface proxies, situational model, genre and rhetorical structure) (Crossley et al., 2007) and that correspond with theory and contemporary models about discourse comprehension. These tools not only analyze the text on its surface, but also rely on natural language processing techniques to analyze text structure and other relevant factors that might impact text readability.

One popular tool is Coh-metrix (Graesser et al., 2004) developed at Memphis University and designed to measure text complexity at multiple levels. Initially its focus was in cohesion, but was extended to focus on linguistic cues associated with

deep levels of comprehension.

Recently, the team from Memphis have developed a new tool known as Text Easy and Readability Assessor or TERA (Jackson et al., 2016), especially designed to help teachers select texts in reading comprehension tasks.

The problem with the mentioned tools is that they are essentially closed-box systems, meaning that no modifications can be done. Moreover, most of them are not even open-source and if they are open-source they are not maintained properly in a repository that would allow people adding contributions, so, in a sense they are still closed-box systems. Another problem is that most of these tools only support the English language. TRUNAJOD (Palma et al., 2021) is an open-source tool for text complexity that follows an approach stated in the Journal of Open-Source software (Katz et al., 2018). This means that the tool is essentially a library that supports multiple readability metrics but it is also an open-box system, allowing contributions from different researchers and even non-researchers interested in fixing issues or improving usability of the tool. Moreover, TRUNAJOD was built on spaCy (Honnibal And Montani, 2017) which supports multiple state of the art natural language processing models for different NLP tasks, such as named entity recognition, dependency parsing, POS tagging among others. Furthermore, since spaCy is open-source it has also been extended to support Stanford NLP models (Quang et al., 2020), and as such, if both libraries improve, it will likely improve TRUNAJOD performance. Finally, TRUNAJOD might be used to build tools for human-computer interaction aiming at different target users, which is one of the contributions of this paper.

AN OPEN-SOURCE COLLABORATION WITH EDUCATIONAL TECHNOLOGIES

The open-source collaboration with educational technologies architecture is shown in Fig. 1. The model is inspired in the journal of open-source software approach (Katz et al., 2018), in which software is available for modifications using a licensing model. This model has the following advantages:

- Some specific purpose libraries can be reutilized, updated or modified according to special needs and as a consequence this broadens the use cases for the library.
- If an open-source library depends on another open-source library, a collaboration is enforced by the dependency. This implies that if the dependency improves (e.g. accuracy, features), some improvements might be passed downstream to the dependent library.
- In an educational system, that is an open-box system, all the enhancements that happen upstream are reflected by the system, so this improves user experience.
- As the dependencies and libraries will be open-source and maintained in a source code version control system, multiple researchers can collaborate in

enhancements, reporting issues, or creating extensions of the lib. This also helps in making research in the field faster, as experiments can be reproduced in a controlled manner.

In this work, we also present a text-complexity assessment tool that was created using this model, which has as target users Spanish-speaking schools, but as we show later, can be extended to other similar tasks in different languages.

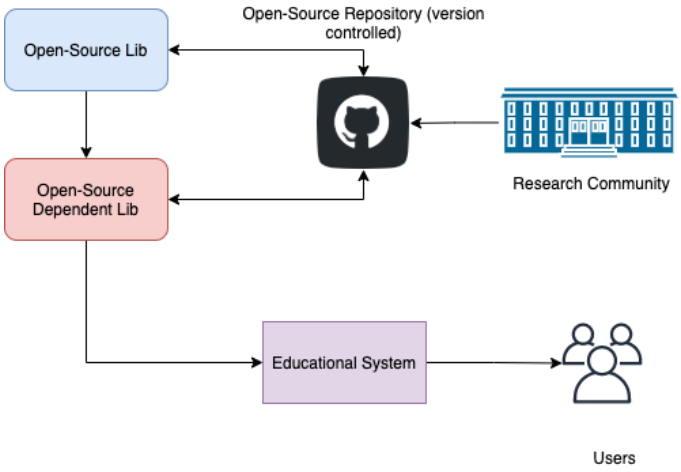

Figure 1. Open-Source collaboration architecture.

