



CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022

## Engage Students in News Writing

Luísa Júlio<sup>a,\*</sup>, Diogo Miguel Carvalho<sup>a</sup>, Alexandre Ulisses<sup>b</sup>, Inês Rito Lima<sup>b</sup>, Telmo Silva<sup>a</sup>

<sup>a</sup>*Digimedia, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.*

<sup>b</sup>*MOG Technologies, R. Eng. Frederico Ulrich 2650, 4470-605 Maia, Portugal.*

---

### Abstract

The technologies evolution impacts how information is produced and consumed by users. Nonetheless, with the spread of information content available on most online news platforms, the misinformation increases alongside the less credible content. In this scope, the present research aims to develop a technological ecosystem to promote students' writing ability. The system will help students, search for credible content to create school newspapers. Thus, in this article, the architecture of the solution for news writing tool for the Portuguese language is presented. This paper aims to introduce a constructive approach that presents the system architecture that will support the development of a news creation tool.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2022

*Keywords:* journalism; school newspapers; fake news; contextualization; school newspapers; platforms; natural language process

---

---

\* Corresponding author. Tel.: +351-234-370-200.

*E-mail address:* [luisaamj@ua.pt](mailto:luisaamj@ua.pt)

## 1. Introduction

As technology and online data transmission evolved, journalism and newspaper companies are expanding to digitized journalism, leading to new ways of producing and consuming information. The growth of online social media platforms allowed quick news publication in a frequently and disorganized manner. In this way, the spread of misinformation and the difficulty of distinguishing the credible content of Fake news can lead to society [1], especially among young people, with alarming levels of disinformation since this audience tends to believe in social media content.

The online newspaper creation tools integration in the classroom context can be one of the solutions to combat misinformation growth. The role of schools is critical to allow the students to develop critical thinking about what is real or fake content. Young people can benefit in two ways from the online newspaper creation tool: (i) how to optimize the search content (e.g., distinguishing what is credible) and (ii) researching for the information fonts in the news. Portugal appears as a context for this study, which endures a lack of construction school newspaper platforms and news production support tools.

This paper aims to describe a potential solution that allows younger generations to write Portuguese news through the construction of school newspapers, which promotes critical thinking development in this audience, specifying the news writing tool. The technology contribute allows to analyzing content's veracity and contextual level.

This article is divided in five sections: (i) introduction; (ii) objectives; (iii) state of the art; (iv) proposed solution; and (v) final synthesis.

## 2. Objectives

The research described in this paper aims to create a technological ecosystem that allows younger generations to write news in Portuguese in a contextualized environment. The proposed technical solution analyzes news information sources and credibility to present them as a base for text writing. This ecosystem identifies keywords inherent to the news writing process allowing students from distinct education levels to tackle the critical thinking gap.

A set of objectives were established based on the project goals: (i) develop a search engine, which provides a set of sites/news with contextualized relevant topics regarding the written news; (ii) create a credibility analysis tool that assesses the credibility of the website content, based on a set of Natural Language Process (NLP) and Machine Learning (ML) metrics and techniques; (iii) create a semantic analysis tool to detect/extract topics in the written news text by using NLP techniques; (iv) design and implement a platform for collaborative construction of school newspapers with content management mechanisms; (v) develop online distribution mechanisms for these newspapers with a component for collecting usage/consumption statistics.

## 3. State of the art

A set of scientific projects and technological platforms were critically analyzed based on technical and conceptual levels: search a few studies about (i) tools and (ii) strategies on fake news detection; (iii) identify news aggregators platforms (e.g., search APIs newspaper); and (iv) analyses Portuguese language news contextualization tools. The scientific databases Scopus and Web of Science were explored to collect a set of studies on “fake news detection strategies” and “contextualization tools”. Other results from Google Scholar and Google were added based on a search with terms such as: “fake news detection”, “fake news aggregation”, “online fake news tools” and “fake news platforms”.

### 3.1. Fake News Detection tools

The term “Fake News” is associated with the dissemination of untrue information intentionally or unintentionally [1], which can be easily done by anyone, namely through social networks [2]. All over the world, currently, a development in misinformation is notorious. An example of this is the international tools to prove the news credibility as FactCheck.org, PolitiFact, Snopes, Emergent and Truth of fiction. In the Portuguese context, the most popular tools used to detect fake news are: “Polígrafo”, “Fact-Check Observador” and “Prova dos Factos”. Through mentioned

international platforms survey, it was possible to identify some constraints regarding: various sources to access the information; most topic discussed is politics; and the news content is in the English language. About the fact-checking websites, the European Portuguese language is only integrated into the Portuguese websites/platforms and the differentiation factor is the credibility scale to classify the Portuguese news.

