

Article



# An AI-Enabled Framework for Real-Time Generation of News Articles Based on Big EO Data for Disaster Reporting

Maria Tsourma \*<sup>D</sup>, Alexandros Zamichos, Efthymios Efthymiadis \*, Anastasios Drosou and Dimitrios Tzovaras

Centre for Research and Technology Hellas, Information Technologies Institute, 57001 Thessaloniki, Greece; zamihos@iti.gr (A.Z.); drosou@iti.gr (A.D.); dimitrios.tzovaras@iti.gr (D.T.) \* Correspondence: mtsourma@iti.gr (M.T.); efthimios@iti.gr (E.E.)

Abstract: In the field of journalism, the collection and processing of information from different heterogeneous sources are difficult and time-consuming processes. In the context of the theory of journalism 3.0, where multimedia data can be extracted from different sources on the web, the possibility of creating a tool for the exploitation of Earth observation (EO) data, especially images by professionals belonging to the field of journalism, is explored. With the production of massive volumes of EO image data, the problem of their exploitation and dissemination to the public, specifically, by professionals in the media industry, arises. In particular, the exploitation of satellite image data from existing tools is difficult for professionals who are not familiar with image processing. In this scope, this article presents a new innovative platform that automates some of the journalistic practices. This platform includes several mechanisms allowing users to early detect and receive information about breaking news in real-time, retrieve EO Sentinel-2 images upon request for a certain event, and automatically generate a personalized article according to the writing style of the author. Through this platform, the journalists or editors can also make any modifications to the generated article before publishing. This platform is an added-value tool not only for journalists and the media industry but also for freelancers and article writers who use information extracted from EO data in their articles.

**Keywords:** Web 3.0; article composition; Earth observation (EO); journalism 3.0; media industry; journalistic workflow; journalistic practices; text generation with artificial intelligence (AI); disaster reporting; EarthPress

## 1. Introduction

With today's use of the web, it is possible to access a huge amount of multi-modal information, such as documents, images, videos, etc. This information is generated from different sources and is often difficult to collect. In this new forming environment, while making it possible for journalists to access this huge volume of information, it is often difficult to collect, correlate, and exploit this heterogeneous data, and therefore, it is difficult for journalists to have a complete overview of an issue of a complex topic. This happens because the majority of the information available on the internet is published in an unstructured way. The collection and correlation of unstructured information is a demanding and time-consuming process. Therefore, there is a need for the creation of tools that will collect data from different sources and convert them into structured data with the use of new technologies, allowing all users to have access to them. These new technologies are shaping a new advanced web form called Web 3.0 or the semantic web.

Web 3.0 can be viewed as an extension of the current form of the web. An interpretation of Web 3.0 is that it is an internet service that contains advanced technological features that leverage machine-to-machine interaction but also human-to-human cooperation [1,2]. Web 3.0 entails multiple key elements, such as the semantic web and artificial intelligence (AI) that can be leveraged in order to create added-value web-based applications. Within



Citation: Tsourma, M.; Zamichos, A.; Efthymiadis, E.; Drosou, A.; Tzovaras, D. An AI-Enabled Framework for Real-Time Generation of News Articles Based on Big EO Data for Disaster Reporting. *Future Internet* 2021, *13*, 161. https://doi.org/ 10.3390/fi13060161

Academic Editor: Gyu Myoung Lee

Received: 12 May 2021 Accepted: 17 June 2021 Published: 19 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). this context, web-based applications that combine AI and Big Data have gained great value in recent years and made possible the retrieval of unstructured data from different sources, as well as the analysis and correlation of the extracted information, in order to use them in decision-making or for the automation of time-consuming processes. In addition, the combination of these technologies has led to the analysis of demanding information with great value, such as Earth observation (EO) data.

EO offers a wide and timely covering solution via the fleet of space-borne sensors orbiting the earth at any time and the Big Data piling up at main storage and satellite facilities and hubs of ESA and NASA. The EO data produced are becoming tremendous in volume and quality, and thus, their retrieval, processing, and information extraction is time-consuming and demanding in terms of skills and resources. Therefore, AI methods and cloud technologies have been leveraged for the processing and data extraction of EO data. Once these data are processed, they can offer information of great value that can be leveraged from any end-user without restrictions or need for special knowledge.

One of the areas of interest, where the extracted information of EO data along with the web-based applications of Web 3.0 could be combined, is the media industry and, more specifically, journalism. Corresponding tools of advanced technologies developed in the context of Web 3.0 can change the way journalists collect, process, and eventually interpret information, helping the transition to Journalism 3.0. Journalism 3.0 can be seen as the direction of using advanced technologies by journalists to automate journalistic practices and workflows. Web-based applications using advanced technologies in this context, can be exploited by journalists and automate or semi-automate some of the journalistic practices and workflows followed. In the context of Journalism 3.0, new added-value technologies are applied in the media industry, such as AI, allowing journalists to use the results of scattered information analysis and fusion in order to create news articles promptly for several topics. In particular, new deep learning and machine learning methods offer new capabilities for a variety of automatic tasks, such as automated content production and others, such as data mining, news dissemination, and content optimization [3]. In general, the use of EO data and images, which contain valuable information, could benefit journalists through the provision of additional information extracted from the EO data processing. Such information might concern the percentage of an affected area from a disaster (e.g., flood, fire, earthquake, etc.).

This paper presents the concept of a new innovative web-based platform, called EarthPress. The EarthPress platform aims to deliver added-value products to editors and journalists, allowing them to enrich the content of their publications. EarthPress leverages EO data to generate automatically personalized news articles using also the information extracted from the analysis of EO data. This platform uses AI methods for the aggregation of data from different sources, such as texts from news articles of different websites and posts on social media, and using their analysis and utilization as input for the automatic generation and composition of a news article. The utilization of EO data within the EarthPress platform gives great value to the information that will be provided in the generated text since journalists and users that are interested in the EO data currently cannot easily have access to them or analyze them, because EO data analysis is a demanding and time-consuming procedure.

From the analysis of the collected data, the most trending topics are extracted and provided to the end-users, allowing them to be selected. For each topic, any available information (e.g., EO images, images from social media, social media posts, news articles, etc.) referring to this event are also retrieved and made available to the end-users. With this information, the user can select the ones that s/he wants to use and generate a personalized article through the EarthPress platform. For the personalization of the generated article, the writing style of the author is extracted using AI methods and is transferred to the generated text using previously written articles of the user as a basis. The personalized article that is provided as output of the EarthPress platform is trimmed and finalized by the journalist within the given layout constraints.

The platform includes AI methods for not only the collection and analysis of the data retrieved and the generation of the final article but also for the evaluation of the quality of the retrieved data. Currently, journalists have access to a pile of information for the generation of an article; however, they cannot easily discriminate which of this information is valid and which sources are trustworthy enough. Therefore, in order to avoid spreading misinformation, one of the key aspects of this platform is to provide qualitative data to journalists from verified sources and also to use this verified information for the synthesis of the final news article. Misleading content and fake news included in news articles and posts from social media are eliminated. High quantity journalistic articles and posts from well-known journalist's profiles on social media are collected and promoted. Non-trustworthy sources are excluded from the list of data sources in order to minimize the risk of providing misleading information. The quality of information is the most important feature of the platform as we aim to implement a solution that will have ethical features and will not promote the production of false news or its reproduction. The news article generated by the EarthPress platform will be also evaluated within this scope.

The focus of EarthPress is disaster reporting (e.g., floods, fires, drought, pest, earthquake, erosion/ sedimentation, avalanche), as journalists are challenged in such cases to find the area of interest, collect and analyze any available information in minimal time, and write a detailed article reporting the disaster's status or outcome. The main scope of this platform is to semi-automate the journalistic workflows and provide journalists with the resources and opportunity to create news articles by having a vast amount of qualitative information available. It will act as an intelligent assistant for the journalist, offering enhanced possibilities for content-rich reports. AI techniques make unsupervised querying through Big Data piles feasible. Moreover, EO data resources and geospatial data will become accessible for exploitation by domain-unexperienced users. Tailor-made solutions will be promoted. Hence, the editor's or journalist's profile will be consulted to allow for a high-level specification of learning AI tasks and reporting.

#### 1.1. Related Work

An important parameter related to the functionality of the platform includes the analysis of EO data and, specifically, EO images. A variety of such platforms exist [4,5], most of which offer tools for the processing and analysis of data from satellites. The platform will offer the feature for automatic processing of EO images in real time.

Global Disaster Alert and Coordination System (GDACS) [6] provides a disaster manager and disaster information systems worldwide and aims to fill information and coordination gaps in the aftermath of major disasters in the first phase. The platform is a cooperative framework of the United Nations and the European Commission. The GDACS provides real-time access to web-based disaster information systems and related coordination tools.

Emergency and Disaster Information Service (EDIS) [7] is another platform that aims to monitor and document all the events on Earth that may cause disaster or emergency. The platform is operated by the National Association of Radio Distress-Signaling and Infocommunications, and it monitors and processes several foreign organizations' data to get quick and certified information.

GLIDE [8] is another platform that provides extensive access to disaster information in one place, as well as a globally common unique ID code for disasters. Documents and data pertaining to specific events can be retrieved from various sources or are linked together using the unique GLIDE numbers. In addition to a list of disaster events, GLIDE can also generate summaries of disasters, e.g., by type, country, or year and also provides export of data sets in charts, tabular reports, or statistics.

Different methods for the processing of EO images were used, according to literature, and particular methods were applied for building footprint extraction from EO images. For this reason, architectures for image segmentation [9–11] have been reviewed. Other

methods of processing of EO images used for the implementation of the platform concern the detection of the affected areas from floods and water in general [12,13] and fire [14,15].

One of the important features supported by the platform concerns the personalization of the generated text. For transferring users' writing style according to literature, methods to extract the user's sentiment [16,17] have been implemented.

In literature, the task of text generation includes a vast number of methods. A summary of multiple texts that maintains the information of these texts can be used for text synthesis. State-of-the-art approaches in the field of summarization, in general, and multi-document summarization, in particular, are all based on the Transformer [18], a model that leverages the mechanism of attention and achieves avoidance of utilizing recurrency and convolution [19–22]. The state-of-the-art PEGASUS model [23] is a large Transformer-based encoder-decoder model pre-trained on massive text corpora with a self-supervised objective and is designed for use in both single and multi-document summarization cases. Notably, some other, more general purpose state-of-the-art NLP models, such as BART [24] and T5 [25], can produce results comparable with those of PEGASUS, in a few-shot and zero-shot multi-document summarization settings, suggesting that unlike single-document summarization, highly abstractive multi-document summarization remains a challenge [23].

Data-to-text generation (D2T) is a subtask for the generation of text using piles of data as input. Data-to-text generation is a promising subtask of study with many applications in media, paper generation, storytelling, etc. In their work, Harkous, H et al. [26] present the DATATUNER, a neural end-to-end data-to-text generation system that makes minimal assumptions about the data representation and target domain. DATATUNER achieves state-of-the-art results in various datasets. Another work presented by Kanerva, J et al. [27] aims to generate news articles about Finnish sports news using structured templates/tables of data and pointer-generation network. In the work of Rebuffel, C et al. [28], a hierarchical encoder–decoder model is proposed for transcribing structured data into natural language descriptions. In other papers, like the one presented by Mihir Kale [29], they use pre-trained models, such as T5 [25] and BART [24], for data-to-text generation, achieving state-of-the-art results in various datasets.

## 1.2. Motivation and Contribution

In the era of Big Data, where a vast amount of data is produced on a daily basis, there is the necessity for the implementation of new tools or platforms that are able to facilitate the selective collection, filtering, and processing of this information so useful information can be extracted and used in different fields of the market. One of these fields is the media industry. Journalism is closely related to the retrieval, evaluation, and filtering of raw information, as well as, to the combination of knowledge retrieved from multiple sources and the publishing of news stories that are based on these data. Thus, Web 3.0 applications that are able to facilitate the above journalistic practices are of great significance and comply with the transition towards Journalism 3.0.