THE COLLABORATION MODEL FOR EDUCATIONAL TOOLS

TRUNAJOD (Palma et al., 2021) is an open-source library that supports several text complexity indices that can be used for different tasks (e.g. automatic summarization, automatic essay scoring, intelligent tutoring systems, etc.). We can think of an open-source collaboration model as a pipeline consisting of transfer functions at each step, as shown in Fig 2. Each step has a performance value of pi that can be thought of as a goodness of fit of transforming an input into an output (e.g. a pre-processing step). In a pipeline consisting of multiple steps, these performance penalties are multiplicative, equation (1) shows what is the total performance of the system consisting of N steps

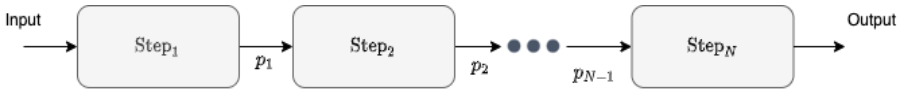

Figure 2. Pipeline of transfer functions.

$$p_{system} = \prod_{i=1}^{N} \quad p_i. \tag{1}$$

The intuition of the collaborative model is that if one of the steps improves in performance, it will imply that the system as a whole will improve in performance. TRUNAJOD follows this collaborative model implementing a software architecture as shown in Fig. 3. In this architecture we make use of already implemented models for general NLP tasks, and make use of these models in order to compute TRUNAJOD readability indices. There are several advantages of doing this: depending on an open-source system, makes easier to manage dependencies and community can help improving the dependency, thus, making TRUNAJOD open-source system eases researchers to build educational tools in language tasks, and make experiments easier to reproduce and easier to extend.

BUILDING AN EDUCATIONAL SYSTEM WITH TRUNAJOD

We built an educational system for reading comprehension tasks. The target users of the system are teachers in the Chilean school system. The tool's goal is helping teachers to select texts for a reading comprehension task, having as target a given school grade, and also check the difficulty of the text for that particular level. This is achieved by creating an adequacy model for the texts that checks if a given text belongs to a desired school level. To build this model, we rely on TRUNAJOD text complexity indices, and human experts in the field. Then, we can think of the problem as a classification task, and we used machine learning techniques to derive such a model. On the other hand, the model itself is not sufficient, as we need to provide some feedback to non-technical users (i.e. school teachers). To achieve this, we used a statistical technique known as factor analysis, and using expert knowledge in the field, we derived five latent variables of text complexity: narrativity, lexical diversity, referential cohesion, concreteness, and syntaxis.

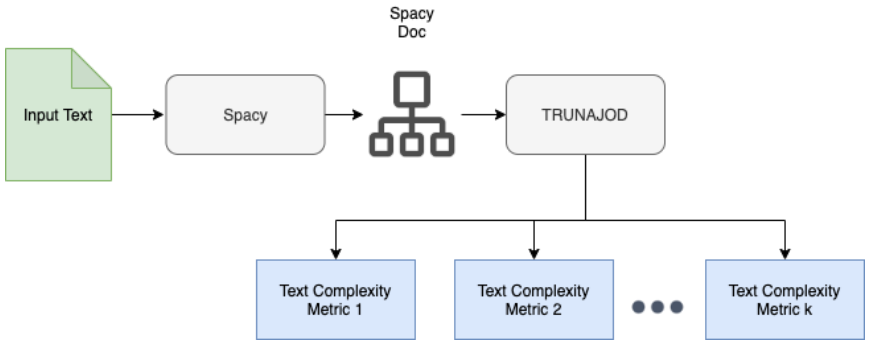

Figure 3. TRUNAJOD pipeline for text complexity.

We collected a corpus of 2133 school texts from the Chilean school system, for different grade levels, as shown in (see Table 1).

Table 1: Distribution of texts analyzed to create TRUNAJOD web app.