### 3.2. Fake News Detection Strategies

Notwithstanding the several tools that contribute to detecting fake news, the main problem is due to the increase in online information consumption, which remains to academic research focus.

Gaonkar et al. [3] proposed a model where it is possible to classify news using NLP and ML techniques. The algorithms analyze the news content and attribute binary variables (i.e., true or false) to obtain a final score. Ahmed, Traore and Saad [4] also developed a system that attributes binary variables to detect fake news, where the incorporation of algorithms techniques, can corroborate and validate the information sources used in school newspaper creation.

Silva, Santos, Almeida and Pardo [5] developed a dataset based on criteria as author association to the news. In this scope, the authors [5] identified the most fact-checking and recommendation algorithms that allow to automatically detect fake news in the Portuguese language, where Random Forest and Logistic Regression are the algorithms that showed the best performance.

Based on the analyzed studies, it was possible to understand that the technological incorporation into fake news strategies assists in detecting credible news content where most of these approaches develop manually datasets in the English language. The Portuguese projects studied, which incorporated the Portuguese language, allow identification of key features to sustain the technological ecosystem as text classification and APIs calls.

### 3.3. News Aggregators

The main goal of several news aggregators was to store news from various sources in APIs forms. This type of aggregator has as its main objective to replicate what is generated by traditional media (e.g., newspapers and radio). In this way, this fusion aids the user by simply displaying relevant information about a specific topic instead of searching across multiple platforms. For this purpose, potential APIs research that integrates the Portuguese language was necessary to identify which can support the web platform development and is a requirement to have in the technological ecosystem.

In this sense, some Portuguese Media makes available the information news through APIs, namely [7, 8]: “News API”, “MediaStack”, “Webhose”, “Público”, “Observador”, “Jornal de Notícias (JN)” and “TSF”. To better understand each APIs characteristics was carried out individually analyses to compare the difference between each other. This analysis concluded that “NewsAPI” is the best aggregator since that returns more information, followed by “MediaStack”, “Webhose” and “TSF”.

### 3.4. Contextualization Tools

News contextualization tools are a crucial method to assist the authors to write and substantiate the text since these tools extract the main keywords referred into the news text. The contextualization tools are based on Natural Language Processing (NLP) and allow word categorization in specific topics [9]. This extraction is based on software that can recognize the written/mentioned named entities and extract them.

In the Portuguese context, these kinds of tools were available in 2004 by Linguateca: Recognition Assessment of Mentioned Entities (HAREM) [6]. Over the past years, the extraction entities tools in the Portuguese language have appeared, such as: (i) Priberam, an API that identifies dates, locations and persons; (ii) Microsoft Azure, whose algorithm allows to identify into unstructured text some categories and customize the algorithms identification; and (iii) Linguakit, an NLP linguistic analyzer focused on the text structural components, which also provides an online keyword or multiword extraction.

## 4. Proposed Solution

The state-of-the-art identifies the principal tools and technical solutions in the market that supports the writing of journalistic texts in Portuguese of Portugal and supports the system architecture. Thus, the proposed solution consists of an interactive and collaborative platform to assist journalistic texts written in Portuguese through a contextualized search and a spelling error detection system.

In the next section, the architecture of the system (i) is explained. Following the news writing component (ii) is detailed.

### 4.1. System architecture

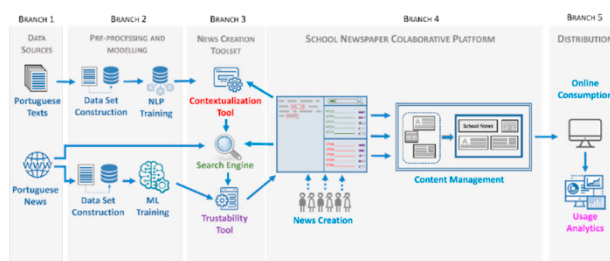


Fig. 1. High level architecture of the project.

