

WriteWise: software that guides scientific writing

Eduardo N. Fuentes^{1,3*}, Hector Allende-Cid², Sebastián Rodríguez¹, Rene Venegas², Juan Pavez¹, Wenceslao Palma², Ismael Figueroa, Sofia Zamora², Brayn Diaz¹, Ashley VanCott³,

¹WriteWise, WriteWise Research Group, Artificial Intelligence Unit, Santiago, Chile

²Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

³BioPub, Scientific Writing Unit, Santiago, Chile

*Corresponding author: Dr. Eduardo N. Fuentes; email: ef@writewise.cl

1. Introduction

Writing scientific articles is traditionally learnt through an inefficient, slow, and arduous process of trial and error. This despite the fact that writing a scientific article is not trivial and in general there is a lack of formal training for learning this task. Scientific articles are inherently complex to write, mainly due to the fact that they are highly technical and present a linguistic/stylistic format specific to the academic genre.

Currently, a scarce number of technological approaches have tried to solve this problem. Most recognizable softwares used by scientists to write papers are Word, GoogleDocs, Grammarly, Authorea, Overleaf, among others. However, none of these solutions satisfactorily provide specific feedback on academic genre, teaches about it, and aids in writing high linguistic levels (academic discourse).

To address this fundamental issue in the academic world we have developed a new software based on artificial intelligence that tackles all the aforementioned problematics; therefore, helping and teaching researchers how to write high-quality English manuscripts.

2. Methods

2.1. Corpus

Several corpora were used, but the most relevant for the development of the software were: 1) An unlabeled “Gold Standard” which consists of 2116 papers; 2) A labeled “Gold Standard” which consists of approximately 13,000 sentences. The “Gold Standard” consists of papers in life-sciences that contain top scientometrics, have in-housed copy editors with rigorous editorial processes, and that present the discursive structure prototypical of well-written papers.

2.2. System for tagging texts at the discursive level

A linguistic model for writing scientific texts in the life-sciences was generated. In parallel a computational system was used to subsequently tag sentences at rhetorical-discursive, structural, lexical-grammatical, stylistic, and functional elements prototypical to this genre.

2.3. Machine learning models

Several shallow and deep learning models were used for text representation (e.g. word and sentence embeddings), visualization (graphs), and extraction of the most relevant information needed for discourse segmentation, and to assess text coherence and cohesion.

2.4. Back and Front end development

Write Wise is presented as a modern web application implemented in proven technologies: Vue.js and Django for frontend and backend development, respectively. The system has a scalable architecture where the machine learning models are consumed as microservices inside the web application.

3. Results

The Write Wise platform can be subdivided into three modules that tackle the main problems while scientist write papers: 1) sentence length; 2) text order and organization; 3) discourse.

The first module (sentence length) provide information on the microstructure of a text, identifying the number of word per each sentence. The second module (text order and organization) automatically detects the similarity between consecutive sentences in the text. We identified an optimal N° words/sentences and semantic similarity threshold based on analysis of the unlabeled Gold Standard. With these modules the user can compare each sentence with the unlabeled Gold Standard. If the user exceeds the threshold, is notified; thus, this person can make changes in the text. Thanks to these two modules the system helps the user to transmit one idea or message per sentence using clear, precise, and concise sentences, which are consecutively connected. The third and most sophisticated module (discourse) provides the user with examples of different “discursive steps” (i.e. functional linguistic unit that fulfills a communicative purpose in a sentence). Thanks to this the user can write a logical and well-structured text at a high linguistic level.

Conclusions

WriteWise represents the first commercially available advanced platform that provides user’s help and feedback to improve scientific papers writing. This is thanks to the development of and advance textual data representation at different linguistic levels (e.g. words, sentences) through using cutting-edge machine-learning models and applied linguistics research.