Level	# of Texts
1	121
2	180
3	181
4	186
5	191
6	185
7	180
8	180
9	181
10	181
11	183
12	184

We tested different configurations for the adequacy model, and the fact that there were no significant differences between adjacent school levels in terms of linguistic features, the best performance was obtained targeting at a group of school levels, and to not lose granularity, we decided to reduce the levels to six groups (i.e. level 1-2 belong to group 1, level 3-4 belong to group 2, etc.). Then we trained a Random Forest model, in a one-vs-all model, where the target was 1 if the text belonged to the target level and 0 otherwise. The performance metric was accuracy score, which is the ratio of correctly classified records vs the total number of records. We obtained the following accuracies per group: 0.9 for group 1, 0.85 for group 2, 0.8 for group 3, 0.82 for group 4, 0.75 for group 5 and 0.70 for group 6.

To generate feedback for the user, we do not use the complete set of predictors. We used a factor analysis to obtain latent variables that, using human expert knowledge, we categorized as: lexical similarity, referential cohesion, concreteness, connectivity, and narrativity. To improve user experience, the factors are re-scaled into a 0 to 1 range, and transformed to text ease measurements, meaning that, the closer to 1, the easier the text. This will help users to calibrate texts given to a target group of students for a reading comprehension task.

Finally, we theorize that at a higher level of education, readability differences are less

clear than for example lower levels, in which students are actually acquiring new words and learning how to use the language. This would also explain why we get better performances when comparing texts in lower school levels. We also theorize that the feedback provided by the latent variables, will help teachers in calibrating their texts and understanding better how their students will perform for a given task.

CONCLUSIONS

Open-Source collaboration design for educational systems act as an open-box system and enable researchers and community to collaborate directly or indirectly in a particular component of an educational system. The collaborative pipeline enables indirect improvements if different components of the system are improved. We showed how this could be achieved with the TRUNAJOD text complexity system. We also showed that TRUNAJOD could be used to build an educational system for reading comprehension tasks. Future work will involve extending this to writing tasks, automated essay scoring and even helping text editorials to improve their texts on a readability dimension.

Finally, we believe that TRUNAJOD might be a good contribution for the research community, not only in educational systems but also on reading comprehension tasks. As the TRUNAJOD repository is open-source, the tool might get contributions and even new indices for different languages.

ACKNOWLEDGMENTS

Thanks to the FONDEF ID20I10290 project: “Literator: An intelligent tutor that enhances reading and writing skills” funded by ANID, Chile.

REFERENCES

Crossley, S., Dufty, D., McCarthy, P., and McNamara, D. 2007. Toward a new readability: A mixed model approach. In Proceedings of the Annual Meeting of the Cognitive Science Society.

Crossley, S.A., Greenfield, J. and McNamara, D.S. (2008). Assessing Text Readability Using Cognitively Based Indices. TESOL Quarterly, 42(3), pp.475–493.

Haviland, S.E. and Clark, H.H. (1974). What’s new? Acquiring New information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 13(5), pp.512–521.

Honnibal, M., and Montani, I. 2017. Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Unpublished software application. https://spacy. io.

Katz, D., Niemeyer, K., and Smith, A. 2018. Publish your software: introducing the journal of open source software (JOSS). Computing in Science & Engineering, 20(3), p.84–88.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. Psychological Bulletin, 85(3), pp.532–554.

Kintsch, W. and Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 5(3), pp.257–274.

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G. and Keenan, J.M. (1975). Comprehension and recall of text as a function of content variables. Journal of Verbal Learning and Verbal Behavior, 14(2), pp.196–214.

McNamara, D.S., Kintsch, E., Songer, N.B. and Kintsch, W. (1996). Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. Cognition and Instruction, 14(1), pp.1–43.

Miller, J.R. and Kintsch, W. (1981). Knowledge-based aspects of prose comprehension and readability. Text - Interdisciplinary Journal for the Study of Discourse, 1(3).

Palma, D., Soto, C., Veliz, M., Karelovic, B., and Riffo, B. 2021. TRUNAJOD: A text complexity library to enhance natural language processing. Journal of Open Source Software, 6(60), p.3153.

Tanner Jackson, G, Allen, LK & McNamara, D 2016, Common core tera: Text ease and readability assessor. in Adaptive Educational Technologies for Literacy Instruction. Taylor and Francis, pp. 49-68. https://doi.org/10.4324/9781315647500

Thorndyke, P.W. (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9(1), pp.77–110.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. 2020. Stanza: A Python natural language processing toolkit for many human languages. arXiv preprint arXiv:2003.07082.