In this work, we present EarthPress, an interactive web-platform whose scope is to facilitate journalists and editors in synthesizing news articles about disasters. The platform combines various methods from different scientific fields, such as information retrieval and artificial intelligence and, more specifically, natural language processing and image processing. EarthPress services allow its users to have access to multimedia data (e.g., documents, videos, images, etc.) from multiple sources regarding disasters such as floods and fires. Moreover, the EarthPress platform detects automatically, and in real-time, breaking news related to disasters. The latter is achieved by monitoring publications in social media. The detected breaking news is presented to journalists through the platform's interactive user interface (UI). Additionally, the platform consists of models that are able to retrieve, process, and extract useful information and statistics from EO data, which can be used by editors and journalists in order to enrich the content of their publications. Finally,

the platform provides services that can combine all the above and generate automatic ready-to-print news articles, considering also the writing style of the journalist.

In summary, the platform can facilitate the following journalistic practices:

- Data collection: raw data need to be available (search for data on the web).
- Data filtering: the process of filtering relevant information from the news story.
- Data visualization: the process of transforming data and creating visualizations to help readers understand the meaning behind the data.
- Story generation (publishing): process of creating the story and attaching data and visuals to the story.

## 2. Materials and Methods

This section presents the architecture of the platform, along with a description of the modules and sub-modules included in it. Initially, and before the presentation of the plat-form's architecture, the method for collecting user requirements and specifications in order to define the functionalities of the platform is presented.

# 2.1. Collection of User Requirements

In software development, an important step before the implementation of a platform is the collection and analysis of user requirements. Thus, before the implementation of the EarthPress platform and the definition of its architecture, an online international workshop was conducted in March of 2021, aiming to determine the needs of journalists, current practices followed, and attitudes towards the EarthPress platform. The target audience, 134 participants, of the workshop was both journalists and academics. Many of the participants had experience in using EO data and in using other platforms for article generation. In the workshop, the use and importance of EO data were presented, along with the concept of the EarthPress platform. For this workshop, a questionnaire including both closed and open-ended questions aiming to identify the opinion, needs, and current practices followed by the participants when writing an article was designed. The questionnaire was available online for one week, and the link was sent to all participants at the end of the workshop. The results of the collected data analysis are presented in this sub-section.

The designed questionnaire included two parts. In the first part, demographic information of the users, such as their gender, age, background, etc., was requested (Table 1). Additionally, one of the questions in the first part asked the users to answer if they previously used or are currently using information from EO data in their articles and if they have previously used other platforms for article generation. This was general information requested in order to have a better view of the sample used.

Question	Answer-Distribution				
Gender	Male (19.4%), Female (79.1%), Prefer not to say (1.5%)				
Age	19–21(82.6%), 22–51(17.4%)				
Education	Students (95.5%), Other (4.5%)				
Experience in using EO data	Yes (26.1%), No (73.9%)				
Experience in using article generation platforms	Yes (7.5%), No (92.5%)				

Table 1. Demographic information of the Participants.

The second part of the questionnaire concerned the retrieval of information about the type of data and current practices used during the generation of an article and the flow that is followed for the retrieval and analysis of EO data. In this part, the participants highlighted the main problems or issues currently existing in the field of journalism related with the scope of the platform. The issues presented by the majority of the participants concerned the fact that they did not have the means to collect adequate information for a certain topic and that the collection of all information for the generation of an article was a

time-consuming task. As a second issue, they mentioned the restricted time frame they had for writing an article. Additionally, most participants responded that it was difficult to collect EO images related to a specific area of interest where the event had occurred and analyze them in order to extract any data included. Regarding the usefulness of the EO data, they mentioned that they would be interested in using such information when creating an article; however, this information should be processed. Moreover, another answer provided by some of the participants highlighted the existence of fake news among the news stories published by other journalists for an event as a problem.

Additionally, the participants mentioned that it was crucial for the journalists to receive this information in a timely fashion and have access to all available multimedia and data (e.g., videos, analytics, local news posted, etc.) on a specific topic in order to use them for the article composition. In this context, the EarthPress platform will provide images and local news retrieved for a certain event/disaster to the end-users, allowing them to select the information that is taken into consideration for the text generation.

From the open-ended questions, the suggestions proposed mentioned that the platform would be a useful tool for journalists and other professionals and that by using this platform, journalists would have a chance to be more precise, avoid fake news, and provide qualitative news articles to the public. The elimination of misleading information and data sources was mentioned in more than one answer provided, making the importance of this issue a high priority for the EarthPress platform.

The requirements and specifications extracted from this survey were valuable for the definition of the EarthPress platform's architecture and functionality. Initially, the fact that users presented the text generation process as a time-consuming task and that EO data were not easily retrievable on time was very important, since these two major issues are the basis of the platform's scope. The platform strives to solve the aforementioned issues through the early detection of trending issues, the presentation of qualitative data to the user, and the direct extraction of information produced by the analysis of EO data. Furthermore, it is important that all information that is currently useful for the journalists during the composition of an article, such as posts from social media platforms and articles from websites, is available to the users through a platform in order to be easily accessible at all times.

Considering the generation of an article and based on the fact that most of the participants had previous experience with such platforms, the requirement extracted concerns about the ability of a user to edit an article and receive qualitative information extracted by the analysis of EO data on time in order to use them for the information extraction. Regarding the multimedia used for the final article, images, and statistics, along with video, were highly rated options by the participants. Therefore, the platform should be able not only to retrieve this information when it is available but also to provide it to the users and allow them to select which information will be used for the final article.

According to the analysis of the data provided in this sub-section, the architecture of the EarthPress platform was enhanced in order to include not only the basic modules but also to include the corresponding modules and meet the requirements extracted.

#### 2.2. System Architecture

The system's architecture is presented in Figure 1. The proposed architecture consists of eight main modules, which are the: (a) user interface, (b) user manager, (c) breaking news detector, (d) data fusion (e) database manager (f) image processing, (g) journalist profile extractor, and (h) EarthBot module. The architecture of the platform includes two parts, the front-end, and the back-end. The front-end concerns the design of the platform's user interface (UI), while the back-end includes the platform's database, the implemented AI methods for image processing and text synthesis, as well as, algorithms for data collection and data filtering. A detailed description for each sub-module is provided in the following sub-sections, while a description of the user interface is available in Section 3.



Figure 1. Platform's architecture.

#### 2.3. User Interface

The user interface module refers to the interactive interface of the platform, which allows users to interact with the other modules of the system, as well as visualize the results of the user's searches and the generated article. The purpose of this module is to allow the smooth interaction of the users with the system. Through this interface, users will be able to interact with the system in a user-friendly way without any restrictions or requirements.

## 2.4. User Manager

The main scope of this module is the registration of new users to the EarthPress platform, as well as the logging-in of the already registered users. It is responsible for (a) the registration of new users to the platform, (b) the authentication process of users when they log in to the platform, and (c) the updating of the users' profile. The corresponding APIs are included in this module, aiming at the exchange of all data required among the related modules.

#### 2.5. Database Manager

This module is a data storage mechanism for storing all the available information from all modules of the platform. The types of data stored vary according to the needs of each module. The input data for this module are outputs from other modules, such as tweets, articles, and any additional information provided by the user via the user interface, such as any previously written documents by the user that will be used for the personalization of the generated text.

## 2.6. Breaking News Detector

The instant detection of breaking news, as well as the collection of data related to the breaking news, is of great significance for journalists. In this scope, a module that detects breaking news and collects relevant news data is included within the platform's architecture. Within the EarthPress platform, the important news that we are trying to detect is related to either natural or human-made disasters. To detect breaking news related to disasters, the platform uses the Twitter social network.

The breaking news detector is composed of several sub-modules that perform different functionalities. Two of the sub-modules are related to the data retrieval from Twitter. At first, the trending topic-monitoring sub-module downloads tweets that contain certain, pre-defined, keywords related to disasters, and second, the topic (disaster) classification sub-module monitors the trending topics of certain areas and recognizes trending topics related to disasters.

From the tweets that have been extracted and are related to a natural disaster, it is important to extract information such as the type of disaster, the location, and the date of the event. For this reason, the keywords and main entities extraction sub-module is used to extract the most important words and phrases from the posts retrieved, along with the aforementioned information. For this task, the YAKE algorithm has been used [30]. These keywords are used as input by the article retrieval sub-module, whose aim is to collect news articles that will contain these certain keywords.

Finally, to ensure that the output of the breaking news detector does not contain articles and tweets with false/misleading content, the collected information is checked for their reliability and the detection of fake news. It is essential to ensure that the output information of the breaking news detector does not contain articles and tweets with false/misleading content, as these data will be used to compose the final news article in the EarthBot module. For this reason, deep learning methods have been developed that automatically detect and filter out any text with misleading content to ensure the quality of the output information.

## 2.7. Data Fusion

As mentioned previously, the information that will be used for the generation of the final article comes from different sources. This information needs to be correlated and stored in a grouped way in the database so that it can be used more easily when requested. To store the information in a grouped way so that it refers to common events for every user in the database, a mechanism for comparing and correlating the different information has been implemented.

#### 2.8. EO Image Processing

The scope of this module is to analyze EO images and extract any available information included. The EO images are retrieved using a sentinel-2 parser responsible for downloading sentinel-2 images from the Copernicus Open Access Hub, according to an area of interest and a timestamp. The users select these input parameters when they select a topic of interest from the user interface presented in Section 3. This image is further processed for the detection of affected areas and changes in buildings that have occurred from a disaster. Two distinct sub-modules with different scopes have been included in the EO image processing module: the first one, called the "building footprint extraction and damage detection sub-module" aims to detect damages within an urban area using RGB images and is based on the buildings that exist within a certain area. The second one. Called the "affected area detection" sub-module, is independent of the type of area and can detect floods and burned areas based on sentinel-2 images.

The outcomes of the processing are the processed EO images retrieved, accompanied by rough statistics on the impacted area (e.g., surface, type of land cover/land use affected). The numerical information, provided as output, is taken into consideration during the generation of the news article, giving an added value to the generated news article. On the other hand, the processed EO images are made available to the end-user through the User Interface, allowing them to select if they would like these images to be presented in the final article or not.

The sub-modules included in this module are:

#### 2.8.1. Sentinel-2 Parser

The sentinel-2 parser has been implemented for the retrieval of EO images. This parser receives input coordinates showing the location of the event selected by the user and a timestamp. A before and an after depiction of the location provided are retrieved from the Copernicus Open Access Hub in order to be processed through the sub-modules included in the image processing module.

This sub-module is not only responsible for the retrieval of the sentinel-2 images but also for their conversion to RGB images in order to be used as input in the building footprint extraction and damage detection sub-module. It should be mentioned that the sentinel-2 data that will be retrieved will include 13 spectral bands: four bands at 10 m, six

## 2.8.2. Building Footprint Extraction and Damage Detection Sub-Module

bands at 20 m, and three bands at 60 m spatial resolution.

This sub-module implements a key EO data processing function by providing the capability to detect buildings in satellite imagery and the capability to detect changes in the footprint of the building(s) given images from a time point preceding and a time point following a (disastrous) event (e.g., an earthquake).

The building footprint extraction and damage detection sub-module uses RGB image(s) as input and provides as output a PNG image file of the segmentation map depicting the two classes, "building" and "not building", along with a text file presenting the building footprint area in square meters. Both sets of data will be used for the synthesis of the final article.

#### 2.8.3. Affected Area Detection

The sub-module of affected area detection aims to detect the affected areas using the images retrieved from the sentinel-2 parser as input. Using two different methods, this sub-module can detect both the areas affected by water, through the water change detector and the burned areas, through the burned area detector. The water change detector processes images from the sentinel-2 and outputs processed images and numerical data relating to flood disasters. Similarly, the burned area detector also processes images from the sentinel-2 sub-module and extracts processed images and numerical data relating to fire disasters.

Both modules provide the processed images as output, depicting the affected areas with the use of additional layers over the initial image and numerical data including the percentage of the affected area. The images and the numerical data resulting from this analysis are used for the composition of the final article. The images can be included in the final article, while the numerical data will be used within the context of the generated text, aiming to provide additional information to the user, which in other cases might not be feasible to be retrieved.

## 2.9. Journalist Profile Extractor

The scope of the EarthPress platform is not only to generate articles that include EO data but also to provide personalized articles that follow the writing style of each author. Each author has a unique way of writing and presenting news. In order to personalize the generated text and, in particular, to transfer the style of the writer to the generated text, a module that creates the user's profile from their previously published articles has been designed. This module is intended to define the user's profile by extracting the words/phrases that a journalist uses more frequently on his/her articles, along with the sentiment that the journalist usually uses when composing an article on a certain subject. This module consists of the following submodules:

## 2.9.1. Journalist Specific Keyword Extractor

As mentioned previously, the aim of the journalist specific keyword extractor is to extract the most significant and more frequently used keywords from the documents provided by the user when creating their profile. The most frequent words that a user tends to use in the document are extracted using the keywords extraction method, similar to those used in Section 2.6. These words, named keywords in this context, are an important asset for the article generation as with the words used by the user in an article, the dynamics of the text are alternated.

#### 2.9.2. Sentiment Analysis

The sentiment is an important asset for writing an article. Journalists, and authors in general, tend to write an article having a certain sentiment according to the event's outcome. The tone of each article can be retrieved by analysis of a document, resulting in the extraction of the document's sentiment. This characteristic is important for the personalization of the article through text style transfer since the tone of the article will be adjusted to the sentiment generally used by the author in similar articles.

Using the database manager, the documents provided by the user as input during the user's profile creation are retrieved from the database. These documents are further analyzed in the sentiment analysis in order to extract the sentiment used by the user in the articles. For this purpose, the BERT [16] model has been developed and integrated within the platform. For each journalist, his/her previously published articles are considered. The sentiment that is more frequently used by a journalist for describing a news story related to a certain type of disaster will be finally extracted and passed to the EarthBot module for the final article creation process.

## 2.10. EarthBot

The EarthBot module is the last module of the EarthPress architecture and is responsible for the automatic generation of news articles. The EarthBot is responsible both for the text synthesis of the article as well as the final article composition that integrates both the generated text and multimedia content selected by the user. In order to achieve that, EarthBot integrates different AI technologies, such as multi-document summarization, name entity recognition and keywords extraction, and data-to-text generation, for the generation of the textual part of the article.

The input data of EarthBot are posts from social media and news articles relevant to a disaster, along with statistics that are extracted from the processing of EO data by the EO image-processing module. The social media posts and the news articles are processed, so the most important keywords and main entities can be extracted from them. Additionally, these posts are inserted into the multi-document summarization method supported in order to have a common summary of all documents as output. These two types of data are passed as input to the news article text synthesis sub-module, which is responsible for generating the text content of the final article. This sub-module supports two options that are further described in sub-Section 2.10.3.

Moreover, the news article generation sub-module receives the journalist's profile as its input for the personalization of the generated text. The composition of the final article is achieved by combining the produced text and images selected by the user by the article composition sub-module. In order to achieve all the aforementioned tasks, this module consists of the following sub-modules:

#### 2.10.1. Additional Keywords Generation

The terms keyword and key phrases are used interchangeably in this context. Apart from the extraction of the keywords that are present in a text (i.e., in posts and documents retrieved), additional, absent keywords that are not present in the initial text can potentially be generated using deep learning sequence-to-sequence models. These additional keywords are used as input for the text synthesis sub-module. For the implementation of the additional keywords generation, the method described in [22] is used.

#### 2.10.2. Multi-Document Text Summarization

Multi-document summarization is the task of producing summaries from collections of thematically related documents, as opposed to the single-document summarization that generates the summary from a single document. In the scope of the EarthPress platform, for the text generation task, a multi-document summarization method is used initially to generate summaries of groups of thematically related articles or tweets. For the implementation of multi-document summarization, the PEGASUS [24] model has been

developed and integrated within the platform. The produced summary will be further provided as input for the text generation.

# 2.10.3. News Article Text Synthesis

This sub-module implements the task of generating the textual part of the news article. The generated article describes the event that the user chooses from the available topics resulting from the breaking news detector. The generated text should be consistent, coherent, and syntactically correct. Once the text is generated, the user can edit it and make changes or additions according to his/her preferences.

The user has the opportunity to choose one of the two options supported for the generation of the news article. For the first option, all available information and data extracted from the EO data processing are mapped to predefined text templates. For this reason, a rule-based system will be used to translate the data to map them to the templates. For the implementation of these methods, the method described in [31] is used.

For the second option, two deep learning approaches are followed. For the first approach, a deep neural network is deployed. This method uses data keywords/keyphrases and data extracted from the EO image processing module and the journalist profile extractor as input. The processing of the data is based on a tokenizer that encodes the data into useful representations for computational purposes. Finally, the processed data are used as input to the model used for the text generation. For this approach, the T5 model [25] is used.

For the second approach, a deep neural network is implemented that uses a summary from different textual data as it was produced by the multi-document summarization sub-module. Subsequently, the output summary is aggregated with information from the EO image-processing module and the journalist profile extractor and forms the final news article.

A key difference between the two approaches lies in the different nature of the input data for the deep learning models. In the first approach, keywords/key-phrases and other non-textual data extracted from tweets and articles will be used as input data. In the second method, the tweets and articles themselves will be used for input data without an intermediate stage of extracting information from them. The platform will support both approaches.

## 2.10.4. Article Composition

This submodule is responsible for the composition of all the individual elements that make up the final article. Data from different modalities (text, images) should correspond to each other (e.g., the content of the first paragraph of the text should correspond with the following image, captions should correspond with the images). If the user chooses the first method to produce the text, as described in Section 2.10.3, then the texts and the corresponding images will be placed in specific predefined positions.

If the user chooses the second method for the final news article generation, a heuristic algorithm is used to correlate data from different modalities and synthesize a wellpresented article. The end-users will be able to edit the final article (e.g., rearrange the text and the image places) according to their own needs and preferences. The users will be able to make changes in the content of the generated text. The input data for this submodule are the generated news article, processed EO images resulted from the image processing module, links, hashtags, and image captions.

#### 3. System Presentation

The EarthPress platform is a web-based application for Web 3.0, allowing users to receive information about events occurring in a certain location, retrieve analyzed EO data, and generate a news article by selecting one of the two options supported for the text generation. This platform provides a user-friendly interface through which the users can

interact with the EarthPress platform and its functionalities. In particular, a user through the user interface of the EarthPress platform has the following abilities:

- Log in and register to the platform;
- Create and edit their profile and upload any previous articles that they want to be used for text generation;
- Receive a list of current breaking news, filter and select a category;
- View the retrieved EO images and the features extracted from them;
- Generate an article, edit, and store it.

The user can fill in the profile when s/he has already registered in the system. The user can set the language that s/he uses when writing an article. Moreover, they can upload their previously published news articles (Figure 2) in order for the system to recognize and extract their writing style. The latter will be used as input to the module that is responsible for the text synthesis (EarthBot) in order to personalize the automatically generated article. The module that is responsible for the journalist profile extraction is described in Section 2.8. The EarthBot module is described in Section 2.9. The final article will be composed of the automatically personalized generated text, along with any retrieved images selected from the user. The images that will be included in the final article will be both processed EO images and images retrieved from Twitter's posts.



**Figure 2.** User profile view, depicting the ability of a user to upload and remove any uploaded documents stored in his/her profile.

When the user has created his/her profile, s/he is transferred to the main view of the platform. The main view includes initially a table presenting a list of the breaking news events, along with additional information (Figure 3) resulting from the breaking news detector module, presented in Section 2.5. This table is updated periodically aiming to provide any new information to the user. The user is allowed to interact with this table and filter this list of breaking news according to the features provided for each event, such as the location of an event, etc.

The user is able to be select the area of interest from the interactive map. The next step is to click on the corresponding button in the main view and receive related information about the selected event. Based on the requirements presented in Section 2.1, users tend to search on social media platforms for information about a topic; they are using such information for the creation of an article, along with other news articles and images posted on social media. Therefore, all this information should be available to the user on the related material section in the user interface (Figure 3). The user should be able to look at all the available material and select the articles and posts that s/he wants to be included in the final article.

13 c	of 18
------	-------

≡ EarthPress	Breaking	g News Manage	r					u	seremail@mail.com 🔱
<ul> <li>Uve</li> </ul>	Search for events		Q						
Breaking News 12 Europe X Asia X Tsunami X Volcano X Results 52 Event	4 Australia X Earthquake 5.000 mentions X Location	Tod	ay Last Week	📧 Tsunami Bali, Indonesia		Publ	icity 300 250 150 150	140 1 11	Total mentions: 20.000
Earthquake	Japan	10.000	30 min					WWWW	MMW
Fire	Australia	5.000	45min				02.35.00 02.35.20	02:35:40 02:36:00 02:36:20	3236.40 02.37.00 02.37.20
Fire	France, Europe	200	1h	🗎 Related Ma	terial Found				
Tsunami	Bali, Indonesia	7.000	Zh	Articles found 5		Generate Rep	lort		
Earthquake	Japan	10.000	30 min	Related Texts Select at least o	ne article from the list to generate	report.			
Fire	Australia	5.000	45min		Text Source	Mentions	Popularity	View Source	
Fire	France, Europe	200	1h		Aricle text source 1	100	100	Ø	
Relatively High	Dall Indexada	7 000	tively Low		Aricle text source 2	90	90	Ľ	
					Aricle text source 3	80	80	ß	
					Aricle text source 4	20	20	Ø	
					Aricle text source 5	2	2	ß	
							items per page:	10 - 1 - 10 of 12	< < <b>&gt;</b> >
				Images found 6					
				Show: Latest First		ý (j)	Color	t at land one immediate the	lint to move to report
				Image Name thestaauri com o 00, 00/00/00	аранана Саранананананананананананананананананана	Image Name thissauf con GC 80, 03/00/0000	ere page	Index of a mage Name mage Name mage Name mage Name mage Name name or 0.0000000	
				Related EO Imag Show:	jes	6			
				Latest First			Solect Control of the solect Control of the solect	t at least one Image from the	Ilet to generate report.
						Generate Rep	items per page:	3 🕶 1-3of8	( <b>C &gt; )</b>

Figure 3. Visualization of all the available material and EO images based on the selected event.

As presented in Figure 3, the user can select the posts and articles that will be used for the generation of the final article. Additionally, the user can select from two different categories of multimedia. The first category, called "related images" in Figure 3, includes images retrieved from social media and are the ones posted by users. The second cate-gory, called "related EO images" in Figure 3, includes the EO images retrieved and the processed ones resulting from the image processing module. From the related articles, posts, and images provided, users can select at least one of them to be used in the final article.

Once the user has selected the images that s/he wants to be included in the final article and at least one of the related articles and posts provided, s/he can proceed to the generation of the article. Aiming to increase user experience, a new view is created (Figure 4), allowing the user to view, edit, store, and update the generated article. In particular, through this view the user can remove the keywords that were initially used for the generation of the text, include new keywords from the list provided and re-generate a text and view the generated article. An important element for the generation of the text is the order in which the keywords are entered so that there can be coherence and flow in the generated text. Therefore, the user will be able to re-arrange the order of keywords. In this list, additional keywords can be added. The additional keywords can be retrieved from another list of keywords extracted from the corresponding modules presented in Section 2.9.1.



Figure 4. Additional view presenting the automatically personalized generated text.

This final article will be trimmed and finalized by the users themselves within the given layout constraints. Furthermore, the user is allowed through this view to save the generated article on his/her user profile. In addition, the final article will be downloaded upon request through this view.

# 4. Discussion

In recent years, the amount of data (documents, images, videos, etc.) that are produced daily has increased tremendously. This has raised many challenges in different fields, such as journalism, where the development of new methods that are able to collect and extract useful information from these vast amounts of information is necessary. Journalistic practices, such as data-driven journalism, follow four basic steps toward the creation of a news story, which are: (a) the data collection, (b) data filtering, (c) data visualization, and (d) story generation. Data collection refers to the collection of raw data while data filtering refers to the process of filtering and maintaining only the information relevant to the news story. Data visualization is the process of transforming data in such a way and creating visualizations to help readers to understand the meaning behind the data. Finally, story generation (publishing) refers to the process of creating the story and attaching data and visuals to the story.

However, in each of the aforementioned steps, several difficulties are raised. Multimedia data collection can be laborious due to the huge amount of available data and to the multiple data sources. Moreover, in many cases, programming knowledge may be required in order to collect data from certain sources. Additionally, the collected data may come in different specialized formats and may be difficult to be checked and filtered from a person or a group of persons. Furthermore, much of the collected information may refer to fake news that has to be filtered out. In addition, the extraction of additional knowledge from data, such as the EO data, and the creation of easily understandable visualizations may require expert knowledge. Finally, the combination of the available data and the limited response time for publishing breaking news makes it a very challenging task. The whole process is a time-consuming procedure that requires the development of systems that can tackle the above difficulties.

Towards this scope, this paper presents an innovative Web 3.0 platform, called Earth-Press, aiming to facilitate professionals of the Media industry by automating many of their journalistic practices. More specifically, the platform is intended to act as a supportive tool in many steps of the journalistic workflow, from data collection and data filtering to the extraction of useful information from the collected data and the automatic synthesis of news stories. In addition to the provided services, the platform aims to deliver value-added products to the editors and journalists based on EO data, allowing them, thus, to enrich their publications and news articles. Such information is important in the news that is related to disasters such as floods and fires.

EO data are freely available in large quantities. However, the main obstacle to their wide use by journalists relates to the difficulties in accessing and even processing them so as to extract additional meaningful information without expert knowledge. Through this platform, users have the possibility to retrieve and extract valuable information from EO data and, specifically, from EO images, without having prior knowledge of processing satellite images. The collected data are automatically processed and provided to the journalists as valuable information regarding the effect of a disaster on a certain area.

The major market segments that EarthPress targets are: (a) local newspapers, which are interested in providing breaking news with a high level of personalization; (b) nationwide newspapers, which are more interested in producing worldwide information; and (c) ePress, which are generally more specialized and could be interested in accessing more elaborated information regarding disasters and EO data.

In this paper, the platform's architecture and its components were presented and analyzed, as well as, the requirements received by journalists and the platform's user inter-face. The basic services provided by the platform are: (a) the detection of breaking news related to disasters, (b) the collection of EO data from Copernicus and multi-media data from social media and news sites, (c) the filtering of the collected data based on their relevance with a disaster event and their credibility, (d) the extraction of useful information related to a disaster and its effect on an area from EO data through the use of image processing techniques and (e) the automatic text synthesis of news stories through the use of AI models that are personalized according to the writing style of each journalist. Each of these services includes many challenging tasks, from multisource and multi-media data acquisition and filtering to the fake news detection and the processing of the EO data, as well as the generation of personalized news articles. The latter deals with various challenging tasks of the NLP field, such as the extraction of the most important keywords from a given text, the writing style transference for the personalization of the generated text, and the text generation of the news article. Text generation is an open research topic whose scope is to provide human-like written text that is coherent and meaningful. The training of AI models that deal with text generation is usually a quite demanding process that requires resources, a huge corpus of training data, and many hours of training. In the scope of EarthPress, the generated text should be also relative to the news story's topic and should avoid containing misleading or erroneous information.

All the aforementioned services included within the integrated platform of EarthPress can facilitate the journalists significantly in order to reduce the needed time for collecting, evaluating, filtering, combining, and finally publishing news stories ready-to-be printed.

However, there are some limitations in the proposed platform that should be taken into consideration. As it was mentioned previously, text generation is a challenging NLP task of great research interest. The final generated text should be paid attention to in order to avoid including misleading or false information, while it should be also coherent and relevant to the news topic. The fake news detector should check the generated text so the credibility of the generated text can be ensured. Moreover, the retrieval and processing of EO data may be a time-consuming process that may require some hours of processing. Additionally, the accuracy of the results may vary depending on the characteristics of the available EO data (e.g., different Spatio-temporal resolutions, cloud cover, etc.).

# 5. Conclusions

With the advent of Big Data, the need for innovative tools that are able to handle this vast amount of data and extract information of added value has emerged. In the field of journalism, the latest journalistic practices require the automatic collection and filtering of information related to a news story from multiple sources, as well as the automation of the procedure of the news stories generation.

In this work, the concept of the EarthPress platform was presented, an assistive Web 3.0 platform that targets to facilitate professionals of the media sector in each step of the journalistic procedure. The EarthPress platform intends to facilitate journalists in collecting data and automatically composing articles related to disasters such as floods and fires. Moreover, it aims at extracting added value information from EO data that can be included in the final news story. Following this, in this paper, the system's architecture of EarthPress is presented, which consists of several modules and sub-modules, each of which serves a different purpose. The main modules of the EarthPress as they are presented in Figure 1 are, namely, (a) the user interface, (b) the user manager, (c) the database manager, (d) the breaking news detector, (e) the data fusion, (f) the EO image processing, (g) the journalist profile extractor, and (h) the EarthBot.

For the definition of the architecture and the user interface of EarthPress, a workshop with more than 130 participants of the media sector was conducted, aiming to identify the requirements and the specification of the EarthPress platform. The collected requirements indicated that a tool or a platform that can facilitate the journalistic procedure in its different steps is of great interest to the media sector. Additionally, the presented platform will include AI methods for the validation of the collected data to avoid using misleading information for the generation of the final news article.

The future steps regarding EarthPress are the implementation and the testing of each of the architecture's modules, as well as the evaluation of the platform in real-time with real users of the media sector.

**Author Contributions:** Conceptualization, A.D.; methodology, A.D., A.Z., M.T., and E.E.; writing original draft preparation, E.E., A.Z., and M.T.; writing—review and editing, E.E., A.Z., and M.T.; supervision, A.D.; project administration, D.T.; funding acquisition, A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by grants from Horizon 2020, the European Union's Programme for Research and Innovation under grant agreement No. 870373-SnapEarth. This paper reflects only the authors' view, and the Commission is not responsible for any use that may be made of the information it contains.

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Pileggi, S.F.; Fernandez-Llatas, C.; Traver, V. When the Social Meets the Semantic: Social Semantic Web or Web 2.5. *Future Internet* **2012**, *4*, 852–864. [CrossRef]
- 2. Fuchs, C.; Hofkirchner, W.; Schafranek, M.; Raffl, C.; Sandoval, M.; Bichler, R.M. Theoretical Foundations of the Web: Cognition, Communication, and Co-Operation. Towards an Understanding of Web 1.0, 2.0, 3.0. *Future Internet* **2010**, 2, 41–59. [CrossRef]

- 3. Kotenidis, E.; Veglis, A. Algorithmic Journalism—Current Applications and Future Perspectives. J. Media 2021, 2, 244–257. [CrossRef]
- 4. Gomes, V.; Queiroz, G.; Ferreira, K. An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sens.* 2020, *12*, 1253. [CrossRef]
- 5. Romeo, A.; Pinto, S.; Loekken, S.; Marin, A. Cloud Based Earth Observation Data Exploitation Platforms. Available online: https://ui.adsabs.harvard.edu/abs/2017AGUFMIN21F.03R/ (accessed on 17 June 2021).
- 6. De Groeve, T.; Vernaccini, L.; Annunziato, A.; Van de Walle, B.; Turoff, M. Global disaster alert and coordination system. In Proceedings of the 3rd International ISCRAM Conference, Brussels, Belgium, 1 January 2006.
- Lee, J.; Niko, D.L.; Hwang, H.; Park, M.; Kim, C. A GIS-based Design for a Smartphone Disaster Information Service Application. In Proceedings of the 2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering, Jeju, Korea, 23–25 May 2011; pp. 338–341.
- Nishikawa, M.S. GLobal Unique Disaster IDEntifier Number (GLIDE): For Effective Disaster Information Sharing and Management. In Proceedings of the International Conference on Total Disaster Risk Management 2003, New York City, NY, USA, 2–4 December 2003; Volume 2, p. 4.
- Badrinarayanan, V.; Handa, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv* 2015, arXiv:1505.07293.
- 10. Gupta, R.; Shah, M. Rescuenet: Joint building segmentation and damage assessment from satellite imagery. *arXiv* 2020, arXiv:2004.07312.
- Gupta, R.; Goodman, B.; Patel, N.; Hosfelt, R.; Sajeev, S.; Heim, E.; Doshi, J.; Lucas, K.; Choset, H.; Gaston, M. Creating xBD: A dataset for assessing building damage from satellite imagery. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 10–17.
- 12. Kordelas, G.A.; Manakos, I.; Aragonés, D.; Díaz-Delgado, R.; Bustamante, J. Fast and Automatic Data-Driven Thresholding for Inundation Mapping with Sentinel-2 Data. *Remote Sens.* 2018, *10*, 910. [CrossRef]
- 13. Kordelas, G.A.; Manakos, I.; Lefebvre, G.; Poulin, B. Automatic Inundation Mapping Using Sentinel-2 Data Applicable to Both Camargue and Doñana Biosphere Reserves. *Remote Sens.* **2019**, *11*, 2251. [CrossRef]
- 14. Parsons, A.; Robichaud, P.; Lewis, S.; Napper, C. Field Guide for Mapping Post-Fire Soil Burn Severity. Available online: https://www.fs.fed.us/rm/pubs/rmrs\_gtr243.pdf (accessed on 17 June 2021).
- 15. Emergency Mapping Guidelines. UN-SPIDER. 2018. Available online: https://www.un-spider.org/sites/default/files/IWG\_ SEM\_Guidelines\_Fire\_chapter\_SERTIT\_2\_0.pdf (accessed on 17 June 2021).
- 16. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
- 17. Gao, Z.; Feng, A.; Song, X.; Wu, X. Target-Dependent Sentiment Classification with BERT. *IEEE Access* 2019, 7, 154290–154299. [CrossRef]
- 18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* 2017, arXiv:1706.03762.
- 19. Liu, Y.; Lapata, M. Hierarchical transformers for multi-document summarization. In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5070–5081. [CrossRef]
- Goodwin, T.R.; Savery, M.E.; Demner-Fushman, D. Flight of the PEGASUS? Comparing Transformers on Few-Shot and Zero-Shot Multi-document Abstractive Summarization. In Proceedings of the COLING—International Conference on Computational Linguistics, Barcelona, Spain (Online), 8–13 December 2020; Volume 2020, p. 5640.
- 21. Liu, P.J.; Saleh, M.; Pot, E.; Goodrich, B.; Sepassi, R.; Kaiser, Ł.; Shazeer, N. Generating wikipedia by summarizing long sequences. *arXiv* 2018, arXiv:1801.10198.
- 22. Chen, W.; Chan, H.P.; Li, P.; Bing, L.; King, I. An integrated approach for keyphrase generation via exploring the power of retrieval and extraction. *arXiv* 2019, arXiv:1904.03454.
- 23. Zhang, J.; Zhao, Y.; Saleh, M.; Liu, P.J. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv* **2019**, arXiv:1912.08777.
- 24. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequenceto-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv* **2019**, arXiv:1910.13461.
- 25. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv* 2019, arXiv:1910.10683.
- 26. Harkous, H.; Groves, I.; Saffari, A. Have your text and use it too! end-to-end neural data-to-text generation with semantic fidelity. *arXiv* 2020, arXiv:2004.06577.
- 27. Kanerva, J.; Rönnqvist, S.; Kekki, R.; Salakoski, T.; Ginter, F. Template-free data-to-text generation of Finnish sports news. *arXiv* **2019**, arXiv:1910.01863.
- 28. Rebuffel, C.; Soulier, L.; Scoutheeten, G.; Gallinari, P. A hierarchical model for data-to-text generation. In *European Conference on Information Retrieval*; Springer: Cham, Switzerland, 2020; pp. 65–80.
- 29. Kale, M. Text-to-text pre-training for data-to-text tasks. arXiv 2020, arXiv:2005.10433.

- 30. Campos, R.; Mangaravite, V.; Pasquali, A.; Jorge, A.; Nunes, C.; Jatowt, A. YAKE! Keyword Extraction from Single Documents Using Multiple Local Features. Information Sciences. Available online: https://www.researchgate.net/publication/335766438\_ YAKE\_Keyword\_Extraction\_from\_Single\_Documents\_using\_Multiple\_Local\_Features (accessed on 17 June 2021).
- 31. Gunasiri, D.Y.; Jayaratne, K.L. Automated cricket news generation in Sri Lankan style using natural language generation. *Eur. J. Comput. Sci. Inf. Technol.* **2019**, *7*, 42–56.