Fig. 1 illustrates the project's architecture, which five branches reflecting key objectives, components and respective interactions: (i) branch 1 - aims to describe work related to the data information compilation from Portuguese news websites, so that it can be analyzed in (ii) branch 2, which is responsible for pre-processing and data modeling operations through NLP and ML techniques; (iii) branch 3 - tools to support the writing of news in a school context by providing a contextualization and trustability tools (e.g., search engine, related news and metrics associated and published information content's veracity); (iv) branch 4 - platform for the effective creation of news and its aggregation to build school newspapers; (v) branch 5 - dissemination and consumption of newspapers (including consumption analytics).

All these branches, such as a search engine for news search, which includes continuous feedback about the searched content credibility and provides related content based on semantic analysis of the student's writing, integrated into a tool to develop in this research benefits students.

The writing news process included in the school newspaper platform must have content management features allowing the students to customize their newspaper elements: format; frequency, newspaper layout; and students access.

After the school newspapers are created, (v) branch 5 – present on architecture system - represents the online newspaper distribution and the integration of a mechanism that provides newspaper usage statistics.

### 4.2. News writing tool

The news writing tool was developed for students to be able to take advantage of the tools developed in branch 3, resorting to the search engine for news search, having continuous feedback on the credibility of searched content/source (trustability tool) and being provided related content based on semantic analysis of their writing (contextualization tool).

For the elaboration of this tool's interface, a screen was designed that summarizes its main functionalities, namely: (i) incorporation of writing fields for students to write the title and body of the news; (ii) integration of a spelling error detector; (iii) suggested synonyms; (iv) presentation of the list of entities associated with the news produced; (v) representation and research of related news from a Portuguese newspaper; (vi) and also the incorporation of a section for researching news from various news sources, in which a credibility index on the content sought is presented.

According to a set of guidelines and heuristics on User Interface (UI) design [7], it was possible to build a graphic design, visually, consistent and centered on user needs, to optimize the design of the user experience on the platform. In this sense, a palette of colors in shades of blue was selected alluding to fidelity and truth, factors that fit the objectives of the platform.

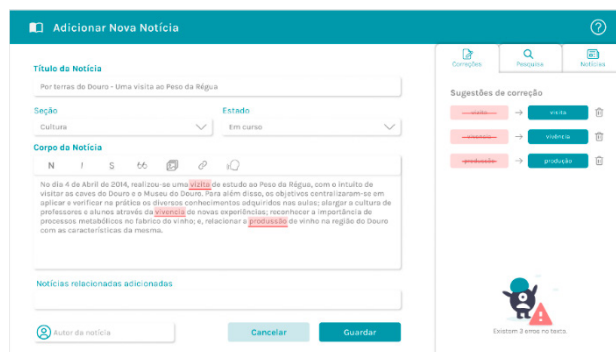


Fig. 2. News writing user interface with detection of spelling errors.

#### 4.2.1. News writing interface' structure

To assist the user in writing Portuguese journalistic texts, the news writing interface, consists: on the left side, of text fields, such as the **title** and **body** of the news, and the **author's name**. There are also **section** and **news status** fields that allow the user to identify the **section** for which he will write their news and select the **status** of their news (in progress, reviewed, published). On the right side of the interface, there are three sections: **corrections**, **search**, and **news**. In all these sections, the platform mascot will be present, giving suggestions to the user (e.g.: an indication of the number of errors present in the text).

#### 4.2.2. Detection of spelling errors

Regarding the **spelling error detection system**, the design conceived for the platform integrates a structure oriented to the presentation and identification of spelling errors. In this interface (Fig. 2), errors are identified in the text with a red underline, and the error and the correction suggestion are displayed in the corrections tab. In this way, the user can click on the correct word to correct it. Near the correct word, there is also an icon to delete the identified error. In addition, the mascot provides feedback to the user by indicating the number of errors that the user has in the text.

#### 4.2.3. Suggested synonyms

Another feature of the news writing interface is the **synonym suggestion**. When selecting a word, options are presented to the user to change the selected word, which allows him to improve his text. Thus, when selecting a word, it is highlighted with a blue underline and a balloon appears with the suggested words. Afterward, the user can choose any of the suggestions, and the word will be automatically replaced in the text.

#### 4.2.4. Identification of key concepts of written news

The news writing support component includes the functionality to identify key concepts of written news. In this regard, a request is made to an entity detection algorithm, developed with the spaCy library [8], which returns a set of entities related to the content text (i.e. written in the news creation platform). This algorithm incorporates a Portuguese dictionary based on the Bosque Constraint Grammar converted version – part of Floresta Sintá(c)tica treebank [9]. Besides, with this algorithm was possible to develop an API that grants access to introduce new terms to the algorithm's dictionary (e.g., SARS and Covid-19 are two recent terms in Portuguese vocabulary).

#### 4.2.5. Search for related news from a Portuguese newspaper

The identification of the key concepts of written news, through the previously identified tool, allows the presentation of news from a Portuguese newspaper related to recognized entities. Thus, on the right side of the interface, in the News tab, the key concepts identified are present. The user can choose to deactivate some concepts, starting to be presented with a gray background. Below are the results of the search for news from the newspaper related to the identified concepts. Any of the results can be associated with the news being written, through the plus icon, thus becoming identified in the field of related news added and changing its icon to a check.

#### 4.2.6. Search for news from different sources

Finally, it is considered important to allow students to **search for news from different newspapers**, related to a topic of interest. That said, in the interface, there is a search tab, in which the user can search for news. For each result is presented: the source, the title, the date of the news and a bar that indicates the credibility index of each news.

## 5. Final Synthesis

The evolution of newspaper writing to a more “digitized journalism”, lead to a new way of producing and consuming information. This easy access to information enhances misinformation and the growth of Fake News amongst credible content.

In this context, this research aims to stimulate students’ critical thinking and help younger generations to write news, in Portuguese, thus contributing to the revival of school newspapers. Therefore, this technology will help write news, providing tools that allow students to assess the veracity of the published news and identify its context.

With the state-of-the-art review was possible to identify which tools already exist either to detect fake news, aggregate news, or contextualize news. In addition, it also allowed to start the development of the system architecture of the final solution, and, specifically, of the news writing tool for the Portuguese language. In this way, the solution proposal meets the defined objectives, guaranteeing flexibility to add new functionalities.

In future work, the usability and user experience of the solution will be evaluated with students belonging to the intended target audience. Furthermore, it is intended to arrive at a solution that acts in education to sensitize young people to be more critical of the information they see and to contribute to the writing of news in Portuguese.

## Acknowledgments

*TRUE* project developed in partnership with the newspaper Público and the company MOG Technologies, funded by the European Regional Development Fund COMPETE 2020 (POCI-01-0247-FEDER-046923).

## References

- [1] Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* 19, 22–36 (2017).
- [2] Allcott, H., Gentzkow, M.: Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* 31, 211–236 (2017). <https://doi.org/10.1257/jep.31.2.211>.
- [3] Shabani, S., Sokhn, M.: Hybrid machine-crowd approach for fake news detection. In: *Proceedings - 4th IEEE International Conference on Collaboration and Internet Computing, CIC 2018*. pp. 299–306. Institute of Electrical and Electronics Engineers Inc. (2018). <https://doi.org/10.1109/CIC.2018.00048>.
- [4] Gaonkar, S., Itagi, S., Chalippatt, R., Gaonkar, A., Aswale, S., Shetgaonkar, P.: Detection Of Online Fake News : A Survey. In: *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. pp. 1–6 (2019). <https://doi.org/10.1109/ViTECoN.2019.8899556>.
- [5] Ahmed, H., Traore, I., Saad, S.: Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore, I., Woungang, I., and Awad, A. (eds.) *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*. pp. 127–138. Springer International Publishing, Cham (2017).
- [6] Silva, R.M., Santos, R.L.S., Almeida, T.A., Pardo, T.A.S.: Towards automatically filtering fake news in Portuguese. *Expert Syst. Appl.* 146, 113–199 (2020). <https://doi.org/https://doi.org/10.1016/j.eswa.2020.113199>.
- [7] Silva, D.: OpenNews is a REST API made in Python to extract news from Portuguese journals., <https://github.com/diogosilva30/OpenNews>, last accessed 2022/08/01.

- [8] gpirescampos: Agregador de APIs e dados públicos, ou semi-públicos de entidades Portuguesas., <https://github.com/devpt-org/public-data-portugal>, last accessed 2022/08/01.
- [9] Cambria, E., White, B.: Jumping NLP curves: A review of natural language processing research. *IEEE Comput. Intell. Mag.* 9, 48–57 (2014).
- [10] Santos, D., Cardoso, N.: Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área. *Linguatca*, Lisboa, Portugal (2007).
- [11] Nielsen, J.: 10 Usability Heuristics for User Interface Design.
- [12] Explosion: spaCy · Industrial-strength Natural Language Processing in Python.
- [13] Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., de Paiva, V.: Universal Dependencies for Portuguese. In: *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling)*. pp. 197–206. , Pisa, Italy (2017).