social sciences

MDPI

Article

A Participatory Journalism Management Platform: Design, Implementation and Evaluation

Kosmas Panagiotidis [1,*], Nikolaos Tsipas [2], Theodora Saridou [1] and Andreas Veglis [1]

[1] Media Informatics Lab, School of Journalism & MC, Aristotle University of Thessaloniki, Thessaloniki 541 24, Greece; saridout@jour.auth.gr (T.S.); veglis@jour.auth.gr (A.V.)
[2] Laboratory of Electroacoustics and TV Systems, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 541 24, Greece; nitsipas@auth.gr
* Correspondence: kpanagioti@jour.auth.gr

Received: 20 January 2020; Accepted: 19 February 2020; Published: 22 February 2020

Abstract: During the last two decades, citizens' participation in news production process has attracted significant interest from both academia and the media industry. Media production and consumption have been altered considerably and traditional concepts, such as gatekeeping, have been under discussion. Many news organisations include in their websites tools and applications that allow users to be active consumers or even co-producers of journalistic content, by liking, sharing, commenting and submitting material. At the same time, large amounts of user-generated content are uploaded every day on social media platforms. Subsequently, media organisations must deal with continually available information which requires management, classification and evaluation not only to keep high journalistic standards, but also to avoid problems. The latter category can include grammar mistakes, fake or misleading information and hate speech. All the above-mentioned parameters highlight the obvious need for platforms that can support journalism manage practice. Such a platform should utilise semantic technologies, which can support organised collection and moderation of content in an effective way and in short time. This study discusses the design and the implementation of a participatory journalism management platform.

Keywords: participatory journalism; semantic analysis; semantic web

1. Introduction

In the current media landscape, news production has moved into a diversifying ecosystem, where the evolutionary changes turn mere consumers into collabourators (Gillmor 2009) and Internet users into prosumers (Deuze 2003). Mainly in the form of text, photos, video and graphics (Anderson 2007), user-generated content (UGC) is adopted by media organisations, which integrate non-professional contributions in their daily work routine, either directly or indirectly (Maniou and Veglis 2016; Maniou et al. 2017). Thus, professional journalists and amateur users co-produce news within a mainstream platform in a participatory journalism context (Aitamurto 2013).

Therefore, journalistic tasks encompass content curation, and especially summarisation, clarification and verification of the vast amount of information circulating around particular issues and stories (Bradshaw 2013). Semantic Web services can be a useful ally in the direction of managing participatory journalism in an effective way and in short time, thanks to their aggregation features, regarding the advanced ability of collecting information (Heravi et al. 2012). Although the exploitation of structured data enables the development of new tools and methods for information processing (Heravi et al. 2012), such semantic techniques have not been widely adopted by media organisations (Brandtzaeg et al. 2016; Pomp et al. 2018), mainly due to the immature semantic technology and the lack of a unified platform in which journalists can concentrate in the utilisation of UGC (Saridou et al. 2018).

For this reason, the study discusses the design, implementation and evaluation of a participatory journalism platform, which aims at augmenting UGC management, moderation and discovery, based on existing state-of-the-art semantic technologies. The prevailing circumstances are the ones that defined the reasons behind this research. Namely, although the exploitation of the available UGC content by media organisations can enrich the content provided, its high volume and mass production raises obstacles in managing the disseminated information and evaluating its quality. Nowadays, journalists' low level awareness of tools and practices to control UGC information flows, don't allow them to tackle the problem. Many of the proposed solutions (i.e., semantic technologies, techniques, tools, web services, etc.) require some expertise. However the absence of the accompanying training activities or even their involvement in the design and evaluation of the aforementioned solutions, makes it difficult to fully understand how they work and therefore effectively utilise them in real-world scenarios. Another reason why this research took place was because although significant steps have been taken at the research level, the progress at application level is still limited. The improvement, simplification and homogenisation of an integrated computational solution demands the coexistence of journalists and citizens in a web environment where the formation and interconnection of technologies, tools and methods is supported. The introduced workflow is optimised for simplicity of use and reduced operational overhead in terms of time and resources. A prototype of the platform is evaluated by a group of professionals with a focus on the assessment of the usability of the proposed workflow and the identification of future improvements and research directions.

2. Literature Overview

2.1. Participatory Journalism

The drastic changes in the news media environment have led media organisations to devote significant amounts of resources in order to embrace and incorporate UGC into their journalistic workflows. Participatory journalism is considered as a process in which citizens contribute to professional journalists' news production (Abbott 2017). Drawing on a classification of citizen participation, scholarship often focuses on the different stages of news production (Singer et al. 2011), on the different propositions of the communication process (Franquet et al. 2011) or on the concept of interactivity (Suau and Masip 2014). According to Spyridou (2019), the most prevalent and widespread participatory tools are (i) content rating, (ii) polls, (iii) sharing through social networks, (iv) audience footage in the form of audiovisual material, (v) collaborative content, (vi) comments, (vii) discussion forums, (viii) submission of textual material and (ix) citizen blogs.

Over the last few years, leading media organisations focused on building heterogeneous platforms providing the means to collect, report, analyse and disseminate UGC-driven news and information. Notable examples are the CNN iReport (Hellmueller and Li 2015), the BBC UGC Hub (Harrison 2010), the Reuters UGC Newsgathering platform (Wilks-Harper 2018), the Associated Press UGC Video Hub (Shanley 2017), the Guardian Witness (Geary 2013) and the Huffington Post Firsthand (Sheffield 2012). At the same time, social media become central to the way people experience news as networked media technologies are extending the users' ability to create and receive personalised news streams (Hermida et al. 2012). Nowadays many people use social media like Facebook and Twitter as a primary source of information. People tend to exchange more and more personal data and give their friends and followers the chance to stay updated on what is going on (Spettel and Vagianos 2019) not only in their lives, but also in the lives of others. Instead of actively choosing to visit a news website or explicitly searching for a news topic, now news is passively found in posts shared from friends, family or news sources that users follow (Bentley et al. 2019).

While managing users' contributions, media organisations often face a long series of ethical, legal and responsibility issues that stem from audience engagement (Saridou and Veglis 2016; Singer et al. 2011), as the incorporation of UGC by the journalists is not an easy task (Veglis 2013). Specifically, many professionals underline the excessive use of inappropriate language, flaming, stereotyping, superficial

discourse and incivility that impede constructive public discourse (Coe et al. 2014; Manosevitch 2011). Furthermore, dark participation ranges from misinformation and hate campaigns to individual trolling and cyberbullying (Quandt 2018), whereas the spreading of fake news, disinformation and conspiracy theories in UGC are forms of deviance as well (Frischlich et al. 2018).

Pre-moderation is one of the methods used in order to check UGC before publication. Although it can ensure high quality levels, this option is considered as labour-intensive and costly (Santana 2014; Singer et al. 2011). On the other hand, post-moderation leads to simpler and more open control policies, but can negatively affect the quality of the website; this kind of moderation is nearly always accompanied by mandatory prior registration (Hille and Bakker 2014). News outlets often involve users in the control process via distributed moderation (Hille and Bakker 2014) or employ purely technical methods, such as automated moderation, under predefined filters that detect and replace problematic content (Veglis 2014). Additionally, Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHA) can also be adopted for preventing automated bots from conducting nefarious activities (Sivakorn et al. 2016). Apart from the requirements for adequate human, financial and time resources, the journalists' moderation decision is affected by newsroom routines, by the type of media organisation they work for, by societal institutions and the social system in which they operate, their personal experiences or gut feelings (Boberg et al. 2018).

2.2. Contributing Semantic Web Services and Tools

The visionaries Tim Berners-Lee, James Hendler and Ora Lassila introduced the concept of Semantic Web in an article for Scientific American Magazine, in May 2001. They stressed that Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Hendler et al. 2001). Simply put, the aim of the Semantic Web is to make the Web more accessible to computers (Antoniou and Van Harmelen 2004). The current Web is a network of unstructured interconnected forms of information (documents, images, text) accessible via Internet. Its content is designed only for humans to read and manipulate (Shadbolt et al. 2006) leaving the beneficial properties of computer programs unexploited. In the Semantic Web where the meaning of every piece of information will be clearly defined and transmitted, computers will be able to interpret information like humans and intelligently generate and distribute useful content tailored to users' needs (Saridou et al. 2018).

Given the features of today's data as Web-based information, semistructured, numeric and totally unstructured (video and audio), certain related tasks can become difficult to complete. This is the point where Semantic Web steps in (Gross 2014) and provides suitable solutions. The development of the Semantic Web will introduce a Web of data (Choudhury 2014) in which the relations of every data item with others, will be fully clarified and structured. As a result, the interconnection of concepts rather than just documents will be feasible. Such a technological evolution will definitely unlock various prospects of data exploitation for journalists and media professionals. Consequently, this progress along with the technologies and services involved, will surface new investigation methods and tools. Journalists seem to need them more than ever before in order to deal with today's exploding Web landscape, where vast amounts of information are published every minute from various sources across the world.

Just recently, some of the aforementioned media organisations started to work towards the utilisation of semantic services, technologies and tools in various ways, from speeding up research to accumulating and cross referencing data (Underwood 2019). For example services, providing the means to extract information from UGC by semantically analysing the aural, visual and textual components, have become increasingly popular. Spectee (Hawkes 2017) is a service that automates the extraction of semantic information from video data in order to simplify categorization and analysis processes performed by journalists. Aural content can be exploited by services like Trint (Gershgorn 2017), which provides the ability to generate transcriptions from audio to enhance audiovisual content search and classification mechanisms. More generally, there is an ongoing trend to democratise access to AI through a number of algorithms provided as services on cloud platforms (Amazon 2019) or as open

source projects (Loper and Bird 2002). Overall, these tools and services enable the automated and semi-automated verification of content uploaded by citizens.

Last, services like Google Street View (Anguelov et al. 2010) are used to correlate content with geolocation data, whereas tools like Twitteraudit (TwitterAudit 2019) can be employed to verify the authenticity of a social media account. Furthermore, the rise of Fake News has accelerated the development of novel detection mechanisms employing the aforementioned services as building blocks. Noteworthy examples include Factmata (Factmata 2019) and Fabula AI (FabulaAI 2019), services that provide journalists with ability to better understand multimedia content.

2.3. The Greek Context

The Greek media system is part of the Polarised Pluralism model, sharing the basic characteristics of other Mediterranean countries (Hallin and Mancini 2004). Although the rise of blogs and online media platforms fundamentally transformed the news media landscape (Michailidou et al. 2014), early research in Greek online newspapers showed little interactivity, whereas attempts that render the public as active receivers and producers of news become part of established journalistic practices with difficulty (Spyridou and Veglis 2008). Professionals seem attached to traditional practices of producing news, unwilling to surrender gatekeeping control (Siapera and Spyridou 2012). Greek professionals have to deal with a number of problems such as spelling and syntactical mistakes along with intellectual property violations, fake news, personality insults, hate speech, defamation, spamming and trolling. At the same time, heavy workload, high content heterogeneity and low content standardisation seem to be the biggest obstacles to UGC exploitation (Saridou et al. 2019). However, Greek journalists state eagerness towards evolution in the context of participatory journalism, as they are not satisfied with the functionality of the UGC management tools they use and believe in the need for further education in this field, whereas an overwhelming majority would spend time to get trained as well (Saridou et al. 2019).

Thus, the synthesis of accumulated literature regarding participatory journalism and semantic Web services emphasises the potential to facilitate a better management of UGC within the professional context by utilising semantic solutions. The employment of existing tools and applications could help journalists handle the high volume and lying problems of users' contributions without burdening their daily work schedule. Against this background, the SEMPRATO project was set up with the procedure described below.

3. Methodology

Envisioning the production of a semantic-oriented participatory journalism management platform, our research extended over a period of nine months and was carried out in three steps, the design, implementation and evaluation of the platform. The first step, focusing on the platform design, consisted of two stages: As part of the first stage, two online surveys were conducted to define and formalise the basic characteristics of such a model (Saridou et al. 2019), whereas during the second, the available semantic services, applications and tools in the context of journalism were identified and analysed (Panagiotidis et al. 2019). As illustrated in Figure 1, during the design step, the model requirements and features were articulated. During the second step, the implementation of these features led to the deployment of a prototype of the platform, based on a collection of tools and processes providing support and guidance to journalists towards the effective management of UGC. The introduced workflow was optimised for simplicity of use and reduced operational overhead in terms of time and resources.

After its implementation, and during the third step, the prototype was evaluated by a group of experts who completed a series of actions on the platform, acting as users and journalists in the context of a collaborative journalism project. Finally, the participants completed an online questionnaire, with a focus on the assessment of the usability of the proposed workflow, while future improvement possibilities and research directions were identified.

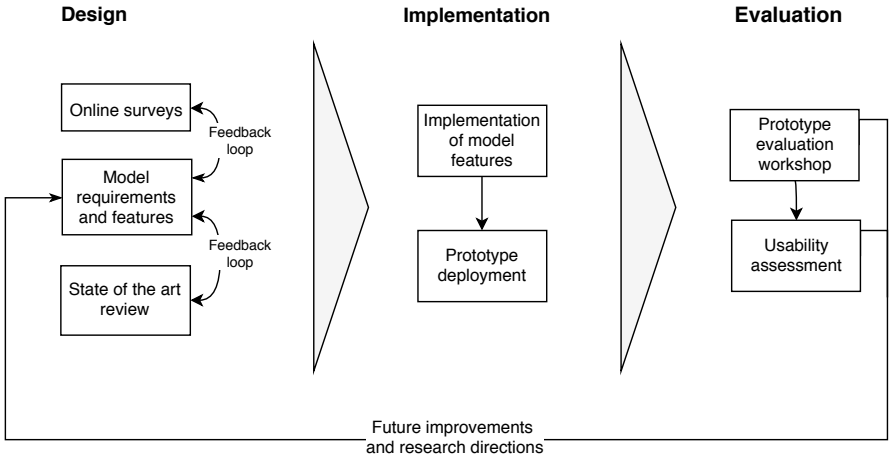

Figure 1. Illustration of the employed methodological approach.

4. Design

To design a new model of managing participatory journalism, two main steps were followed: The first was the definition of the requirements and the description of the basic characteristics that such a model should have (Saridou et al. 2019), and the second one was the analysis of the available semantic services, applications and tools in the context of journalism (Panagiotidis et al. 2019). For the initial purpose, a requirements gathering stage was undertaken, employing agile software development techniques (Kavitha and Thomas 2011). Two online surveys of journalists and users in Greece were conducted, to gain some insights concerning the development of the model (Saridou et al. 2019). The findings were fed back to the requirements gathering stage in order to refine the feature selection process (Saiedian and Dale 2000). Research interest was focused on UGC management in journalists' work routine and their perceptions of the importance of education in this process, whereas users' participation in content production in media organisations was also studied. Finally, all survey participants were asked to indicate their attitude towards a semantic-oriented participatory journalism management model.

A questionnaire consisting of 22 questions was distributed to journalists of the Journalists' Union of Macedonia-Thrace Daily Newspapers, of the Journalists' Union of Thessaly-Sterea Ellada-Evia Daily Newspapers and to professionals of online media 3 (magazines, newspapers and news portals) via email. In total, 52 people responded. At the same time, the second survey was conducted among undergraduate, postgraduate and PhD students of the School of Journalism and Mass Media Communications of the Aristotle University of Thessaloniki. A hyperlink that led to a questionnaire consisting of 11 questions was uploaded on the relevant Facebook group. Overall, 52 users responded. The data were collected via anonymous responses, in a one-month period, between October, 9 2018 and November 9, 2018, using Google Forms. Data analysis accentuated a clear agreement with a new, collaborative way of managing UGC in a semantic context, as user engagement was deemed effective both from journalists and from users themselves. Participants' responses highlighted the need for a model that is fast to use and part of the existing journalistic workflow in order to facilitate daily newsroom routines. It should allow automatic categorization of UGC, detection of malicious content, moderation by journalists and problem flagging by users. Furthermore, to serve verification purposes, it should support features for data export related to the place and time that multimedia content was acquired. Both surveys also indicated that it would preferably be embedded on the organisation's website, able to work both through computer browsers and mobile applications. Additionally, it should incorporate some characteristics inspired by social networking platforms, such as "like" and "comments", as the majority of the participants seemed familiar with such social media functions.

By examining specific applications and tools adopted by global media organisations, the second step detected the existing use of semantic Web technologies in journalism. As explained above, several

organisations are experimenting with the integration of semantic technologies aiming at various objectives. The authors of Panagiotidis et al. (2019) categorised these objectives in three distinct levels: data, information and content. Specifically, at the first level, data are treated as abstract information (raw data), the second level includes tools with acquirement, management and process properties, and the third one refers to content exploitation (exploitation of already published or ready to be published information). Figure 2 indicates the objectives as a series of multiple semantic oriented actions classified in the aforementioned levels.

To design a new model of managing participatory journalism, two main steps were followed: The first was the definition of the requirements and the description of the basic characteristics that such a model should have (Saridou et al. 2019), and the second one was the analysis of the available semantic services, applications and tools in the context of journalism (Panagiotidis et al. 2019). For the initial purpose, the authors of Saridou et al. (2019) conducted two online surveys of journalists and users in Greece to gain some insights concerning the development of the model. Research interest was focused on UGC management in journalists' work routine and their perceptions of the importance of education in this process, while users' participation in content production in media organisations was also studied. Finally, all survey participants were asked to indicate their attitude towards a semantic-oriented participatory journalism management model.

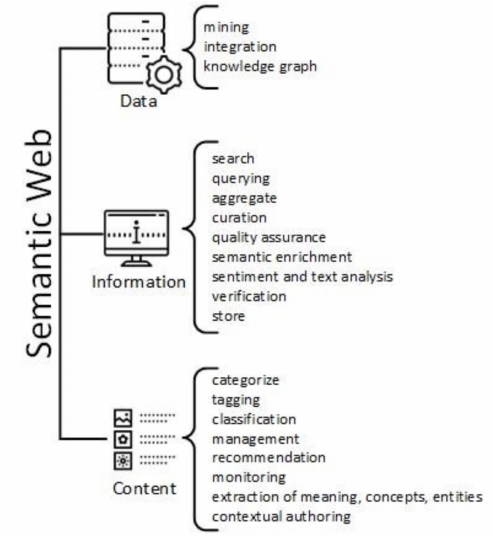

Figure 2. Objectives achieved in the semantic context.

5. Implementation

The proposed participatory journalism management platform focuses on augmenting the adoption of UGC by news and media organisations by enhancing day-to-day workflows employed by journalists. SEMPRATO (Semantic participatory journalism platform) is organised around a main workflow which provides a collection of tools and processes to support and guide the journalist with suggestions on how to approve, edit and publish UGC. A high level diagram of the proposed platform is illustrated in Figure 3 where the main steps employed for the enrichment of raw UGC are indicated. Based on the implement workflow, UGC goes through three different stages: the submission stage, where it is still in raw form as submitted by the user; the enriched stage, where the content has been augmented to simplify the approval, editing and publication steps performed by the journalist; and, finally, the published stage, where UGC becomes available for interaction with the audience.

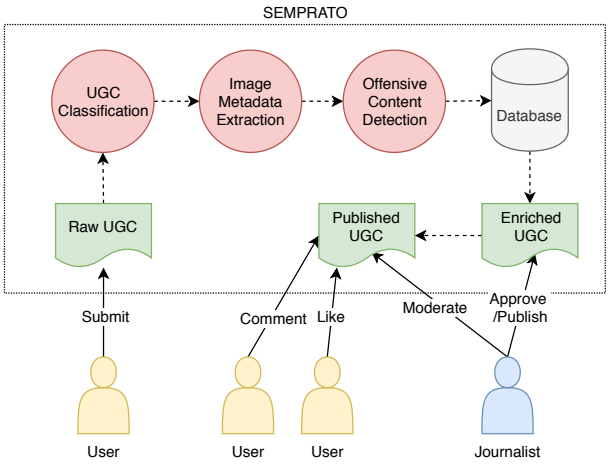

Figure 3. User and journalist interactions with SEMPRATO.

5.1. An Extensible Open-Source Platform

SEMPRATO is an extensible platform that can be used as a starting point in similar scenarios/experiments. The implemented participatory journalism management platform is based on the python web-framework Django and utilises an SQLite database for state storage and retrieval. The application was deployed on a Kubernetes (Bernstein 2014) cluster running on the Okeanos cloud-platform (Koukis et al. 2013) using Docker containers. The application can be easily executed locally using Docker, which reduces the operational and infrastructure management overhead during evaluation sessions/workshops and enables reproducible research (Boettiger 2015). As part of this work, SEMPRATO (https://github.com/nicktgr15/semprato) is released as a publicly available open-source project.

5.2. Content Submission

SEMPRATO is driven by content uploaded by the users, thus the UGC submission step is considered very critical to capturing as much raw information as possible during the initial interaction with the user. As indicated in Figure 4, the submission mechanism relies on a web form where the user provides information about a number of available fields. Apart from the basic ones which are the title, the main text and images, the user can also provide an indicative category as well as descriptive text tags that can be used for UGC retrieval at a later stage. In conventional platforms, UGC normally goes through a manual verification process that is carried out by a journalist.

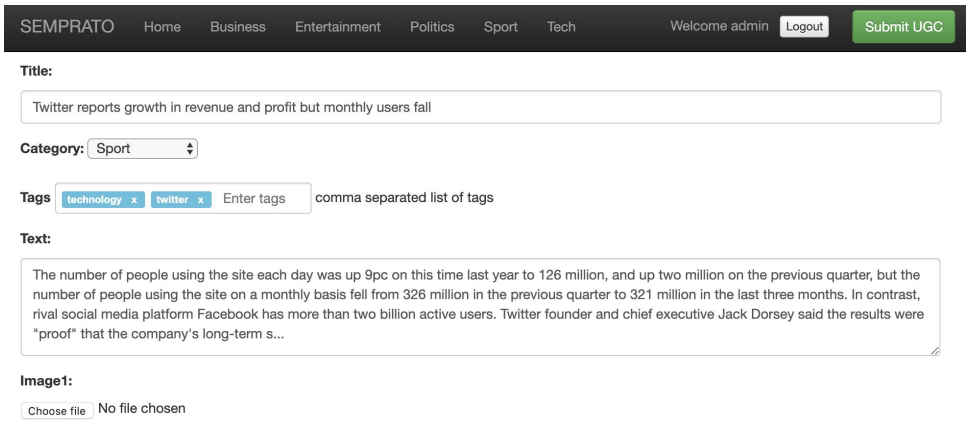

Figure 4. The UGC submission form employed in SEMPRATO.

5.3. Content Enrichment

To support and simplify the manual verification process performed by the journalist, SEMPRATO employs a number of postprocessing steps for the enrichment of raw UGC. A natural language processing (NLP) based machine-learning algorithm is employed for the category classification of submitted UGC. A publicly available BBC dataset (Greene and Cunningham 2006) consisting of 2225 documents from the BBC news website derived in 2004–2005 is employed. The documents cover five topical areas (classes): business, entertainment, politics, sports and tech. The term frequency-inverse document frequency (tf-idf) statistic is used as a feature vector for every document whereas a multinomial Naive Bayes classifier (Pedregosa et al. 2011) is used for the categorisation of the documents. Using a train/test split approach, with 33% of the data retained for testing, the classification algorithm achieved a weighted f1-score of 0.953. The inferred category for the submitted UGC is used to enrich the submitted content and help the journalist during the editing and verification phase before publication.

A subsequent enrichment step focuses on exploiting metadata available on the submitted image data. Two metadata fields are analysed: the gps coordinates and the date-time information. These two fields are almost always available on images captured using smartphones, which is the main scenario investigated as part of this work. With this information made available to the journalist, the cross-referencing and validation of images can become significantly simpler. As indicated in Figure 5, the gps coordinates visualisation allows the selection of the appropriate image for a story created from UGC.

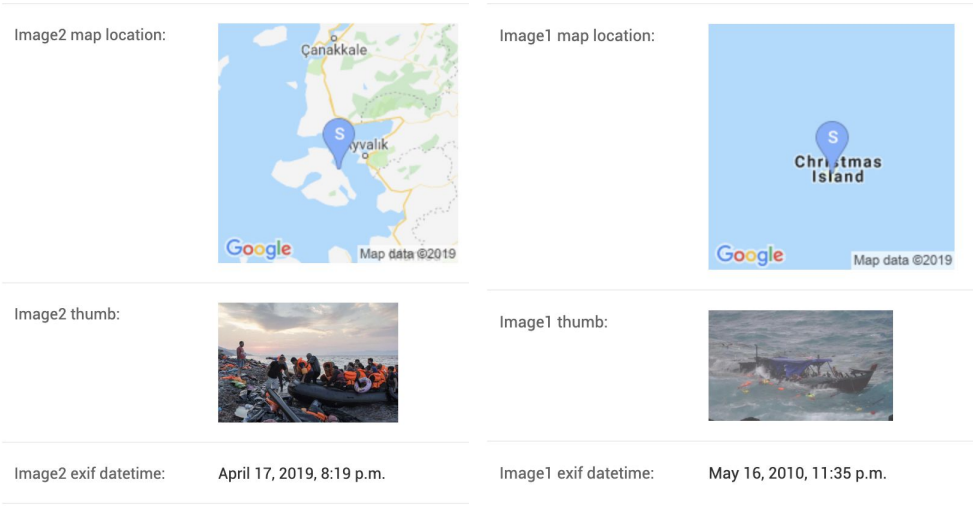

Figure 5. Two images available as part of UGC submitted for an article entitled "In Greece, most immigrants arrived in Europe in the first five months of 2019". The right image was captured in the Pacific ocean in 2010 while the left one in the Aegean sea in 2019.

Additionally a rule-based approach is followed to flag irregular date-time information in UGC. An example is presented in Figure 6.

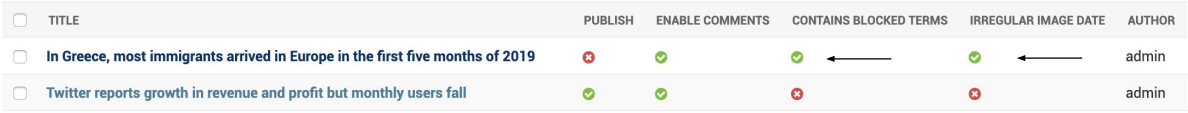

Figure 6. Irregular date-time information identified for the uploaded images during UGC submission.

The detection of offensive content is another mechanism of raw UGC enrichment that simplifies the validation process for the journalist during the editing stage. SEMPRATO comes with a long list of offensive keywords and any uploaded content is compared against that list. The administrators

of the platform can easily modify the list of offensive terms in order to achieve the desired level of moderation. The offensive content is not moderated automatically, it is only highlighted and the journalist is able to make an ad hoc decision based on the context. An example of a highlighted blocked term is illustrated in Figure 7, whereas an example of an article containing blocked terms is presented in Figure 6.

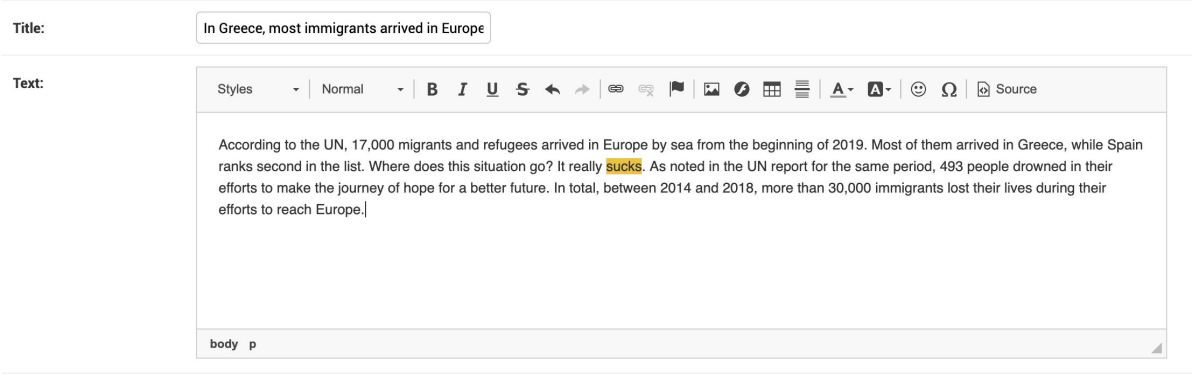

Figure 7. Offensive content highlighted in submitted UGC.

5.4. Content Publication

During the publication stage, the journalist accesses the enriched UGC and decides which parts of the content are appropriate for publication. The information generated during the enrichment stage allows journalists to categorise content, avoid misinformation and maintain high quality standards by blocking offensive terms. After publication, the journalist can enable commenting on UGC and use the provided tools to handle comments submitted by the readers. SEMPRATO provides the means to automatically block comments containing offensive content or manually hide specific comments not conforming to the organisation rules.

6. Evaluation

The evaluation of the SEMPRATO platform was conducted by a group of 28 individuals, which was comprised by professional journalists, postgraduate and phd students in the field of journalism, communication and media technologies. Thus, the sample could be considered as experts, as they are all very competent users of internet related tools and applications. The evaluation was conducted during the first 10 days of June 2019, and it took place in a computer lab where the SEMPRATO platform was presented in detail by the members of the research team. The participants were instructed to complete a series of actions on the platform initially in the role of a user that is contributing to a collaborative journalism project and then in the role of a journalist that is supervising a collaborative journalism project. Finally, the participants completed a web questionnaire, based on Google Forms. Fifty-seven percent of the participants were male and 43% female; 39.3% of participants belonged to the 25–30 age group and the second biggest group (28.6%) were the over 40 age group. The rest of the participants were organized as follows, 17.9%, 31–35 age group, 7.1%, 18–24 age group and the rest of the sample belonged to the 36–40 age group. As far as the educational level is concerned most of the users in the sample (53.6%) had or were attending postgraduate studies and 28.6% had a university degree. 39.3% of the participants worked at the private sector, 35.6% were students and 17.9% stated that they were unemployed.

6.1. Evaluation as uSers Contributing in a Collabourating Project

As it was previously stated, initially the participants were directed to work on the SEMPRATO platform as contributors to a collaborative journalism project. Thus, they were asked to evaluate the ease of use of several actions. All actions were evaluated very positively, as indicated in Figure 8.

Comment submission, document submission and photo submission received the highest scores whereas account activation collected the lowest score.

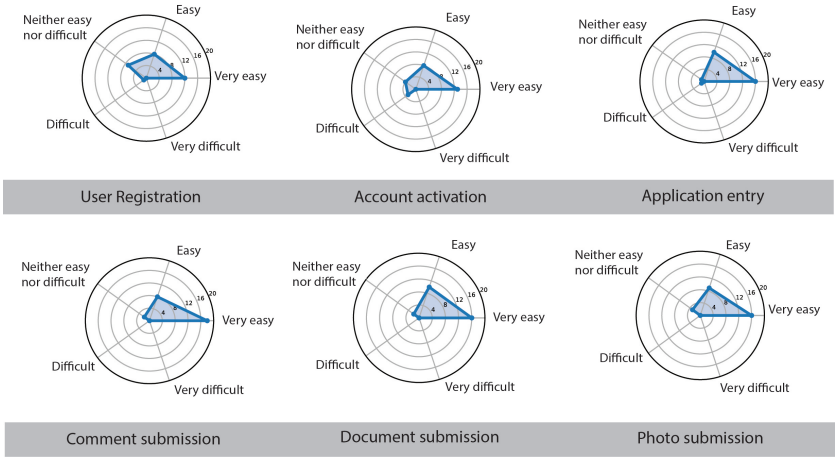

Figure 8. Evaluation of the ease of use of different actions (contributors).

Next, they were asked to evaluate the usefulness of certain features (article categorization, tag definition, comment submission, document submission, photo submission and positive evaluation). Most of the participants found the features useful and very useful (detailed results can be found in Figure 9).

In the last section of the evaluation of the SEMPRATO platform as contributors, the participants were asked to rate from 1 up to 5 (1 the lowest and 5 the highest) the usefulness of the platform (average 3.8), their satisfaction in regards the management options provided (average 3.8), the supported workflow (average 3.8), and the organisation of the platform (average 4). Finally, they were asked how likely was to recommend SEMPRATO to a friend and the average rating was 3.9. The above results indicate a quite positive attitude of the participants towards the SEMPRATO platform.

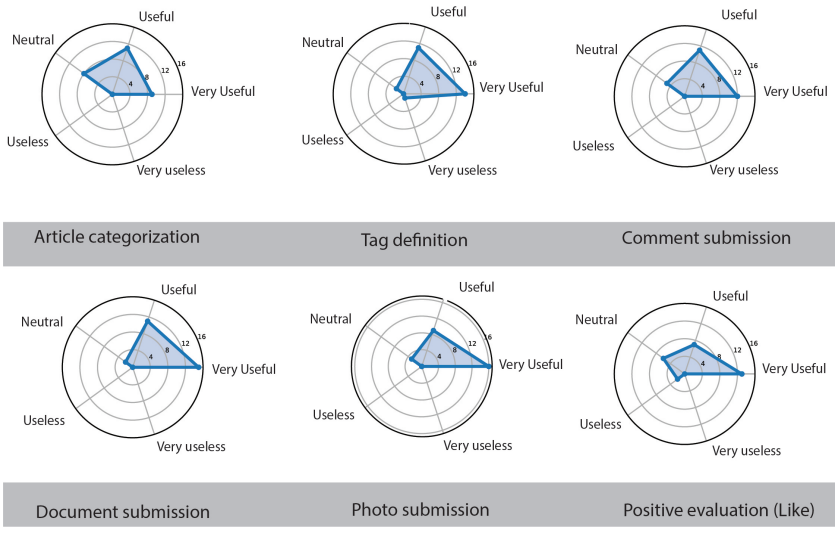

Figure 9. Evaluation of the usefulness of certain features (contributors).

6.2. Evaluation as Journalists Supervising a Collabourating Project

During the second part of the evaluation, the participants were asked to work with certain scenarios in the role of journalists, supervising a project through the SEMPRATO collaboration platform. As depicted in Figure 10, they were asked to evaluate the ease of use of several actions, namely, text correction, disabling comments, changing and correcting content categorization, article

publishing, deletion of malicious content, photo authentication, managing submitted comments, adding a term in the blacklist. Most of the actions were evaluated very positively (easy and very easy), with article publishing and adding a term in the blacklist considered as the easiest actions while photo authentication and disabling comments were considered as difficult.

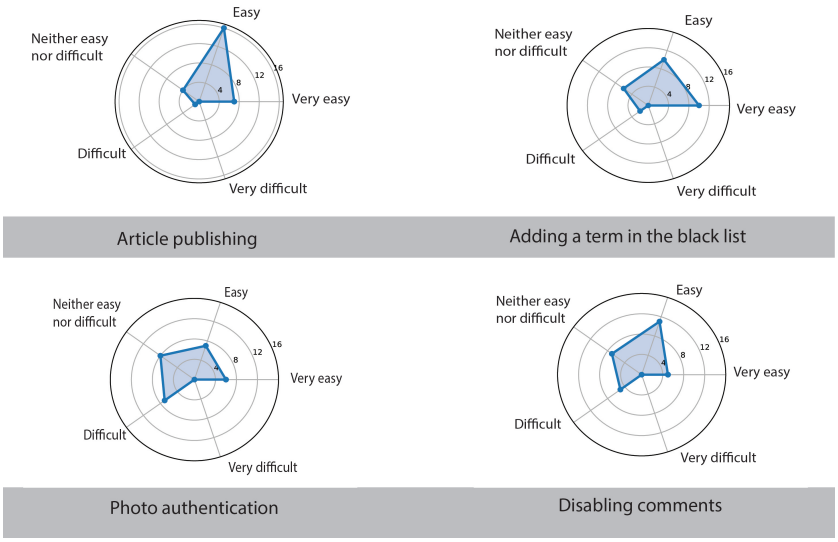

Figure 10. Evaluation of the ease of use of different actions (supervising journalists).

When the participants were asked to evaluate the usefulness of the previously mentioned actions, they rated all of them very positively. Namely, text correction and article publishing were acknowledged as useful and very useful actions, whereas the deletion of malicious content and disabling of comments had a slightly higher number of participants evaluating them as less useful. Figure 11 illustrates this subtle difference in an indicative manner.

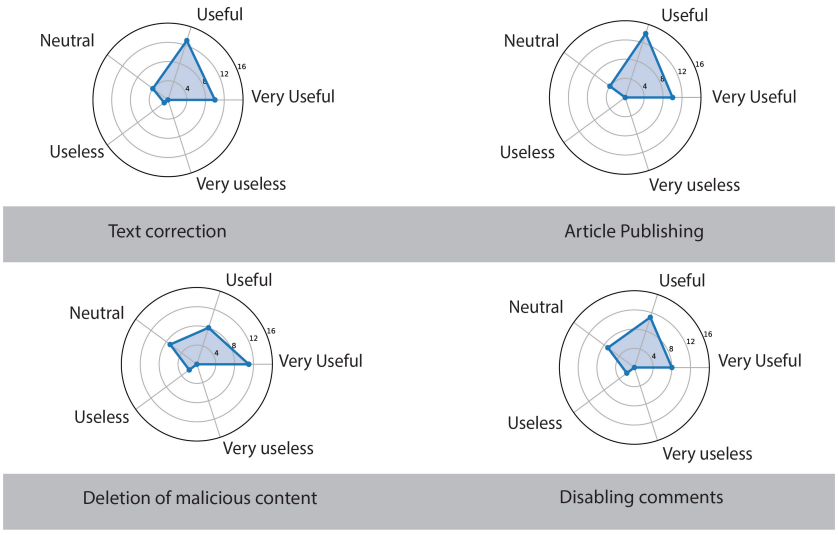

Figure 11. Evaluation of the usefulness of certain features (supervising journalists).

Overall, the functionality of the navigation and the ease of use of the platform were evaluated with 4 and 3.9, respectively (with 5 being the highest score). Based on the above findings, we can conclude that the features implemented in SEMPRATO were found to be useful for the supervising journalists. The same results were found in the ease of use with only one exception regarding photo authentication for which the software development team had some platform limitations in making

this feature easy to use. Finally, the participants were asked to rate from 1 to 5 (1 the lowest and 5 the highest) how likely would be to recommend SEMPRATO to a fellow journalist which received an average rating of 3.8, and how likely would be to use SEMPRATO if it would be made generally available which received an average rating of 3.9.

Overall, the results from both scenarios were quite positive, if we consider the fact that the platform is at a prototype stage offering a basic set of features and the user interface is not as refined as the user interface of a professional web application. Note that 42.9% of the participants stated that they had previous experience working in a collaborative journalism platform and that indicates that the SEMPRATO platform would be able to compete with other available systems.

7. Conclusions and Future Extensions

This paper has studied the design, implementation and evaluation of SEMPRATO, a platform enabling users to manage UGC, using intelligent semantic web tools and analysis strategies. Based on the findings of the study, a number of conclusions can be drawn.

The difficulties of dealing with large volumes of data and the need to maintain high journalistic quality standards, led media organisations to explore and incorporate semantic technologies in their workflows. After exploring the current media landscape, this trend was identified and pursued in this work. From the initial theoretical analysis, it was concluded that semantic web technologies can be employed in various ways and will become a critical component in participatory journalism platforms. Additionally, through the assessment of a small number of use cases around UGC management, it was concluded that the introduction of a new upgraded model of participatory journalism was required.

The thorough theoretical investigation set the foundations for the implementation of a prototype participatory journalism environment. The assessment of available semantic services, applications and tools in the context of journalism, along with the conducted surveys mentioned in the design chapter, defined the structure of the proposed model. To evaluate the usefulness of the introduced model a workshop was organised where participants were asked to use and evaluate SEMPRATO under two different scenarios, as contributing users and as journalists.

Based on the results of the evaluation, it can be concluded that a technologically-augmented journalistic environment can provide a high level of automation and improved functionality in everyday workflows that journalists find beneficial. Its orientation towards practicality and ease of use is a critical factor for the adoption of these new processes by media organisations. The impact of SEMPRATO's framework is bilateral. On the one hand, it is located in the participatory journalism context, where the introduction of technological novelties generates new information exploitation methods for journalists. One the other hand, it is located in the media market context, where the leverage of information systems and services opens up new business opportunities. The proposed environment facilitates dynamic and flexible ways of collecting and analysing UGC. However the prioritisation of features provided in such an environment should be mainly driven by the journalist's needs. Overall the practical application of semantic technologies in the field of journalism will highlight their advantages, leading the way for broader adoption.

Future extensions of this work could include further research in the adoption of semantic technologies by global media organisations, with emphasis on user-friendly workflows. Moreover, the incorporation and evaluation of additional features (e.g., text-based fake-news detection) in SEMPRATO would be useful for the extraction of additional insights using a similar, workshop-based assessment methodology.

Author Contributions: Conceptualization, A.V.; Data curation, N.T.; Formal analysis, K.P.; Funding acquisition, A.V.; Investigation, K.P., N.T. and T.S.; Methodology, N.T. and T.S.; Project administration, A.V.; Resources, A.V.; Supervision, N.T. and A.V.; Validation, K.P., N.T. and T.S.; Visualization, K.P.; Writing—original draft, K.P., N.T. and T.S.; Writing—review and editing, K.P., N.T., T.S. and A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in the context of the project "Semantic services and models of participatory journalism management" (MIS 55184) under the call for proposals "Supporting researchers with emphasis on new researchers" (EDULLL 34). The project is cofinanced by Greece and the European Union (European Social Fund-ESF) by the Operational Programme Human Resources Development, Education and Lifelong Learning 2014-2020

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

References

Abbott, Jennifer Y. 2017. Tensions in the scholarship on participatory journalism and citizen journalism. *Annals of the International Communication Association* 41: 278–97. [CrossRef]

Aitamurto, Tanja. 2013. Balancing between open and closed: Co-creation in magazine journalism. *Digital Journalism* 1: 229–51. [CrossRef]

Amazon. 2019. Amazon rekognition. Available online: https://aws.amazon.com/rekognition (accessed on 24 August 2019).

Anderson, Paul. 2007. *What is Web 2.0?: Ideas, Technologies and Implications for Education*. Bristol: JISC, vol. 1.

Anguelov, Dragomir, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. 2010. Google street view: Capturing the world at street level. *Computer* 43: 32–38. [CrossRef]

Antoniou, Grigoris, and Frank Van Harmelen. 2004. *A Semantic Web Primer*. Cambridge: MIT Press.

Bentley, Frank, Katie Quehl, Jordan Wirfs-Brock, and Melissa Bica. 2019. Understanding online news behaviors. Paper presented at the CHI Conference on Human Factors in Computing Systems, Glasgow, UK, May 4–9; p. 590.

Bernstein, David. 2014. Containers and cloud: From lxc to docker to kubernetes. *IEEE Cloud Computing* 1: 81–84. [CrossRef]

Boberg, Svenja, Tim Schatto-Eckrodt, Lena Frischlich, and Thorsten Quandt. 2018. The moral gatekeeper? moderation and deletion of user-generated content in a leading news forum. *Media and Communication* 6: 58–69. [CrossRef]

Boettiger, Carl. 2015. An introduction to docker for reproducible research. *ACM SIGOPS Operating Systems Review* 49: 71–79. [CrossRef]

Bradshaw, Paul. 2013. *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*. Abingdon: Routledge.

Brandtzaeg, Petter Bae, Marika Lüders, Jochen Spangenberg, Linda Rath-Wiggins, and Asbjørn Følstad. 2016. Emerging journalistic verification practices concerning social media. *Journalism Practice* 10: 323–42. [CrossRef]

Choudhury, Nupur. 2014. World wide web and its journey from web 1.0 to web 4.0. *International Journal of Computer Science and Information Technologies* 5: 8096–100.

Coe, Kevin, Kate Kenski, and Stephen A Rains. 2014. Online and uncivil? patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64: 658–79. [CrossRef]

Deuze, Mark. 2003. The web and its journalisms: Considering the consequences of different types of newsmedia online. *New Media & Society* 5: 203–30.

FabulaAI. 2019. Available online: https://www.fabula.ai/ (accessed on 24 August 2019).

Factmata. 2019. Available online: https://factmata.com (accessed on 24 August 2019).

Franquet, Rosa, Maria Isabel Villa, and Ignacio Bergillos. 2011. Audience participation in online news websites: A comparative analysis. *Observatorio (OBS*)* 5. [CrossRef]

Frischlich, Lena, Svenja Boberg, and Thorsten Quandt. 2018. Comment sections as targets of dark participation? journalists' evaluation and moderation of deviant user comments. *Journalism Studies* 20: 2014–33. [CrossRef]

Geary, Joanna. 2013. Introducing Guardianwitness, Our New Platform for Content You've Created. Available online: https://www.theguardian.com/help/insideguardian/2013/apr/16/introducing-guardianwitness-platform-content-youve-created (accessed on 24 August 2019).

Gershgorn, Dave. 2017. Audio Transcription Is Finally Being Automated. Available online: https://qz.com/931350/trint-uses-artificial-intelligence-tohelp-transcribe-your-hours-and-hours-of-interviews/ (accessed on 24 August 2019).

Gillmor, Dan. 2009. Introduction: Toward a (new) media literacy in a media saturated world. In *Journalism and Citizenship*. Abingdon: Routledge, pp. 19–30.

Greene, Derek, and Pádraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. Paper present at the 23rd International Conference on Machine Learnin, Pittsburgh, PA, USA, June 25–29; pp. 377–84.

Gross, David. 2014. Big Data and the Semantic Web—What Does It Really Mean? Available online: https://www.express.co.uk/life-style/science-technology/529574/big-data-semantic-web-what-it-means (accessed on 24 August 2019).

Hallin, Daniel C., and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Harrison, Jackie. 2010. User-generated content and gatekeeping at the bbc hub. *Journalism Studies* 11: 243–56. [CrossRef]

Hawkes, Rebecca. 2017. Ap Embarks on Content Tie-Up With Tokyo'S Spectee. Available online: https://www.rapidtvnews.com/2017060747513/ap-embarks-on-content-tie-up-with-tokyo-s-spectee.html (accessed on 24 August 2019).

Hellmueller, Lea, and You Li. 2015. Contest over content: A longitudinal study of the cnn ireport effect on the journalistic field. *Journalism Practice* 9: 617–33. [CrossRef]

Hendler, James, Ora Lassila, and Tim Berners-Lee. 2001. The Semantic Web. *Scientific American* 284: 34–43.

Heravi, Bahareh Rahmanzadeh, Marie Boran, and John Breslin. 2012. