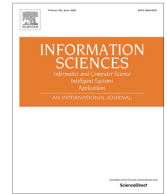




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Coronavirus fake news detection via MedOSINT check in health care official bulletins with CBR explanation: The way to find the real information source through OSINT, the verifier tool for official journals



Sergio Mauricio Martínez Monterrubio, Amaya Noain-Sánchez, Elena Verdú Pérez, Rubén González Crespo*

Universidad Internacional de La Rioja - UNIR c/ Avenida de la Paz, 137, 26006 Logroño, Spain

ARTICLE INFO

Article history:

Received 27 October 2020
Received in revised form 3 May 2021
Accepted 28 May 2021
Available online 05 June 2021

Keywords:

Coronavirus
COVID-19
Fake news detection
OSINT
CBR

ABSTRACT

This research aims to design and prototype a tool to perform intelligence on open sources (OSINT), specifically on official medical bulletins for the detection of false news. MedOSINT is a modular tool that can be adapted to process information from different medical official bulletins. From the processed information, intelligence is generated for decision making, validating the veracity of the COVID-19 news. The tool is compared with other options and it is verified that MedOSINT outperforms the current options when analyzing official bulletins. Moreover, it is complemented with an expert explanation provided by a Case-Based Reasoning (CBR) system. This is proved to be an ideal complement because it can find explanatory cases for an explanation-by-example justification.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

In December 2019, Wuhan, one of the largest cities in Hubei, a province of China, became the center of an outbreak of pneumonia of unknown cause. This pneumonia attracted a lot of attention not only within China but also internationally. Most of the initially identified patients were vendors at a poultry, seafood, and live wildlife market in China. Soon after, confirmed cases were not only limited to Wuhan but also spread throughout the world. The World Health Organization (WHO) declared this disease a pandemic, which was named 'Coronavirus Disease 2019' or COVID-19. Until April 13, 2020, a total of 1,773,084 confirmed COVID-19 cases and 111,652 deaths were documented globally across 5 continents [44].

As the COVID-19 pandemic spreads around the world, the proliferation of online disinformation weakens the world's response to the virus. From miraculous cures to conspiracy theories, social media, news apps and mobile notifications spread messages that contain useless, incorrect, or even harmful information and advice. The spread of false information related to the epidemic has become such a problem that the WHO has labelled it as an 'infodemic': an excessive amount of information about the health emergency spread together with rumors and disinformation, causing confusion and distrust to people [46]. At the same time, conspiracy theories related to the origins of the coronavirus have been created. These theories have been

* Corresponding author.

E-mail addresses: sergiomauricio.martinez@unir.net (S.M. Martínez Monterrubio), amaya.noain@unir.net (A. Noain-Sánchez), elena.verdu@unir.net (E. Verdú Pérez), ruben.gonzalez@unir.net (R. González Crespo).

shared by both, websites known to benefit from disinformation [38], and legitimate news sites falling into the trap of broadcasting incorrect information.

Closely related to the diffusion of falsehood on COVID-19, cases of xenophobia and racial prejudice against Asian people have been publicly reported. Particularly, the prejudices against people from China [38] led to the Twitter campaign #JeNeSuisPasUnVirus in France, a hashtag created to support China's fight against intolerance, which rapidly became a trending topic in various languages.

In this context, governments, agencies and multiple organizations have taken steps to tackle disinformation during the epidemic: The WHO in conjunction with search and media companies like Facebook, Google, Pinterest, Tencent, Twitter, TikTok, YouTube and others, has launched the website 'MythBusters' [47]. Besides, the European Observatory Against Disinformation (SOMA) [39] has decided to collect and debunk the most common false news to provide support to the European community. From the news industry, leading fact-checkers and verification experts have created tools, and implemented web resource hubs offering advice, and free online training to help journalists report on COVID-19 [15]. However, many users are still exposed to unfounded medical advice, rumors and other false information that could risk public health. Within this framework, the detection of fake news, defined as "the prediction of the chances that a particular news article is an intentional deception" [13], has become a central discussion topic. This research develops an OSINT based tool named MedOSINT for the verification of the content of official bulletins such as those of the WHO. MedOSINT contributes to know the veracity of the information that is generated daily on the Internet. Thus, by consulting MedOSINT official websites you will be able to verify whether the information is false or true quickly and reliably. Moreover, by integrating MedOSINT with a Case-Based Reasoning (CBR) system, an explanation of the result will be provided.

This paper is organized as follows: (1) Introduction. The present introduction; (2) State of art. By reviewing existing literature, we define the concepts of both "fake news" and "disinformation" and provide the context to address their dissemination on social media. We also present state-of-art tools for fake news detection and conclude with the advances that MedOSINT provides to overcome their limitations; (3) MedOSINT tool. The proposed tool is described, firstly its architecture is introduced, then its design and working methods are explained from the technical point of view; (4) Results analysis and evaluation. This section shows the operation of the MedOSINT tool from the user's point of view. Besides, it includes a comparison with related tools, and the evaluation showing that the initial requirements are met; (5) Conclusions. Finally, we draw the conclusions and propose future work.

2. State of the art

2.1. The dissemination of disinformation on social media

Although misleading information has existed since the news exists, it has gained attention since the 2016 U.S. presidential election campaign because of the deliberate spread of false information on social media platforms such as Twitter and Facebook [2,20]. During the campaign, the top twenty most-discussed false election contents generated 8,711,000 shares, comments, and reactions on Facebook, whereas the top twenty most-discussed items of news disseminated by the 19 most important news websites generated a total of 7,367,000 [37]. After being present in a large number of electoral processes in countries such as France, the UK, Germany and Spain, concern has increased due mainly to its ability to erode democracy by diminishing trust in public institutions and boosting political polarization [34]. From all the issues regarding the spread of false, manipulated, or inaccurate information, the one that draws the most attention is the so-called 'fake news' defined as 'fabricated information that mimics news media content in form but not in organizational process or intent' [25]. Nonetheless, despite being quite popular, the term "fake news" is very imprecise. In order to deal with the inaccuracy of the term, the Council of Europe refers to "information disorder", a concept which encompasses a) 'Misinformation', related to false content or inaccurate information shared unintentionally; b) 'Disinformation', or false information shared to cause harm; and c) 'Mal-information', when genuine information is shared in order to cause harm, for example by making public information that should stay private [12]. In this work, we focus on 'disinformation', which encompasses two key features: 'Intention' (from the publisher/source) and lack of authenticity (content).

The continuous increase of disinformation online is closely connected to the widespread crisis of confidence in traditional media and public institutions, and the growth in the use of social media to find news. Social media are spaces where consuming news takes less time and is less expensive. Moreover, it is easier to further share and discuss the news with others [36]. As a result, the media landscape has changed: traditional news media have lost their central status as gatekeepers and distributors [9]. The media landscape is now dominated by digital, mobile, and social media platforms that currently shape news consumption, thus acting as the new gatekeepers, so users are exposed to content that has not necessarily verified [4]. This kind of consumption may have other detrimental effects on user's perceptions. On these spaces, 'content flows and reaches web users in an isolated way, with no context and with a weak link to the particular medium which publishes the news', thus there is no formal cue to interpret or detect bias [10]. Moreover, news reaches social media users through automated processes, driven by algorithms that filter a selection of news tailored to each user's personal interests and preferences. As a result, users may only receive information which reinforces their existing views and even their prejudices [7]. This phenomenon, known as "echo chamber" or "filter bubble" [33], challenges to dispel disinformation, and can lead to radicalization, outburst of racial prejudices and xenophobia. Two main motivations lie behind the creation and dissemination of

disinformation: economic (false content fabricated in pursuit of advertising revenues) and ideological (for example, there were more than 100 news sites that made up pro-Trump content traced to Macedonia, according to a BuzzFeed News investigation [2]). Once created, research reveals the role of social bots in spreading disinformation [42]. The low cost of creating social media accounts also encourages the creation of malicious users accounts such as cyborg users, who set automated programs to perform activities in social media, and trolling behaviors [36]. Users themselves may knowingly share false content due to its psychological utility: for example, when they prioritize the expressive functions of language to information. In this case, false content spreads faster since linked to basic feelings such as indignation, fear, or distrust [18]. Moreover, as users prefer to receive news that satisfy their prior opinions, they tend to favor content that confirms one's pre-existing beliefs, ignoring information that challenges them ('confirmation bias') [49]. Furthermore, on social networks, blogs, and platforms such as YouTube, users can start a rumor thanks to the so-called 'cascade effect' [42].

The deliberate spread of online false content has been subject of study in relation to its prevalence in the political area [3]. However, as the COVID-19 crisis escalates, political and health misleading information co-exist in the media ecosystem. In response to the growth in the volume and diversity of disinformation, the number of fact-checks concerning the pandemic has increased [8]. Though previous research has laid the groundwork, disinformation about the pandemic poses new challenges: 1) the characteristics inherent in scientific information, that sometimes may appear obscure and confusing for a part of the audience (even the journalists who 'translate' scientist information into items of news may find it challenging); 2) the lack of previous experience in a crisis of such magnitude; and 3) the generalized climate of mistrust generated from the myriad of contradictory and confusing information emanating from official sources.

2.2. Fake news detection

The detection of inaccurate or misleading content in traditional news media has relied mainly on the analysis of the meta information related to a piece of news. The attributes analyzed are divided into two general categories: 1) source (publisher or author of the news article); 2) content (headline, body text, image/video, or the language, amongst others). These attributes match the key features of disinformation: 'Intention' and lack of authenticity. On the Internet, technical implementations to detect disinformation are based on the content attributes, while at the same time, analyze extra context information inherent in the digital substrate (engagement or diffusion dynamics, amongst others). Related works focus mainly on social media and provide their specific datasets, own usable set of tools and several detection strategies based on machine learning, data mining, natural language processing, information retrieval and social search. For example, from a data mining perspective, different types of feature representations can be built based on news content attributes, to extract discriminative characteristics that help discriminate false content [36]. At the level of contextual information, multiple aspects can be considered such as users' characteristics and profiles (User-based features), characteristics and language of the post (Post-based) and dynamics of each network (Network-based features) [36]. By implementing machine learning methods, others categorize posts as "false" or "true", comparing the performance of several algorithms [24], or study the probability that a given post is true through the analysis of contextual information such as "likes" or followers [11]. The prediction of the information veracity by identifying the features that characterize the information spread has also been analyzed [41]. Other works study the application of deep learning techniques to discriminate fake contents using only text and propose several neural network architectures [22]; or focus on capturing local features of texts and images of the news, applying Convolutional Neural Networks (CNNs) [48]. Recently, the blockchain technology has been proposed to verify news [23].

2.3. The impact of the COVID-19 in Spain

In Spain, the first COVID-19 case (imported) was reported on 31 January 2020, and the first death, on 13 February 2020. These figures increased to a total of 2965 confirmed cases and 84 deaths on 13 March 2020. On 14 March 2020, the Spanish government declared a State of Emergency to slow down the coronavirus outbreak, which lasted until 11 May 2020. Under this status, people's movement were severely restricted, and most shops and businesses were closed. The impact of the disease in Spain during those first days is shown in Figs. 1 and 2. Given the severity of the pandemic and the ease of virus spread, numerous studies on its propagation dynamics have been conducted [16,14,21].

2.4. Fake news detection via the OSINT technique

Information is the resolution of uncertainty, which is given as a set of data on a phenomenon to increase knowledge about it. It has four main dimensions which are meaning, importance, validity, and value. Information management has a 6-phase life cycle: 1) Planning; 2) Creation of information from the data obtained; 3) Storage and distribution of the information by means of computer media; 4) Maintenance and updating; 5) Decision making using the information; 6) Destruction [43]. The importance of information has been accentuated as technology and access to it by institutions and individuals advance. Given its importance, information is an asset that must be protected. Sometimes cybercriminals access information illegally using hacking techniques. However, sometimes these techniques are not needed as the information is openly available on the Internet. This is where OSINT techniques come into play. The collection and processing of information from public sources to generate intelligence has existed for many years. The United States began conducting OSINT through the Foreign Broadcasting Monitoring Service (FBIS) created in 1941. During the Second World War, this agency carried out OSINT obtain-

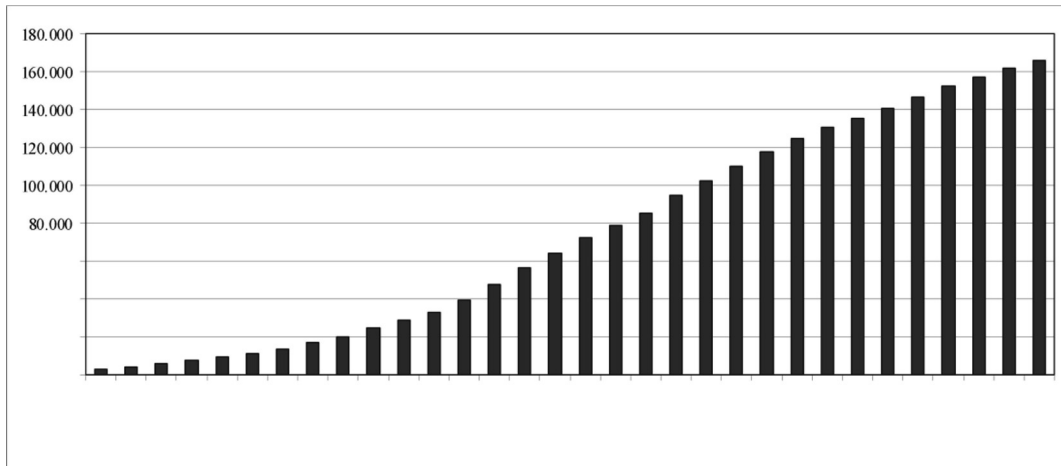


Fig. 1. Confirmed cases in Spain from March 13th to April 13th, 2020. . Source: [44]

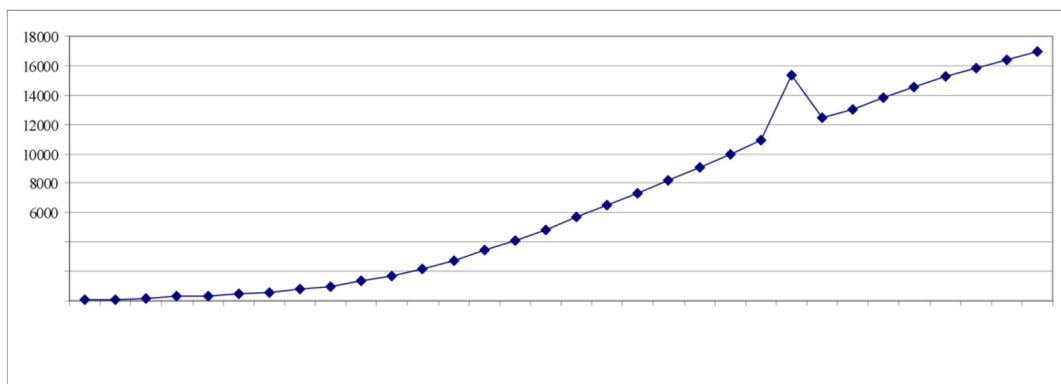


Fig. 2. Deaths in Spain from March 13th to April 13th, 2020. . Source: [44]

ing information as valuable in a war context as knowing whether the railway bridges had been successfully bombed using the price of oranges in Paris as an indicator [5]. With the advance of technology and telecommunications this practice has been facilitated. In fact, in 2005 the US national intelligence agency announced the creation of the National Open Source Center (NOSC) to collect open information available in the media, including the Internet [31]. The private sector (both companies and individuals) has also been involved by developing tools that help collect and analyze the available information. These tools enable the collection of information (legislative, political, geospatial, personal, financial, electronic, telemetric, etc.) on targets (individuals, organizations, companies, and countries) for subsequent analysis and intelligence generation. Some examples of these OSINT tools are FOCA, Shodan and Maltego, as shown in Table 1.

The open-source intelligence is the intelligence that meets these requirements [29]:

- 1) It is produced from information that is publicly accessible (media, public information from government sources, forums, RSS, blogs, papers, online libraries, free or paid video on demand, etc.).
- 2) It provides specific intelligence.
- 3) It is collected, processed, and disseminated to a specific audience.

Table 1
Examples of tools to perform OSINT. Source: Own elaboration.

Name	Description
FOCA Tool	Search for information contained in file metadata.
Shodan	Search for information about computers connected to the Internet.
Maltego	Open source intelligence and forensic software to analyze relationships between social networks, websites, domains, networks, etc.

The first point means that access to information does not require specialized techniques or knowledge, not requiring extraordinary effort or illegality. The technologies and technical knowledge required are for the automation of information collecting and processing. The concepts of data, information and intelligence should not be confused. Data are collected values. Information is formed from the relationships between data. Intelligence arises from information with the knowledge we have about it. From the data, the information is generated, and from the information, intelligence is generated, as shown in Fig. 3.

Examples of open sources are official gazettes. An official gazette is a written medium by which an entity in an administrative region, such as a State, publishes legal regulations of a public nature. The first publication of this type was La Gazette, in 1631 in France by the journalist *Théophraste Renaudot* [5]. The following are official gazettes from Spain:

1. Supranational level: Official Journal of the European Union (OJEU).
2. National level: Official State Gazette (BOE) and Official Gazette of the Commercial Registry (BORME).
3. Sub-national level: Autonomous Community bulletins (autonomous communities and cities) and provincial bulletins (in the case of multi-provincial Autonomous Communities).

The services currently available to carry out OSINT in official bulletins are reviewed below. The first tools are the usual search engines such as Google, using for example dorks to refine the search on the newsletter. You can filter by the newsletter you need to consult and by keywords. This method has several limitations. For example, it does not allow to search for certain information in more than one bulletin at a time. Besides, it does not represent the information making its access easier. Moreover, the way of searching by keywords limits the selection of information. Another disadvantage of this method is that there are numerous bulletins not indexed by search engines, by the will of the government or for legal reasons such as the right to forget. At present, there are approximately 20,000 official state bulletin documents de-indexed from search engines using the robot exclusion protocol (robots.txt file). Another possibility to search for information are the search engines of the web pages of the bulletins themselves, but the search is limited to a specific bulletin. Furthermore, this form of search is not available in all bulletins, such as the official bulletin of the autonomous city of Ceuta. One of the best current alternatives is IBOOF. This service has the advantage that it allows to search in several official bulletins at the same time, doing a search by keywords. As a disadvantage, it is not a generic search engine with the capacity to adapt to any bulletin (for example, it does not consider the official bulletin of the Spanish commercial registry or the bulletin of the European Union). On the other hand, it does not represent the information facilitating its analysis. Another tool is LibreBOR.me. This is a platform to consult and analyze the Official Gazette of the Commercial Registry. While this tool does not facilitate data analysis and does not provide all the historical data, the LibreBOR.me tool does. Table 2 summarizes the characteristics of each of the current alternatives. It should be noted that some bulletin websites do not have an internal search engine. In addition, the Google search engine blocks content using robots.txt [19].

The possibility of obtaining information from open sources and facilitating its subsequent analysis is the motivation of this work. The aim is to design and prototype a modular tool for this purpose. It must be modular to adapt to the different input formats, since each bulletin has its particular characteristics when representing the information (for example, different formats: PDF, TXT, HTML, etc.). It should also be modular when processing the captured information to study it. A prototype is developed, and some example cases show that this tool improves to a greater or lesser extent the deficiencies detected in other tools, maintaining the possibilities that already exist. Moreover, through a Case-based Reasoning (CBR) system, the tool provides an explanation about the classification of the news. The proposed CBR-based tool, called MedOSINT (*Medical OSINT*), is described in the next section.

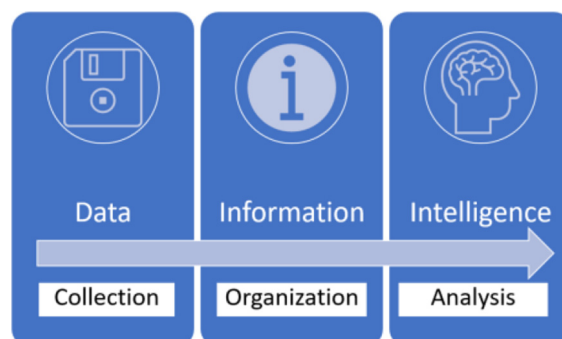


Fig. 3. Relationship between data, information, and intelligence. . Source: [5]

Table 2
Comparison of current tools. Source: Own elaboration.

Feature	Google search	Internal search engine of each newsletter	www.who.int bulletin	IBoOF	LibreBOR.me
Locked content	Yes	No	No	No	No
Restricted search	No	Depending on each bulletin	No	No	No
Indexed search and not just by words	No	No	No	No	Yes
Display information from any newsletter	Yes	No	No	No	No
Representation of the information to carry out OSINT	No	No	No	No	Yes
Supports complementary external information for analysis	No	No	No	No	No

3. MedOSINT tool

3.1. Objectives

Since there is currently no effective tool for analyzing official bulletins, the aim of this work has been to design and prototype a tool for this purpose. The general objective is that MedOSINT can show the complete information of the BOE, BORME and the official bulletin of the European Union in a precise way by means of indexed searches of the information and with a good representation to carry out OSINT.

The specific objectives in terms of capabilities of MedOSINT are the following:

1. Displaying information without blocking it.
2. Precise search through indexed searches.
3. Using a representation that facilitates the analysis of the information.
4. To be able to use information external to bulletins to improve the outcome of the analysis (charts and tables).
5. Through a CBR explanatory system all news are classified as false (4 types) or true, enabling an explanation.

3.2. Architecture

The architecture of the proposed coronavirus news verification tool has 3 stages. Fig. 4 shows the architecture of MedOSINT. The first stage is the verification of the news by MedOSINT. News is taken randomly from the internet. This news may be false or true. MedOSINT first classifies the news. Then it checks the veracity of the news on authorized websites. Once its veracity is detected, the news item is sent to the case database (second step). In the third step the CBR feeds on cases in the

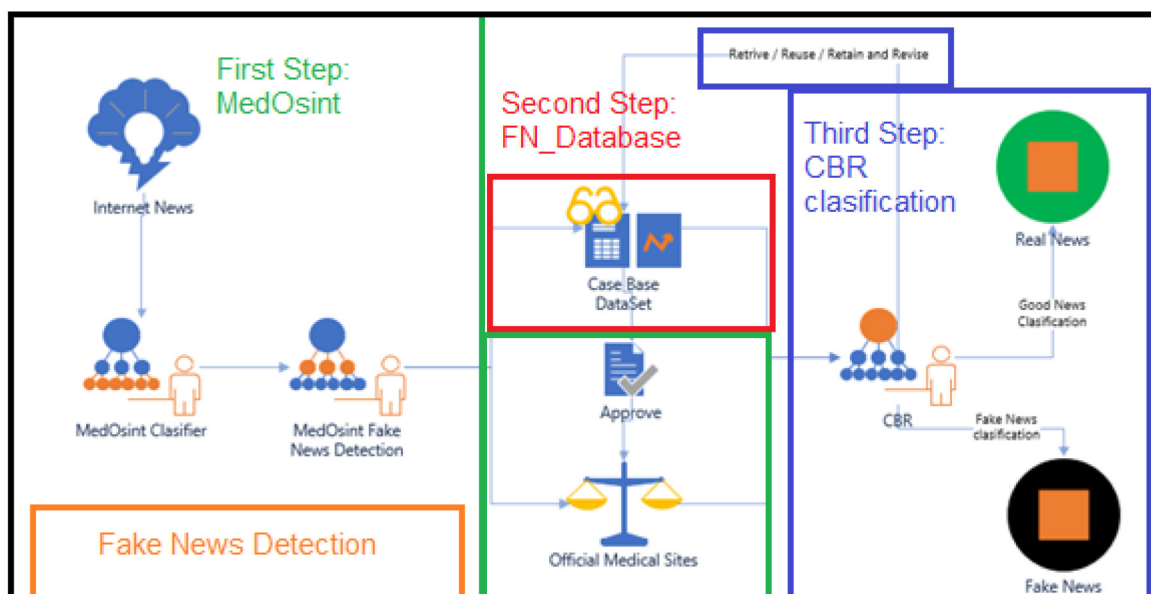


Fig. 4. Architecture of MedOSINT. Own source.

database. The CBR, using the reuse, retrieve, retain and revise technique, classifies news when it finds one with a close similarity index in its knowledge database. Fig. 4 shows the general diagram of fake news detection. MedOSINT appears in the first place. The function that MedOSINT has is the detection of fake news, verifying the news with official sources such as the WHO. There are five official sources with which the news is commonly contrasted. In this step it is necessary for a systems analyst to take the news and process it in MedOSINT to detect whether the news is false or not. In a second step, the systems analyst takes this news to the database, which stores both true and false news. This database is used, in a third step, by the CBR. The CBR takes all the case studies and classifies them as true or false. If a news item has already been validated, it omits it to work faster. The CBR does not detect the veracity of the news because this is done by MedOSINT, but rather classifies them. Fake news tends to reproduce in several portals or unreliable sources such as social networks, and thus become viral, so the CBR is very useful because it establishes whether the news tends to be more false or true using a similarity metric.

3.3. Design and working of the MedOSINT tool

In this and next sections, the MedOSINT tool is described and its design and working method are explained from the programmer and technical points of view. MedOSINT is a tool to analyze information from official bulletins for intelligence purposes. It collects a series of information, processes it and presents it properly. It allows queries on the information loaded, which has been previously processed to organize it and facilitate consultation. The development of the prototype has two distinct parts. First, a solution is needed to store the information and do the processing. This is a development on an Oracle Database Express Edition 18c, which is a modern technology capable of supporting this implementation. On the other hand, the MedOSINT tool must show the information to the user. This information is intended for a professional intelligence analyst. The second part of MedOSINT is a web application to show the information. The Ionic and Angular frameworks have been used for the web application because they are state of the art technologies, with great support, that allow an easy maintenance and that are flexible allowing the execution in multiple devices. Table 3 shows a list with all the technologies used and their versions.

3.4. MedOSINT flowchart

The steps and the operating characteristics of the tool are described below. Fig. 5 shows the MedOSINT flowchart. This is the process to verify the news. To explain the detail of the diagram it has been divided into 11 steps. They are described below from sections 3.5 to 3.16. When stage 11 finishes, the next step is to enter the classified information into the oracle database. Then, this news feeds the CBR so that, when a false or true news is repeated, it is searched in the repository and classified based on the similarity metric.

3.5. Step 1) Selection of the information to be analyzed

In this paper we show five examples of system use to facilitate the reader's understanding, but any other data could be analyzed. These examples provide a guide to repeating the process with any other topic. The classifications chosen for the analysis are as follows: 'Economy/Companies', 'Politics', 'Legal' and 'Miscellaneous'. Within 'Economy/Companies' the topics selected are: 'Announcements of company incorporation by province' and 'Relationship between company incorporation announcements and COVID-19 information search'. For the other three classifications, three topics have been selected, as shown in Table 4.

Table 3
Technologies used for the development of MedOSINT. Source: Own elaboration.

Technology	Version	Remarks
body-parser Library	1.19	Data Parsing (JSON)
cors Library	2.8.5	HTTP Access Control
Framework Angular	8.2	Web application development
Framework Express	4.17.1	HTTP requests
Framework Ionic	5	Functionality and visual details
Lenguaje Typescript	3.4.5	Programming language
Materials Module	8.2.3	Visual details and charts
ng2-charts Module	2.3.0	Charts
Node.js Environment	12.13	Asynchronous execution environment
npm package manager	6.12	Package management for Node.js
Oracle Database Express Edition 18c	18.0.0.0.0	Database
Oracle SQL Developer	19.2	Oracle Development Environment
oracle-db Module	4.2	Access to Oracle Database
Visual Studio Code	1.44.2	Angular development environment

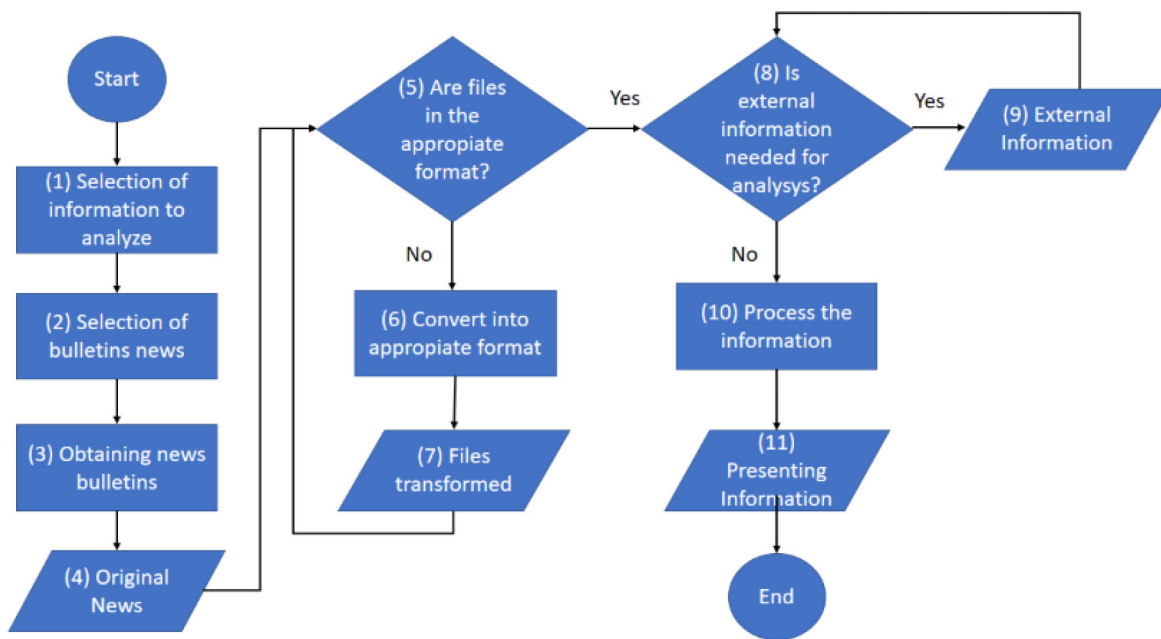


Fig. 5. MedOSINT Flow chart. Source: Own elaboration.

Table 4

Classifications and subjects of the analysis. Source: Own elaboration.

Classification	Topics
Economy & Companies	Company Formation Announcements by Province (Topic A, TA) Relationship between company incorporation announcements and COVID-19 information search (Topic B, TB)
Politics	Evolution of policy type development (Theme C, TC)
Legal	List of pardons (Theme D, TD)
Miscellaneous	Evolution of the development of the amount of legislation (Theme E, TE)

3.6. Step 2) Selection of the bulletins involved

The bulletins involved are the Official State Gazette, the Official Gazette of the Commercial Registry, and the Official Journal of the European Union. These are selected because they contain the information needed to make the analysis as shown in Table 5. This does not mean that the MedOSINT tool can only work with these 3 bulletins, but for this concrete test development, modules for these 3 bulletins will be implemented. It is perfectly feasible to use MedOSINT with any other bulletin considering the same particularities that are considered for these 3 bulletins in the following parts of the process. For TA and TB, the information is in the BORME. For TC and TE the information is in the BOE and the DOUE, since they are the official legislative communication journals. For TD, the information is in the BOE, in provisions of the Ministry of Justice.

3.7. Step 3) Obtaining the bulletins

For each of the 3 bulletins indicated in Table 5, the sample files to be uploaded to the database are obtained during the first 4 months of the year 2020 as a reliable sample. It should be noted that it is unfeasible to do this development without

Table 5

Bulletins from which the necessary information is obtained by subject. Source: Own elaboration.

Topics	Bulletins from which the necessary information is obtained
TA	BORME
TB	BORME
TC	BOE and DOUE
TD	BOE
TE	BOE and DOUE
WHO	World Health Organization Newsletter

obtaining the bulletins locally instead of processing the files directly from the repository website of each bulletin service because of the huge volume of existing files.

3.7.1. Obtaining bulletins from the BOE

In the case of the BOE, this institution offers open access to the archives. It also offers, in its open data section, a PHP script for downloading by date.

Obtaining bulletins of the DOUE:

For the DOUE, the EUR-Lex institution also offers open access to the archives. In fact, it is possible to download packages by language and year, but currently it is not possible to download a file with all the bulletins of the year 2020. Therefore, it is necessary to create a script for automatic retrieval. Then, you have to get all the links of the year 2020 bulletins from the repository. The repository search engine looks for the year 2020 newsletters and their download URL, finds the newsletter identifier of the link, and lists all the identifiers of all the newsletters obtained. In this way it is possible to create a list of all the download links of all the newsletters from the general link and the specific identifiers of each one that are available in the source code of the page.

3.7.2. Obtaining bulletins from the BORME

For the BORME the process is the same as for the BOE, except for the script to be used. The PHP download script is available on the website of the Official State Gazette [6].

3.8. Step 4) Loading of raw bulletins

This section explains how the raw data obtained is stored. Regarding the table scheme for loading the raw bulletins, a table is required for the coding of the bulletins. In this table, you define the types of bulletins to be processed and assign a unique identifier for them. The second table is for sub-coding bulletins, since there are several types of bulletins, such as supplemental bulletins. The third table for coding provinces is required specifically for TA. The numbering chosen is the one proposed by the INE. These 3 tables are preloaded with the necessary data (Table 6 shows the required data for bulletin coding and sub-coding tables).

The fourth table shown in Table 7 is where the sample bulletin files are loaded. The 'ID' field takes auto-incremental numeric values to be used as unique file identifiers. The code fields are for sorting data according to the 3 tables above mentioned. We also have a field for the name of the original document (e.g. 'BOE-A-2020-4738.pdf') and another field of type BLOB to load the file. The other 3 fields are a CLOB field for uploading the adapted file (which will be explained in steps 5 and 6), a field for the date of the document publication/signature, and another field for the date when the upload is made.

Once the tables have been created, all the documents that will be in a directory on the machine are loaded. With their name it is possible to know the bulletin code and the bulletin detail code. For example, with the file name 'BOE-A-2020-4738.pdf', you can know that the bulletin code is 1, corresponding to the BOE according to the bulletin coding table, and you can know that the bulletin detail code is A, corresponding to 'General' according to the bulletin detail coding table. This is processed at the technical level with regular expressions. The field for the adapted document and the publication/signature date will also be filled in, if the file is in a valid format for uploading, e.g. in plain text. This will be known in advance by the file extension and the bulletin code.

3.9. Step 5) Valid format for processing

For each type of newsletter downloaded, there is a file extension and a presentation format for the information it contains. If these are not suitable for reading and processing, an adaptation will be needed.

3.10. Step 6) Convert to a valid format

The downloaded files of the BOE and the BORME bulletins are in PDF format, while the DOUE file is in HTML format. These formats are not adequate for reading text from the files automatically. For example, HTML contains an excessive number of tags, spaces, tabs and other characters that increase the size of the file and do not help in loading and reading efficiently. Therefore, the files are converted to *.TXT format. This conversion is done with available tools that allow massive file conversions.

3.11. Step 7) Uploading bulletins in a valid format for processing

Once converted to the valid *.TXT format, each document is loaded into the CLOB-type field with the name 'ADAPTED_DOCUMENT' adding the signature date field by reading the line number where the date is and using regular expressions. If the gazette code is 3, the province code field is also filled in with a regular expression using the province code table, since this gazette has editions by province and will be used for TA.

Table 6
Bulletin file upload tables. Source own.

Bulletin coding table:		
Nombre	¿Nulo?	Tipo
CODIGO_BOLETIN	NOT NULL	NUMBER(3)
DESCRIPCION		VARCHAR2(200)
CODIGO_BOLETIN DESCRIPCION		
1		Boletín Oficial del Estado
2		Boletín Oficial del Registro Mercantil
3		Diario Oficial de la Unión Europea
Bulletin sub-coding table:		
Nombre	¿Nulo?	Tipo
CODIGO_BOLETIN	NOT NULL	NUMBER(3)
CODIGO_DETALLE_BOLETIN	NOT NULL	VARCHAR2(3)
DESCRIPCION		VARCHAR2(200)
CODIGO_BOLETIN COD DESCRIPCION		
1		A General
1		S Suplemento
2		A Actos
2		C Convocatorias
2		S Suplemento
3		L Legislación
3		C Comunicaciones e informaciones

Table 7
Sample bulletins table. Source own.

Documents table:		
Nombre	¿Nulo?	Tipo
ID	NOT NULL	NUMBER(6)
CODIGO_BOLETIN		NUMBER(3)
CODIGO_DETALLE_BOLETIN		VARCHAR2(3)
CODIGO_PROVINCIA		NUMBER(2)
NOMBRE_DOCUMENTO		VARCHAR2(200)
DOCUMENTO_ORIGINAL		BLOB
DOCUMENTO_ADAPTADO		CLOB
FECHA_FIRMA		DATE
FECHA_CARGA		DATE

3.12. Step 8) External information required for analysis

At this point it is necessary to identify whether it is necessary to add external information to the database to carry out the analysis. For TB it is necessary, as this topic is about the relationship between certain data in the BORME with data obtained from Internet searches.

3.13. Step 9) Loading external information

The necessary data on Internet search trends identified in the previous section will be obtained from Google Trends as shown in Fig. 6. From the website the data is exported in CSV format. Then, this data is inserted into a new auxiliary table called 'GOOGLE_COVID_19'. The data is inserted as offered by Google and making the monthly average. This will be explained in steps 10 and 11.

```
Nombre ¿Nulo? Tipo
-----
ENERO_2020 NUMBER(8)
FEBRERO_2020 NUMBER(8)
MARZO_2020 NUMBER(8)
ABRIL_2020 NUMBER(8)
```

Nombre	¿Nulo?	Tipo
ENERO_2020		NUMBER (8)
FEBRERO_2020		NUMBER (8)
MARZO_2020		NUMBER (8)
ABRIL_2020		NUMBER (8)

ENERO_2020	FEBRERO_2020	MARZO_2020	ABRIL_2020
0	2	53	62

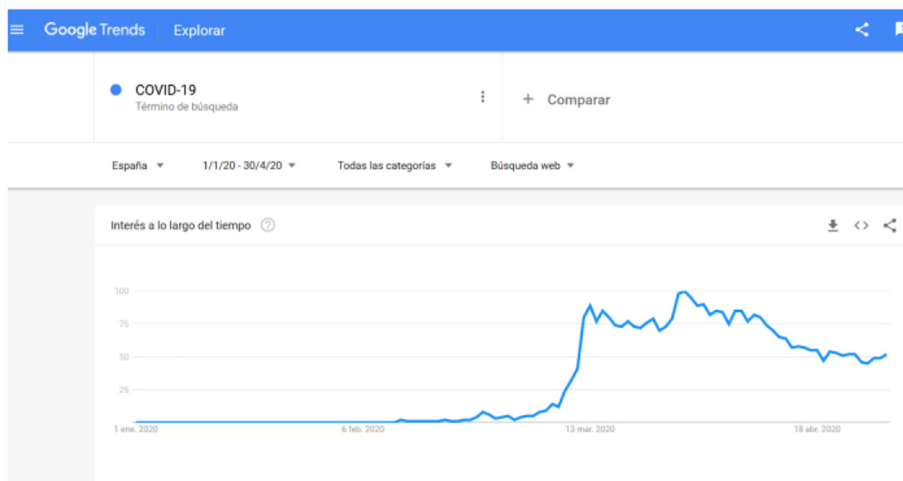


Fig. 6. Search trend of the term 'COVID-19' in relation to interest over time . Source: [19]

ENERO_2020	FEBRERO_2020	MARZO_2020	ABRIL_2020
0	2	53	62

3.14. Step 10) Processing all the information.

The following explains how to implement the processing of the information loaded in MedOSINT for the topics described in step 1. To obtain the number of company incorporation announcements from each BORME bulletin and from each province, the regular expression shown below is used.

```
REGEXP_COUNT(DOCUMENTO_ADAPTADO, ' Constitución. ', 1, 'i')
```

Each coincidence in the expression will correspond to a declaration of company incorporation as shown in Fig. 7. The PDF document in the figure shows two company incorporation announcements, which when converted to a valid format for processing, remain as a character string that complies with the above regular expression.

In this way, considering that the bulletin files are already classified by date and province, a list is obtained with the number of company incorporation announcements by province and date. Note that the date of the beginning of the company's activity and the date of the announcement of its incorporation in the BORME do not have to coincide exactly.

3.15. Relationship between company incorporation announcements and information search

The data obtained in the previous step and those obtained in step 9 in Google Trends are processed to check if there is a relationship between them. For this purpose, we obtain the total number of company incorporation ads from all the provinces grouped by month. This development has been limited to 4 months. At the end, the percentage that each month represents with respect to the statistical average of the 4 months is shown. Fig. 8 shows an example of the statistical average resulting from the 4-month sample.

(23.12.19).

4 - EVOSEGUR SL.

Constitución. Comienzo de operaciones: 5.12.19. Objeto social: 1. Acti mantenimiento de aparatos, equipos, dispositivos y sistemas de seguridad cor de control o de videovigilancia. 2. Fabricación, comercialización, venta, entreg GAUDI 5 - PARQUE EMPRESARIAL (ELCHE). Capital: 3.000,00 Euros. PRUDENCIO;ESCRIBA VALERO JEAN PIERRE. **Datos registrales.** T

5 - PEDRO INIESTA SA.

Reelecciones. SecreNoConsj: CANDELA ALVAREZ PABLO. Consejer DIEGO;INIESTA ESCRIBANO PEDRO. Presidente: INIESTA MOREN PEDRO;INIESTA BOTELLA DIEGO. **Datos registrales.** T 1511 , F

6 - RA-EL CALZADOS SL.

Nombramientos. Apo.Sol.: FERNANDEZ JIMENEZ ANA BELEN;FERNANDE F 170, S 8, H A 37134, I/A 4 (23.12.19).

7 - MARINA ALTA AEREOVISUAL SL.

Nombramientos. Apoderado: REIG MORA ROMAN. **Datos registrales.** T 42:

8 - REHABILITACIONES Y CONSTRUCCIONES COSTA ALICANTE SL.

Constitución. Comienzo de operaciones: 13.12.19. Objeto social: La cor

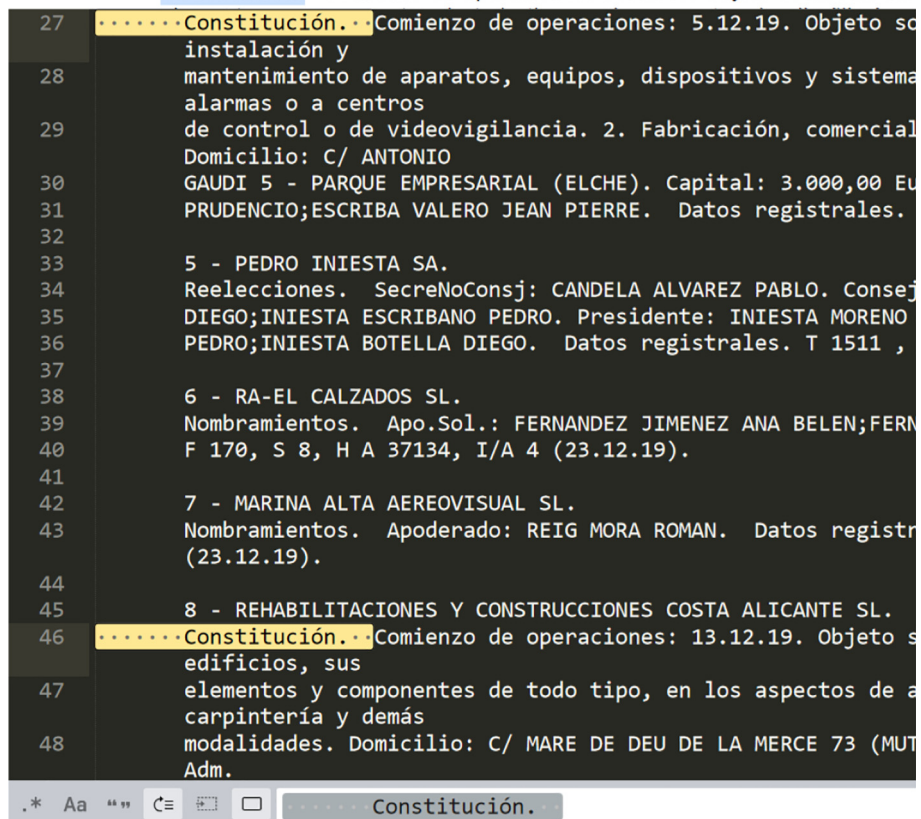


Fig. 7. Identification of company incorporation announcements in the BORME. Source: Own elaboration.

3.15.1. Evolution of policy developments

To analyze policy development over time, three different types of policy have been selected as a sample: health, economic and environmental/ecological. One way to approximate what kind of policies are being developed in official journals is to find keywords related to each of these policies. For each of these three policies, the keywords that appear in [Table 8](#) have

```

Value of the month 1 = 10
Value of the month 2 = 20
Value of the month 3 = 10
Value of the month 4 = 40

4 months average =  $\sum (4 \text{ months value})/4 = (10+20+10+40)/4 = 20$ 
Result of the statistical average of month 1 = Value of month 1/Average of 4 months =  $10/20 = 0.5$ 
Result of the statistical average of month 2 = Value of month 2/Average of 4 months =  $20/20 = 1$ 
Result of the statistical average of month 3 = Value of month 3/Average of the 4 months =  $10/20 = 0.5$ 
Result of the statistical average of month 4 = Value of month 4/Average of the 4 months =  $40/20 = 2$ 

```

Fig. 8. Example of a sample calculation of the statistical average and COVID-19 search.

Table 8
Key words associated with each type of policy. Source: Own elaboration.

Policy type	Associated keywords
Health	'health', 'nurse', 'vaccine', 'medical', 'covid' and 'coronavirus'
Economic	'debt', 'employment', 'bonds', 'fund', 'liquidity' and 'provision'
Environment/Ecology	'CO2', 'carbon', 'transition', 'ecology', 'green' and 'environment'

been selected. The keywords are selected considering the whole family of associated words. This means that from the keyword 'medical', all the words in the family will be found as medication or medicines.

Once the words have been selected, it is necessary to obtain the number of occurrences of each group associated with a type of policy, grouped by bulletin and by the study time interval. To find the frequency of occurrence, the query with the following regular expression is used with each group of keywords.

```
SUM(REGEXP_COUNT(replace(UPPER(DOCUMENTO_ADAPTADO), chr(10), '' ), UPPER('health|nurse|vaccine|
medical|covid|coronavirus'), 1, 'i'))
```

Finally, in the same way that it is done for TB, the percentage that each month represents with respect to the statistical average of the 4 months is obtained.

3.15.2. List of pardons

To obtain the name of the persons pardoned from the BOE, the regular expression shown below is used.

```
UPPER(REGEXP_REPLACE(REGEXP_REPLACE(REGEXP_REPLACE(REGEXP_REPLACE(REGEXP_REPLACE(REGEXP_REPLACE(
REPLACE(DOCUMENTO_ADAPTADO, chr(10), '' ), '([[:print:]]+, por el que se indulta a) ([[:print:]]+ ) (
[[:print:]]+)', '\2'), ' don ', '' ), ' doña ', '' ), '[ ]{2,}', '' ), '\.', '' ), ' Visto el expediente de indulto
de.*', '' ))
```

This expression returns the full name of each pardoned person, which will correspond to an announcement. The PDF document in Fig. 9 shows a pardon announcement, from which, when converted to a valid format for processing, the name of the person pardoned with the above regular expression can be obtained.

3.15.3. Evolution of the development of the amount of legislation

In a similar way to TB processing, the amount of legislation is obtained as the percentage of each of the 4 months to be represented with respect to the average. The quantity of legislation is defined as the size it occupies when stored. The size is directly proportional to the number of characters it contains. The function `dbms_lob.getlength()` is used to check the size of each BOE and BORME bulletins.

3.16. Step 11) Presentation of the processed information

For the presentation of the information, a website developed in Angular was implemented. This website presents a simple interface to make queries on the data obtained in step 10. To perform database queries, the 'oracle-db' module available on the NPM website [32] is used. The results of the queries are sent to the web page through the 'express' module also available in NPM. Once on the query web page, the previous data will be accessed through a service in Angular. A series of components from the Material and 'ng2-charts' libraries of Angular are used. To represent the information, the users must select which information they want to consult. To select the topic and the theme to be consulted from Table 3 of step 1, the system uses a

The image shows a screenshot of the BOLETÍN OFICIAL DEL ESTADO website. At the top, there is a search bar with the text 'Buscar (1/1)' and a search input field containing 'por el que se indulta a'. Below the search bar, there are buttons for 'Anterior' and 'Siguiente'. The main content area displays the following information:

Núm. 38 Jueves 13 de febrero de 2020 Sec. III. Pág. 14174

III. OTRAS DISPOSICIONES
MINISTERIO DE JUSTICIA

2139 Real Decreto 367/2020, de 11 de febrero, por el que se indulta a don Erbing Wang Chen.

Visto el expediente de indulto de don Erbing Wang Chen, condenado por el Juzgado de lo Penal número 15 de Valencia, con sede en Alcira/Alzira, en sentencia de 15 de abril

The screenshot also shows a search bar at the bottom with the text 'por el que se indulta a' and a search button.

Fig. 9. Identification of pardon announcements. Source: Own elaboration.

component 'mat-horizontal-stepper', components 'mat-select' and a component 'mat-form-field' to filter by name in the case of TD. Finally, the information is represented by tables with the component 'mat-table' and by graphics with the module 'ng2-charts'. For each theme, there is a consultation section with a Typescript module that obtains the information from the database and shows it as a table or a graph, as required by each theme.

3.17. Using Case-based reasoning (CBR) to classify news

Case-based reasoning is the process of solving new problems based on solutions to previous problems. This works to remember previous cases of fake news. For example, if a news item has the same features and is only differentiated by the information source, it is probably a copy-paste from another site. This is how fake news spreads through the internet and social networks. CBR can utilize the specific knowledge of previously experienced cases. A new problem is solved by finding a similar past case and reusing it in the new problem situation. CBR is an incremental approach, sustained learning, since a new experience is retained each time a problem is solved, making it immediately available for future problems. Therefore, CBR allows solving a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation. Let us illustrate this by looking at some typical problem-solving situation. For example, a physician - after having examined a particular patient in his office - remembers a patient he treated two weeks ago because of the similarity of symptoms. Therefore, the physician uses the diagnosis and treatment of the previous patient to determine the disease and treatment for the patient in front of him. As this example shows, reasoning by re-using past cases is a powerful and frequently applied way to solve problems for humans. This claim is also supported by results from cognitive psychological research.

We build a classifier based on CBR to classify Fake News. We use a Case-Based Reasoning (CBR) system proved to be an ideal complement because it finds explanatory cases for an explanation-by-example justification [1]. Common algorithms are trained with large volumes of data, including articles, newspapers records and official bulletins. However, these solutions do not provide an explanation of the classification made. CBR systems have a “natural” transparency as they are based on the reuse of previous experiences or examples.

Case-based reasoning is a paradigm for combining problem-solving and learning that has become one of the most successful applied subfields of artificial intelligence in recent years. CBR is based on the intuition that problems tend to recur, so that new problems are often like previously encountered problems and, therefore, past solutions may be used in the current situation [26]. CBR basically consists of a retrieval stage where the cases similar to the given query are retrieved, and an adaptation stage that combines the retrieved cases to build a solution. When applied to classification problems, CBR is considered a lazy learning approach because, instead of generating an abstract representation of the set of training examples, it uses those training examples closer to the problem example (k Nearest Neighbors) to determine its class.

As CBR is considered a transparent technique, then it is applied to explain opaque machine-learning techniques such as neural networks [40]. CBR systems follow an explanation-by-example approach where the explanation given to the user is based on the comparison of the input with similar examples obtained by the CBR system. This way, every instance in the

dataset can be considered as an explanatory case. However, the performance of the CBR process is influenced by the number of features/variables that represent each case due to the underlying nearest neighbor method.

In our system, the attributes analyzed are divided into two general categories: 1) source (publisher or author of the news article); 2) content (headline, body text, image/video, or the language, amongst others). The prediction is explained by the most similar cases, following the explanation-by-example paradigm. The information given by MedOSINT has been reused to start with explanatory cases of classification of true or fake news as reported in official bulletins. This allows to create a case repository for the CBR system. The classification given by the CBR system consists of 4 types of non-authorized news and one type of authorized news as shown in Table 9. Table 10 shows the relevance percentage from the variables selected to describe the explanatory cases.

Fig. 10 shows the relationship between MedOSINT and the CBR. It starts with the incoming news, whose source can be the internet, an email, or a social network. Then, MedOSINT takes this news and contrasts it with the official media such as the

Table 9
Variables selected to describe the explanatory cases. Source: own.

Acronym	Class	Description
FN1	Fabricated news	Completely fictitious medical facts or events (for example, stories related to disease negation)
FN2	Manipulated news	Generally true basic information, but false conclusions or recommendations coming from the over-interpreted or overly extrapolated results (for example, data from in vitro studies presented as a readily available option for patients)
FN3	Advertisement news	Stories about diseases, often critical towards conventional therapies, designed to advertise ‘miracle products’ (for example, alternative treatments)
FN4	Irrelevant news	Not directly related to health (for example, name of the disease used as a metaphor)
GN	Sufficient news	Generally true and evidence-based information about the disease

Table 10
Relevance percentage from the variables selected to describe the explanatory cases. Source: own.

Acronym	Relevance		Percentage	Explanation
	From	To		
FN1	301	400	100%	Completely fraudulent
FN2	201	300	75%	Falsification of content
FN3	101	200	50%	Mixture of truth and lies
FN4	1	100	25%	True news mixed with false incongruities
GN	0	0	0%	Real News

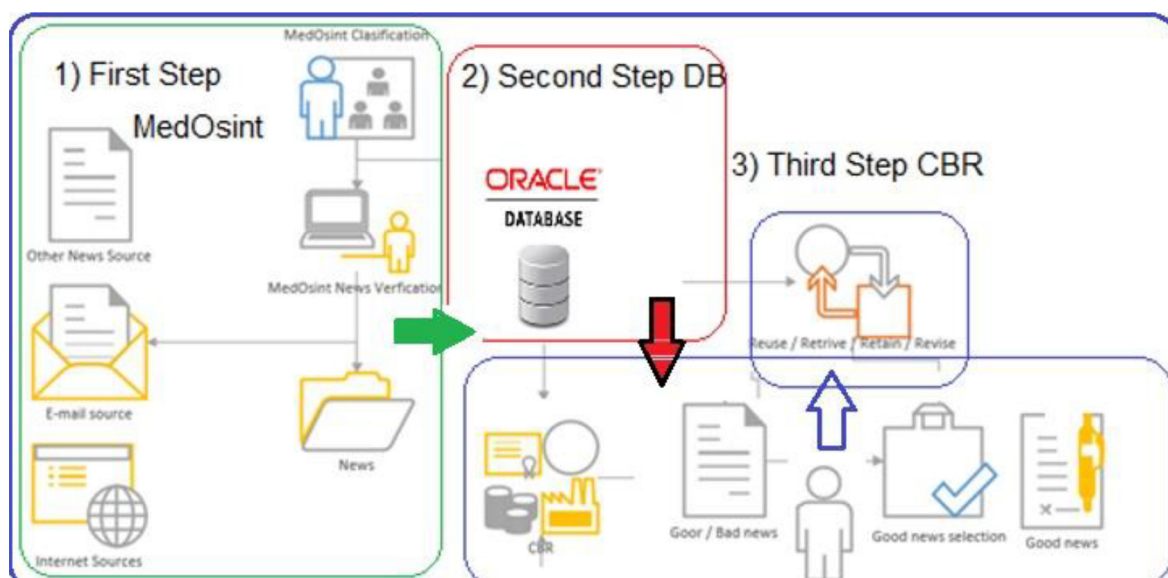


Fig. 10. Relationship between MedOSINT, dataset information and CBR results. Source: own.

WHO. Once the news is classified as true or false, it is stored in a database. From this database the CBR takes the news as an input source for the process of solving new problems based on the solutions of previous problems. This means that from a piece of news, which is false but spread in 5 different newspapers, the case-based reasoning cycle classifies such news into cases in different sub-processes. Specifically, the cycle is divided into 4 clearly differentiated processes:

RECALL cases similar to the one we are analyzing.

RE-USE the information and knowledge of similar cases to solve the problem.

REVIEW the proposed solution.

RETHINK the parts of this experience that can be useful for solving future problems.

All the news items are included within the CBR explanatory system and are classified according to the types indicated in the Table 9. Finally, the CBR explains why the incoming news is false or true and classifies it as a new case according to its falsity or veracity (see Fig. 10).

4. Results analysis and evaluation

This section shows the operation of the MedOSINT tool from the user's point of view. Firstly, the general interface of MedOSINT is presented and secondly, the processing results are presented. Next, a review of the tool and its operation is made to check if it meets the requirements initially established. Finally, it is compared with the other competing tools described in the state of the art section.

4.1. MedOSINT general interface

MedOSINT interface works on the World Wide Web (WWW). MedOSINT has a menu on its left side with several sections. The first section is 'data query', which allows to query data of each topic according to the classifications of Economy/Companies, Politics, Legal and Miscellaneous described in Table 4. Fig. 11 shows the query web page with the left side menu and the data query section.

The second section entitled 'Explanation of processing' shows how the information has been processed. This section and the previous one are the most important as they show the information processed and how the processing was done. This is very important for an intelligence analyst. The third section entitled 'Official Bulletins' shows the bulletins processed by MedOSINT. These bulletins are those described in Table 5. In the fourth section entitled 'About', additional information about the tool is shown.

4.2. MedOSINT fake news study cases

4.2.1. International fake news

According to an article [28] "Nearly 99 per cent of COVID-19 patients who are vitamin D deficient die, according to a terrifying study that adds to mounting evidence that the 'sunshine' nutrient could be a coronavirus life-saver." However, verification shows that this COVID-19 study linking deaths to low vitamin D levels is flawed.

MedOSINT, in a simple query, detects that this is fake news. The article had been analysed by the WHO [45].

4.2.2. Spain fake news

Another source claims that masks create serious health risks [27]. MedOSINT detects that it is fake news, as reported by the WHO web site [45].

4.3. MedOSINT queries

This section shows how to query and the result obtained for the different topics mentioned in Table 4. It should be noted that the analysis, study, and interpretation of the results is not the object of this research. This task concerns only the user.

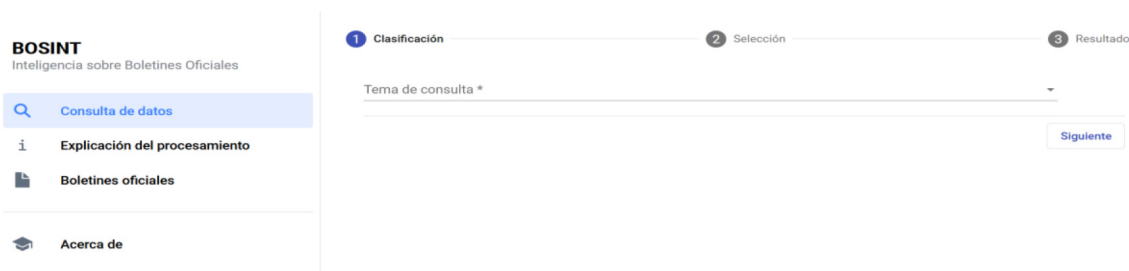


Fig. 11. MedOSINT data consultation section. Source: Own elaboration.

MedOSINT only shows the information. The information contained is the exclusive property of the official gazettes and third-party sources.

4.3.1. Topic A) Announcements of incorporation of companies by province

The classification corresponding to the topic is Economy/Companies (as shown in Table 4). Then, the user must select the category Economy/Companies. Fig. 12 shows the way to consult and the result in a table. The date of incorporation of a company can be up to 2 months prior to its announcement in the BORME. This is indicated to the user with a message and with the reference to the BOE that justifies it.

4.3.2. Topic B) Relationship between company incorporation announcements and COVID-19 information search

This topic is contained in the Economy/Companies classification. To consult about it, the user must select this classification. Fig. 13 shows the query form and the result of the query in a bar chart.

4.3.3. Topic C) Evolution of policy development

This topic is contained in the Policy category, so the user must select this category to view the evolution of policy development. The results are presented by two bar charts, one for the BOE and the other for the DOUE. Each chart shows one bar for each type of policy, as shown in Fig. 14.

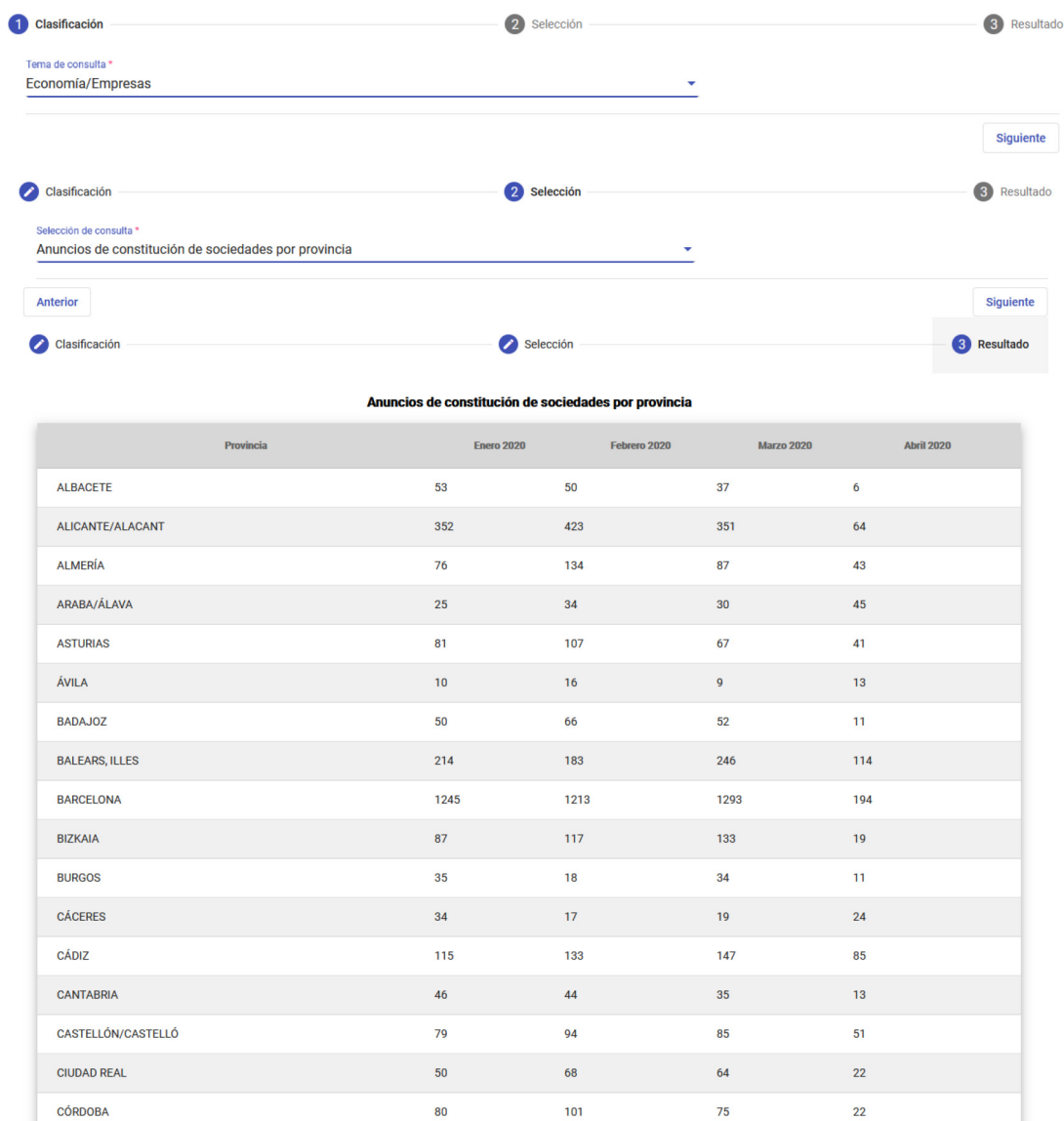


Fig. 12. TA query form and results. Source: Own elaboration.

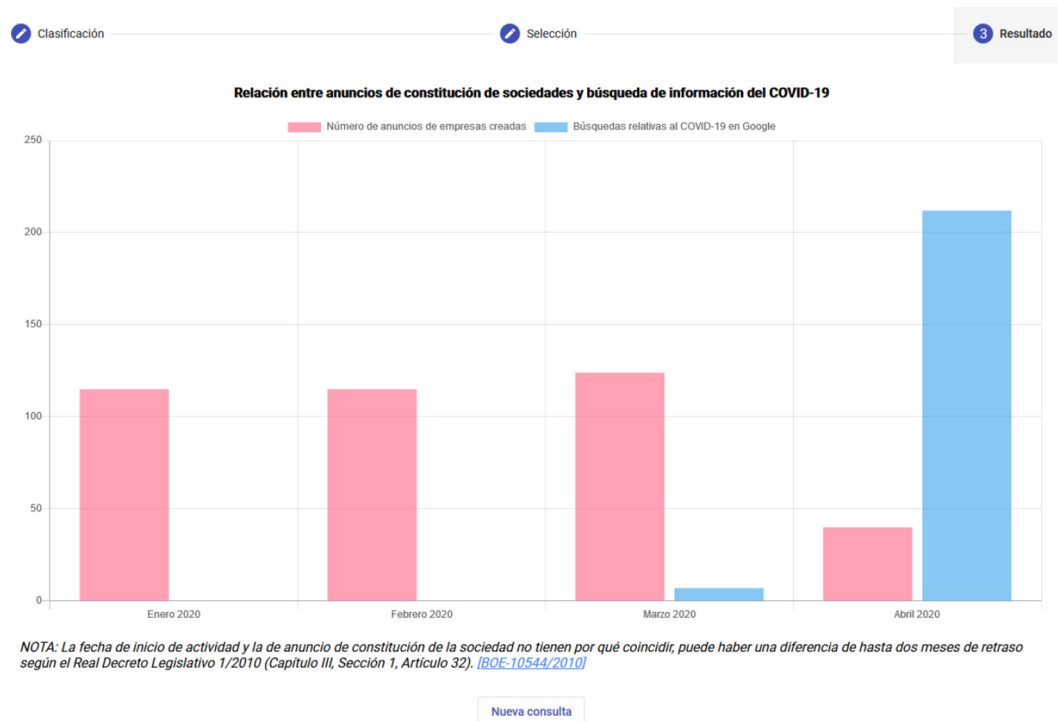


Fig. 13. TB query form and results. Source: Own elaboration.

4.3.4. Topic D) List of pardons

This topic is contained in the Legal classification. Then, the user must select this category to query about this topic. Fig. 15 shows the query form and the result of the query in a table. Fig. 16 shows the result of the query by filtering through the 'Name' field that appears in step 2 of the 'mat-horizontal-stepper' element.

4.3.5. Topic D) Evolution of the development of the amount of legislation

The user must select the Miscellaneous classification to query about this topic. Fig. 17 shows the query form and the result of the query using a bar chart.

4.4. Evaluation

4.4.1. Objective 1) Showing information without blocking it

Search engines, e.g. Google, must remove search results according to the content of the robots.txt file. Then, when using Google to perform OSINT, there will be blocked information. In the case of MedOSINT, since the process of obtaining the bulletins does not consider the restriction of the robot exclusion protocol because it is not a search engine, it does not block information. Therefore, this objective is fulfilled.

4.4.2. Objective 2) Precise search through indexed searches

Indexing is to make an index to order a series of data to facilitate their consultation and analysis. The MedOSINT tool processes the loaded information from bulletins and other sources (as detailed in step 10) and stores it in a series of tables. There is a table for each of the topics. MedOSINT also allows to consult by filtering information. Therefore, this requirement is fulfilled through the processing form described in step 10. The demonstration of this objective is shown in Figs. 16 and 18. In both figures a person's name is consulted to find out if he or she has received a pardon. As shown in Fig. 16, MedOSINT immediately shows that this person has been pardoned thanks to the indexed search. As shown in Fig. 18, the BOE state agency search engine [6], on the contrary, does not show this information directly, and the user has to search through more than 2000 results of the 44 pages.

4.4.3. Objective 3) Using a representation that facilitates the analysis of information

The information should be presented in a format that facilitates consultation and analysis. MedOSINT uses a series of graphs and tables to meet this objective. In addition, the processing method for each topic is explained in a section in the web. It also indicates the bulletins and external sources where the information for each topic has been obtained. This has

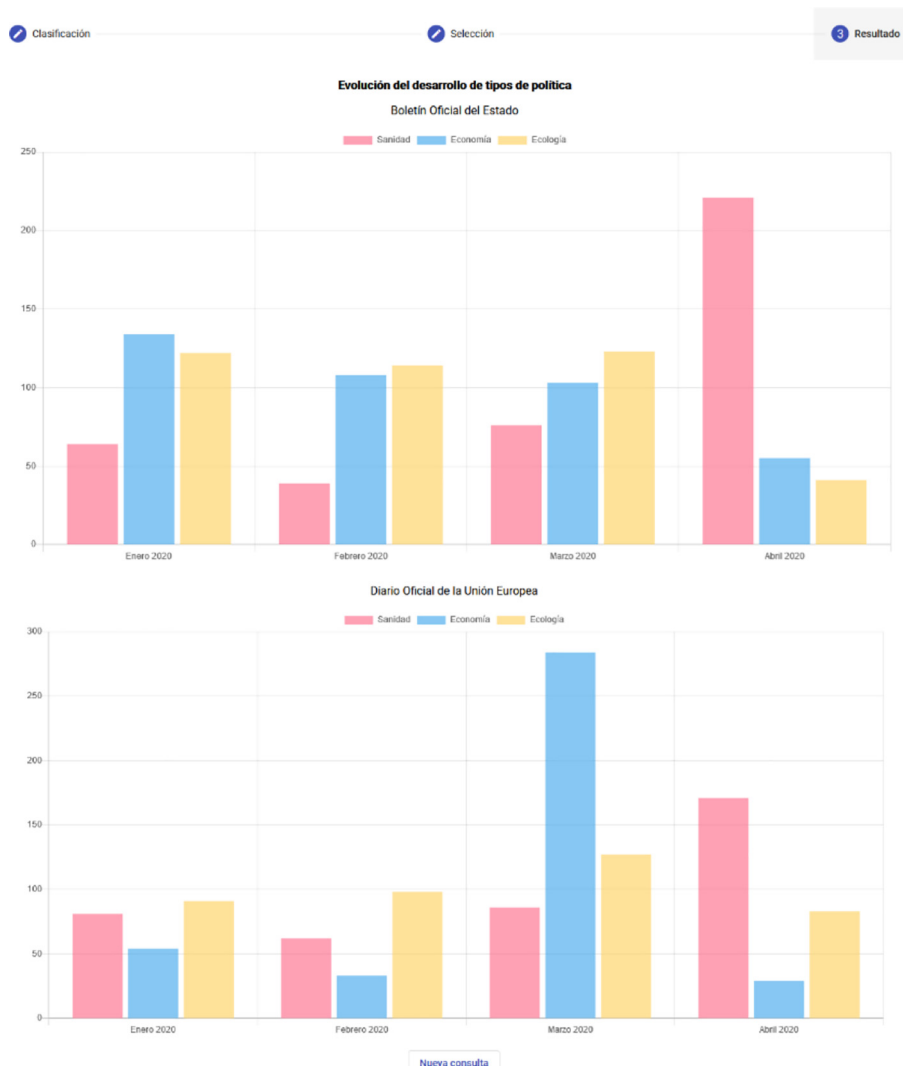


Fig. 14. CT results. Source: Own elaboration.

been shown in previous figures. For example, in Fig. 16 MedOSINT shows the results with a simple table with two columns, one for the date of the pardon and the other for the name of the person pardoned. On the other hand, Fig. 18 shows the results retrieved from the search engine of the state agency of the BOE [6]. These are distributed in 44 pages that do not facilitate the consultation and the analysis of the information.

4.4.4. Objective 4) To be able to use information external to the bulletins to improve the outcome of the analysis

The use of information external to bulletins is an important objective because this provides more complete information for the user. This objective is met by steps 8 and 9. In step 8 it is analyzed if external information is necessary, and in step 9 it is loaded in MedOSINT. Fig. 6 shows how MedOSINT uses external information loaded from Google Trends [19].

4.4.5. Objective 5) Through a CBR explanatory system all news are classified as false (4 types) or true, enabling an explanation

To evaluate the explanation method, we have analyzed the common features between the query – Fake/Good News description- and the header cases, using this metric as an estimation of their quality. The intuition behind this metric is that an explanatory case should be as similar as possible to the query. This way, the explanation given to the user will contain more feature-level justifications as we describe in Fig. 19. It means that the dataset contains two different types of categories. 20,000 are non-fake news from verified official sources as the WHO and the BOE from Spain. In the other side there are 20,000 records obtained from Internet fake news. These records are classified into five different categories (see Table 9), Fake News 1 (FN1) means 1) Fabricated news; Fake News 2 (FN2) means 2) Manipulated news; Fake News 3 (FN3) means 3) Advertisement news; Fake News 4 (FN4) means 4) Irrelevant news; Real News 1 (RN1) means 5) Sufficient News.

1 Clasificación 2 Selección 3 Resultado

Tema de consulta *
Judicial

Siguiente

1 Clasificación 2 Selección 3 Resultado

Selección de consulta *
Listado de indultos

Nombre
Nombre

Anterior Siguiente

Listado de indultos

Fecha	Nombre
11/02/2020	AND [REDACTED]
18/02/2020	GIZE [REDACTED]
13/02/2020	LUIS [REDACTED]
11/02/2020	BAS [REDACTED]
18/02/2020	MAP [REDACTED]

Fig. 15. TD query form and results. Source: Own elaboration.

1 Clasificación 2 Selección 3 Resultado

Selección de consulta *
Listado de indultos

Nombre
GIZE [REDACTED]

Anterior Siguiente

1 Clasificación 2 Selección 3 Resultado

Listado de indultos

Fecha	Nombre
18/02/2020	GIZE [REDACTED]

Nueva consulta

Fig. 16. TD results filtering by name. Source: Own elaboration.

Concretely, we have obtained the three most suitable explanations for every case and analyzed the percentage of common features between them. By using a nearest neighbor retrieval method, we obtain the most similar cases to the given query. Then the adaptation stage combines the most similar explanatory cases to generate the explanation. For example, we found 3,221 fabricated news on the Internet. Header and content of fabricated news are compared with each other to identify similarities. Similarity is a core concept in case-based reasoning (CBR) because case base building, case retrieval, and even case adaptation, all use similarity-based reasoning. The notion of a metric of distance between objects is used in many contexts as a measure of similarity between elements. Similarity and the nearest neighbor algorithm play a major part in CBR. The CBR classifies according to the distance or similarity between a new item and an item in the database. For example, if its distance to a fabricated news item is less than its distance to a manipulated news item, it is classified as a fabricated news item. Firstly, it examines the origin of the news item and, if it is issued by reliable sources such as WHO or the BORME, it directly



Fig. 17. TE query form and results. Source: Own elaboration.

classifies it as true. Note: previously to the CBR stage, the classification is done by the systems analyst with the support of MedOSINT. He/she contrasts this news with official sources such as the WHO. Once the news item is verified, it is classified as fake or true. The CBR provides a more detailed classification of Fake news as shown in Table 9.

Table 11 shows an example of this comparative analysis. For each query, the identifier of the most similar neighbors is provided together with the percentage of common features. As we can observe in this example, the most suitable explanatory case has a higher percentage of common features (81%), corresponding to irrelevant news, the second one also corresponds to irrelevant news. Percentage 0% corresponds to good news and others are in the middle (e.g. 24% is FN1 - fabricated news -). Each case has its own ID and a unique solution. As the database grows, more similarities can be found between the news. The explanatory CBR will find the level of falsehood of a news item based on this percentage.

To understand the global efficiency of the CBR explanatory system, we have computed the average common features between the query and explanatory cases for the whole case base. The average common features for the three nearest explanatory cases are 58% for 3NN (nearest neighbor), 78.88% for 2NN and 97.38% for 1NN in the FN1 cases. This suggests that the reduction of the number of cases from the 40,000 available instances to 1,000 cases has no impact in the performance of the explanation system as the CBR method can still find several explanatory examples. The choice of 1,000 cases for the case base was made based on preliminary experimentations. Future work could analyze the behavior of the system when decreasing that value to reduce the size of the case base and analyze its impact in the performance of the CBR system. To feed the system with the classification reported in Table 9, we used 1,000 cases of which 50% were false news and 50% were true news. The system correctly predicted the falsehood of 93.33% of fake news in Spain.

4.5. Comparison with other tools

4.5.1. Tool 1) Google search engine

Table 2 of the state-of-the-art section shows the following deficiencies found in the Google search engine to carry out OSINT on official bulletins: 1) Lack of information in certain bulletins blocked by the robot exclusion protocol 2) Impossibility to carry out indexed searches 3) Low quality in the representation of the information for analysis 4) Impossibility to add information to this representation to improve the user's analysis. All these points have been improved in MedOSINT. For

The image shows two screenshots of the Agencia Estatal Boletín Oficial del Estado (BOE) search engine. The top screenshot displays the search interface with various filters: Sección (I, II, III, IV, V, TC), Título, Departamento, Texto (with a redacted value), Núm. de boletín, Núm. oficial, Fecha de publicación, and Documentos por página. The bottom screenshot shows the search results for a pardon, listing three entries from the Ministerio del Interior and Ministerio de Justicia, each with a reference to a BOE-A-2020 document and a link to view more details.

Fig. 18. Search for a pardon in the BOE state agency search engine. Source: Own elaboration.

example, MedOSINT does not block information. MedOSINT is focused on intelligence analysis since it facilitates the study of information with an adequate and friendly representation. MedOSINT allows indexed searches and complementary external information. Fig. 20 shows the result of a Google search of a person's name to check if he/she has been pardoned. When comparing this figure with Fig. 16, it is verified that MedOSINT offers the information directly and with a friendly format. Besides, Google may show no results if the web of the state agency of the BOE [6] blocked the information by means of the protocol of exclusion of robots. This would not happen using MedOSINT since it does not use this protocol.

4.5.2. Tool 2) internal search engine of each newsletter

This tool is the least suitable to carry out OSINT on official bulletins. MedOSINT can use any bulletin. Moreover, not all the websites that publish the bulletins have a search engine. The other characteristics described in Table 2 are also improved by MedOSINT. For example, MedOSINT allows to make indexed searches as shown in Figs. 16 and 18 in which the same query is made in MedOSINT and in the internal search engine of the BOE, respectively.

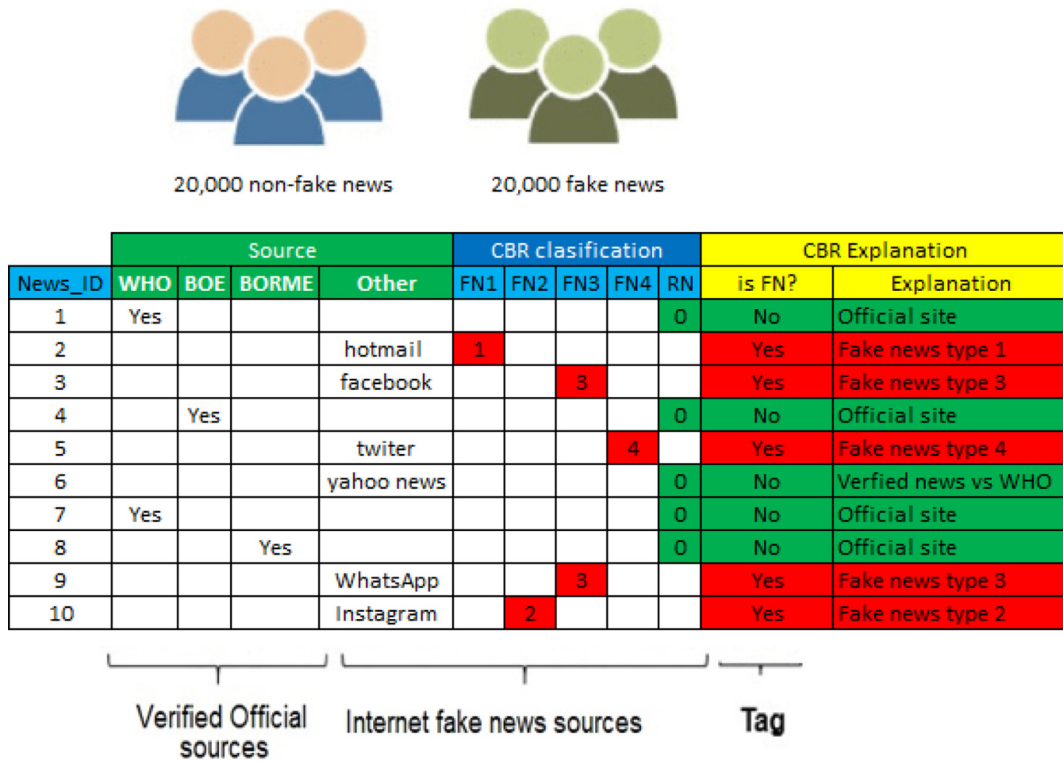


Fig. 19. News database detection for CBR. Source: Own elaboration.

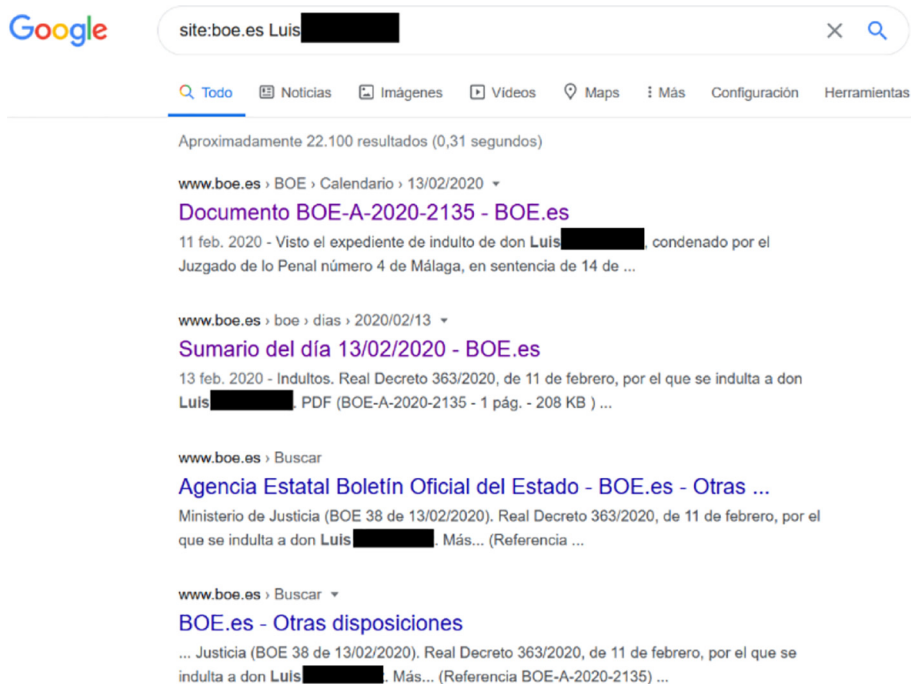


Fig. 20. Google search for a pardon. Source: Own elaboration.

Table 11

Most similar neighbors comparative analysis. Source: own.

Case ID	Notice source	Fake News Type	Fake news quantity	Fake News porcentaje	FN name	Origin Source	Solution
1	WHO	GN	0	0%	Sufficient news	https://www.who.int/es/	S1
2	WhatsApp Web	FN4	325	81%	Irrelevant news	web.whatsapp.com	S2
3	Yahoo news	FN1	58	15%	Fabricated news	https://news.yahoo.com/	S3
4	Facebook	FN4	303	76%	Irrelevant news	https://www.facebook.com/	S4
5	WHO	GN	0	0%	Sufficient news	https://www.who.int/es/	S5
6	BOE	GN	0	0%	Sufficient news	https://www.boe.es/	S6
7	BORME	GN	0	0%	Sufficient news	https://www.boe.es/diario	S7
8	Instagram	FN3	151	38%	Advertisement news	https://www.instagram.com/	S8
9	madridsalud.es	GN	0	0%	Sufficient news	https://www.comunidad.madrid/salud	S9
10	La nación	FN1	78	20%	Fabricated news	https://www.lanacion.com.ar/	S10
11	Clarín	FN2	188	47%	Manipulated news	https://www.clarin.com/salud	S11
12	infobae	FN3	233	58%	Advertisement news	https://www.infobae.com/	S12
13	Ministerio de Sanidad	GN	0	0%	Sufficient news	https://www.mscbs.gob.es/	S13
14	Euskadi.eus	FN1	97	24%	Fabricated news	https://www.euskadi.eus/	S14
15	5 TeleCinco	FN1	5	1%	Fabricated news	https://www.telecinco.es/	S15

Table 12

OSINT tools comparison versus MedOSINT Source: Own elaboration.

Feature	Google search	Internal search engine of each newsletter	www.who.int bulletin	IBOOF	LibreBOR.me	MedOSINT
Locked content	Yes	No	No	No	No	No
Restricted search	No	Depending on each bulletin	No	No	No	No
Indexed search and not just by words	No	No	No	No	Yes	Yes
Display information from any newsletter	Yes	No	No	No	No	Yes
Representation of the information to carry out OSINT	No	No	No	No	Yes	Yes
Supports complementary external information for analysis	No	No	No	No	No	Yes

4.5.3. Tool 3) IBOOF

The deficiencies found in IBOOF to carry out OSINT are: 1) Impossibility to carry out indexed searches 2) Limitation of some bulletins 3) Low quality in the representation of the information for its analysis 4) Impossibility to add information to the representation to improve the analysis. These characteristics have been improved in MedOSINT.

4.5.4. Tool 4) LibreBOR.me

This tool, unlike the 3 previous ones, allows to make indexed searches and represents the information in a way that facilitates the analysis to the user. These two features are also present in MedOSINT. However, LibreBOR.me cannot use complementary external information and cannot work with any bulletin. LibreBOR.me only processes information from the BORME. MedOSINT can work with any newsletter from which information is loaded and with external information.

As a summary, Table 12 compares the tools of Table 2 and MedOSINT.

MedOSINT satisfactorily fulfils all the analyzed characteristics and outperforms the other analyzed tools.

5. Conclusions

MedOSINT has proven to be a reliable and secure tool that contributes to the detection of false news about the coronavirus. MedOSINT also works with other types of news as economic news. Besides it can work with different official bulletins such as those of the European Union and, specifically, Spanish bulletins. The collection and processing of open access information to generate intelligence through its analysis is very relevant to individuals, companies, and institutions. In fact, government agencies have been created exclusively for this purpose. Furthermore, considering that official bulletins are a source of important official information, it is necessary to carry out OSINT on these sources. The hypothesis raised at the beginning of this research was: MedOSINT outperforms IBOOF, LibreBOR.me, Google and the bulletins' search engines in the task of

OSINT in the dimensions of content blocking, restricted search, indexed search, simultaneous search on multiple bulletins, and representation of information, in terms of availability and quality.

MedOSINT is a tool that has two different main parts. The first part covers the acquisition and processing of the information. The second part corresponds to the representation of the processed information. MedOSINT has a technological design allowing to comply with the previous requirements. To perform intelligence, it is necessary to obtain data from official gazettes and MedOSINT classifies them. For each topic, information is obtained from newsletters and other external sources. Once this information is processed, it is presented to the user, usually an intelligence analyst, through a web application. MedOSINT has a modular design that allows, from the input data, to perform multiple processing and representations according to the topic. The system shows the result of the processing and explains the way in which each topic is processed. Sections 3 and 4 show the result of the design and prototyping. In this work, 5 specific topics have been implemented, but it is possible to implement any other topic. It is a tool that, in addition to performing OSINT on official bulletins, has surpassed the specifications of other tools that perform a similar function. It has been demonstrated that MedOSINT does not block information, allows for precise indexed searches, represents the information facilitating its analysis and supports information from external sources. If in the future the entities that publish the newsletters change their presentation format, it will be necessary to change the processing in MedOSINT. The source code of MedOSINT is available for future use and research [30].

The problem of fake news has reached the current European political scene where even a discussion was opened to figure out how to deal with it to prevent its creation and propagation [17]. The CBR classifier demonstrated its ability to learn how to identify fake news and then apply this knowledge to the prognosis of new cases. The system correctly predicted the falsehood of 93.33% of fake news in Spain. Based on the fake news unique identifier, it is possible to monitor the non-authorized news and implement preventive activities that minimize the risk of releasing a false article.

As future work, we would like to explore further ways to combine the explanatory cases and generate the explanations. In this work, we have binarized the features to compare the most similar explanatory cases with the given fake news description. However, we could apply more advanced and knowledge-rich strategies that consider the magnitude of the associated values. Additionally, as the choice of 1,000 cases for the case base was made based on preliminary experimentations, we could analyze the impact in the performance of the CBR system when decreasing that value to reduce the size of the case base. Finally, we would like to conduct a user-based evaluation to validate the generated explanations. As future work we intend to extend this research to medical reports issued by the governments of each country as well as other important bulletins such as the WHO. Besides we are planning a study on the user experience (UX) when using MedOSINT using some combination of existing UX questionnaires [35].

6. Glossary

Abbreviation	Definition
BOE	Spanish Official Gazette
BORME	Spanish Official Gazette of the Commercial Registry
DOUE	Official Journal of the European Union
TA (Topic A)	Spanish Company formation announcements by province
TB (Topic B)	Relationship between company incorporation announcements and COVID-19 information search
TC (Topic C)	Spanish Evolution of policy development
TD (Topic D)	Spanish List of pardons
TE (Topic E)	Spanish Evolution of the development of the amount of legislation
NPM	NODE Package Manager
WHO	The World Health Organization Bulletin
CBR	Case-based Reasoning

CRedit authorship contribution statement

Sergio Mauricio Martínez Monterrubio: Conceptualization, Formal analysis, Supervision, Writing - original draft. **Amaya Noain-Sánchez:** Investigation, Software, Visualization, Writing - original draft. **Elena Verdú Pérez:** Conceptualization, Formal analysis, Methodology, Writing - review & editing. **Rubén González Crespo:** Conceptualization, Validation, Visualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been made possible by the support of SECTEI (*Subsecretaría de Ciencia, Tecnología e Innovación de la Ciudad de México*) for the first author during his postdoctoral studies at the Universidad Complutense de Madrid.

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

References

- [1] A. Aamodt, E. Plaza, Case-based reasoning: foundational issues methodological variations and system approaches, *AI Commun.* 7 (1) (1994) 39–59.
- [2] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, *J. Econ. Perspect.* 31(2). American Economic Association, pp. 211–236, Mar. 01, 2017, doi: 10.1257/jep.31.2.211.
- [3] S. Aral, D. Eckles, Protecting elections from social media manipulation, *Science* 365(6456) (2019).
- [4] A. Bergström, M. Jervelycke Belfrage, News in social media, *Digit. J.* 6(5) (May 2018) 583–598. doi: 10.1080/21670811.2018.1423625.
- [5] Biblioteca Nacional de Francia (BNF). (17 May 2020). <https://catalogue.bnf.fr>. Accessed: 17 May 2020. [Online]. Available: <https://catalogue.bnf.fr/ark:/12148/cb32780021g.public>
- [6] BOE. (3 May 2020). <https://www.boe.es>. Accessed: 3 May 2020. [Online]. Available: https://www.boe.es/datosabiertos/ejemplo_script_borme.php
- [7] E. Bozdag, Bias in algorithmic filtering and personalization, *Ethics Inf. Technol.* 15 (3) (Sep. 2013) 209–227, <https://doi.org/10.1007/s10676-013-9321-6>.
- [8] J.S. Brennen, F. Simon, P.N. Howard, R.K. Nielsen, Types, sources, and claims of COVID-19 misinformation, Reuters Institute for the Study of Journalism, 2020. <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation> (accessed Apr. 22, 2020).
- [9] A. Bruns, Gatekeeping, gatewatching, real-time feedback: new challenges for Journalism, *Brazilian J. Res.* 7 (2) (2011) 117–136, <https://doi.org/10.25200/bjr.v7n2.2011.355>.
- [10] J. Cepeda, Provisional version Social media: social threads or threats to human rights, 2019. Accessed: Apr. 22, 2020. [Online]. Available: <http://assembly.coe.int>.
- [11] A. Chakraborty, B. Paranjape, S. Kakarla, N. Ganguly, Stop Clickbait: detecting and preventing clickbaits in online news media, in: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2016, pp. 9–16, <https://doi.org/10.1109/ASONAM.2016.7752207>.
- [12] M. de Cock Buning, A multi-dimensional approach to disinformation, in: Report of the independent High level Group on fake news and online disinformation, 2018, <https://doi.org/10.2759/739290>.
- [13] N.J. Conroy, V.L. Rubin, Y. Chen, Automatic deception detection: methods for finding fake news, *Proc. Assoc. Inf. Sci. Technol.* 52 (1) (Jan. 2015) 1–4, <https://doi.org/10.1002/pr2.2015.145052010082>.
- [14] M. Dur-e-Ahmad, M. Imran, Transmission dynamics model of coronavirus COVID-19 for the outbreak in most affected countries of the world, *Int. J. Interact. Multimedia Artificial Intell.* 6 (2) (Jun. 2020) 7–10, <https://doi.org/10.9781/ijimai.2020.04.001>.
- [15] Ethical Journalism Network, Tackling misinformation during Covid-19: a journalistic and ethical imperative, 2020. <https://ethicaljournalismnetwork.org/tackling-misinformation-during-covid-19-an-journalistic-and-ethical-imperative> (accessed Apr. 22, 2020).
- [16] S. Fong, G. Li, N. Dey, R.G. Crespo, E. Herrera-Viedma, Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak, *Int. J. Interact. Multimedia Artificial Intell.* 6 (1) (Mar. 2020) 132–140, <https://doi.org/10.9781/ijimai.2020.02.002>.
- [17] Foro Europa Ciudadana. El debate de cómo combatir las fake news en las redes sociales llega al parlamento europeo. April 2018. Accessed: 20 Sept. 2020. [Online]. Available: <https://www.europaciudadana.org/el-debate-sobre-como-combatir-noticias-falsas-en-las-redes-sociales-llega-al-parlamento-europeo/>
- [18] N. Gallant, G. Latzo-Toth, M. Pastinelli, Circulation de l'information sur les médias sociaux pendant la grève étudiante de 2012 au Québec, 2015. Accessed: 22 Apr. 2020. [Online]. Available: <https://www.cem.ulaval.ca/publications/circulation-de-linformation-sur-les-medias-sociaux-pendant-la- greve-etudiante-de-2012-au-quebec/>.
- [19] Google Trends. (3 May 2020). <https://trends.google.es>. Accessed: 3 May 2020. [Online]. Available: <https://trends.google.es>: <https://trends.google.es/trends/explore?date=2020-01-01%202020-04-30&geo=ES&q=COVID-19>
- [20] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election, *Science* 363 (6425) (Jan. 2019) 374–378, <https://doi.org/10.1126/science.aau2706>.
- [21] M. Gupta, R. Jain, S. Taneja, G. Chaudhary, M. Khari, E. Verdú, Real-time measurement of the uncertain epidemiological appearances of COVID-19 infections, *Appl. Soft Comput.* 101 (Mar. 2021) 107039, <https://doi.org/10.1016/j.asoc.2020.107039>.
- [22] C.K. Hiramath, G.C. Deshpande, Fake news detection using deep learning techniques, in: 1st IEEE International Conference on Advances in Information Technology, ICAIT 2019 - Proceedings, Jul. 2019, pp. 411–415, doi: 10.1109/ICAIT47043.2019.8987258.
- [23] F. Jurado, O. Delgado, A. Ortigosa, Tracking news stories using blockchain to guarantee their traceability and information analysis, *Int. J. Interact. Multimedia Artificial Intell.* (2020), <https://doi.org/10.9781/ijimai.2020.06.003>.
- [24] J.Y. Khan, M.T.I. Khondaker, A. Iqbal, S. Afroz, A Benchmark Study on Machine Learning Methods for Fake News Detection, May 2019, Accessed: 22 Apr. 2020. [Online]. Available: <http://arxiv.org/abs/1905.04749>.
- [25] D.M.J. Lazer et al, The science of fake news, *Science* (80) 359 (6380) (2018) 1094–1096, <https://doi.org/10.1126/science.aao2998>.
- [26] D.B. Leake, *Case-Based Reasoning: Experiences Lessons and Future Directions*, MIT Press, Cambridge, MA, USA, 1996.
- [27] J.C. Lopez, V. Lacorzana, ConlaSaludNoseJuega, 18 May 2020. Accessed: 18 August 2020 [Online]. Available: <https://conlasaludnosejuega.org/blog/index.php?las-mascaras-faciales-crean-serios-riesgos-en-nuestra-salud>.
- [28] Mail Online news, 17th June 2020. Accessed: 18 August 2020. [Online]. Available: <https://www.dailymail.co.uk/news/article-8432321/Government-orders-review-vitamin-D-role-Covid-19.html>.
- [29] A. Martínez, (28 May 2014). www.incibe-cert.es. Accessed: 8 Apr. 2020. [Online]. Available: <https://www.incibe-cert.es/blog/osint-la-informacion-es-poder>
- [30] S.M. Martínez Monterrubio, «MedOsint in Google Drive,» 1 March 2021. [Online]. Available: https://drive.google.com/file/d/1RQarOI55EnSMMc_XwxmVnPsPyqsK1jY/view?usp=sharing.
- [31] ODNI News Release No. 6-05. (8 Nov 2005). DNI.gov. Accessed: 10 Apr. 2020. [Online]. Available: DNI.gov: https://web.archive.org/web/20060430212003/http://www.dni.gov/press_releases/20051108_release.htm
- [32] Oracle. (20 Jan. 2015). <https://github.com/>. Accessed: 10 May 2020. [Online]. Available: <https://github.com/>: <https://github.com/oracle/node-oracledb/blob/master/test/README.md>
- [33] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, Viking, 2011.
- [34] A.L. Schmidt et al, Anatomy of news consumption on Facebook, *Proc. Natl. Acad. Sci. U. S. A.* 114 (12) (Mar. 2017) 3035–3039, <https://doi.org/10.1073/pnas.1617052114>.

- [35] M. Schrepp, T. Thomaschewski, Design and validation of a design and validation of a framework for the creation of user experience questionnaires, *Int. J. Interact. Multimedia Artificial Intell.* 5 (7) (2019) 88–95. <https://doi.org/10.9781/ijimai.2019.06.006>.
- [36] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media. A data mining perspective, *ACM SIGKDD Explor. Newsl.* 19 (1) (Sep. 2017) 22–36. <https://doi.org/10.1145/3137597.3137600>.
- [37] C. Silverman, This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, BuzzFeed News, pp. 1–7, 2016, [Online]. Available: <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>.
- [38] Social Observatory for Disinformation and Social Media Analysis, False News Spread More Rapidly Than Coronavirus 2019-nCoV, 2020. <https://www.disinfobservatory.org/false-news-spread-more-rapidly-than-coronavirus-2019-ncov/> (accessed Apr. 22, 2020).
- [39] Social Observatory for Disinformation and Social Media Analysis, "Homepage | SOMA Disinfobservatory," 2020. <https://www.disinfobservatory.org/> (accessed Apr. 22, 2020).
- [40] G.R. Vásquez-Morales, S.M. Martínez-Monterrubio, P. Moreno-Ger, J.A. Recio-García, Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning, *IEEE Access* 7 (2019) 152900–152910. <https://doi.org/10.1109/ACCESS.2019.2948430>.
- [41] S. Vosoughi, M. 'Neo' Mohsenvand, D. Roy, Rumor gauge: predicting the veracity of rumors on twitter, *ACM Trans. Knowl. Discov. Data* 11(4) (Jul. 2017) 1–36. doi: 10.1145/3070644.
- [42] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, *Science* (80) 359(6380) (Mar. 2018) 1146–1151. doi: 10.1126/science.aap9559.
- [43] L.M. Wathen, *Estadística aplicada a los negocios y la economía*. México D.F.: MC GRAW HILL, 2008.
- [44] World Health Organization, Situation reports Coronavirus disease 2019 (COVID-19), 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [45] World Health Organization, Consejos para la población acerca de los rumores sobre el nuevo coronavirus (2019-nCoV), 30 June 2020. Accessed: 22 August 2020. [Online]. Available: <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>.
- [46] World Health Organization, Munich Security Conference, 2020. <https://www.who.int/dg/speeches/detail/munich-security-conference> (accessed Apr. 22, 2020).
- [47] World Health Organization, Myth busters, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters> (accessed Apr. 22, 2020).
- [48] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, P.S. Yu, TI-CNN: convolutional neural networks for fake news detection, *ACM Trans. Knowl. Discov. Data* 11 (4) (Jul. 2018) 1–36. <https://doi.org/10.1145/3070644>.
- [49] D. Yin, S. Mitra, H. Zhang, When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth, *Inf. Syst. Res.* 27 (1) (2016) 131–144. <https://doi.org/10.1287/isre.2015.0617>.



Sergio Mauricio Martínez Monterrubio Received the Ph.D. degree in computer science from the Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) in 2016. The degree in computer science and the master's degree in international business administration from the Universidad Nacional Autónoma de México (UNAM) in 1998 and 2003, respectively. He won the Mexico's First National Prize for his thesis in computer science from ANFECA in 1999. He gains a Doctoral Scholarship from CONACYT in 2009 and a Postdoctoral Scholarship from the SECTEI in 2017. Since 2017 he has been an investigator making his Postdoctoral research with the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM). He is member of the Group of Artificial Intelligence Applications (GAIA). His professional experience includes working in transnational companies as McAfee, Oracle, Entrust Technologies, Colgate Palmolive, ABC Medical Center, and Continental Automotive Systems. His current job is director of the master's degree in computer security at UNIR. He has published research works in several relevant journals. His research is in cyber security and Artificial Intelligence.



Amaya Noain-Sánchez Ph.D. Lecturer at both International University of La Rioja and Rey Juan Carlos University of Madrid. She has been visiting researcher at the Goethe-Universität Frankfurt. Her lines of research include Digital Communication, Social Networks, Privacy vs. Public Sphere, Data Protection, Surveillance, Disinformation and Media Ethic. Her Doctoral Essay obtained the 2015 National Research Award by the Spanish Agency of Data Protection (AEPD). Additionally, she obtained the 2017 Emerald Literati Awards for Excellence for her research on privacy by default and informed consent on the Internet. Apart from her academic experience, she works as a journalist for several mass media.



Elena Verdú Elena Verdú received her master's and Ph.D. degrees in telecommunications engineering from the University of Valladolid, Spain, in 1999 and 2010, respectively. She is currently an Associate Professor at Universidad Internacional de La Rioja (UNIR) and member of the Research Group "Data Driven Science" of UNIR. For more than 15 years, she has worked on research projects at both national and European levels. Her research has focused on e-learning technologies, intelligent tutoring systems, competitive learning systems, accessibility, speech and image processing, data mining and expert systems.



Rubén González Crespo Dr. Rubén González Crespo has received a PhD in Computer Science Engineering. He is also Industrial Engineering and MSc in Computer Science and Project Management. Currently he is Vice Rector of Academic Affairs from UNIR. He is an advisory board member for the Ministry of Education at Colombia and evaluator from the National Agency for Quality Evaluation and Accreditation of Spain (ANECA). He is president from Engineering and Architecture Commission in Fundación Madrid+D. He is patron of the Spain Free Software Foundation. He is member of different committees at ISO Organization. He has published more than 200 research works in several relevant journals. He also manages different research projects in IoT, TEL, and Artificial Intelligence.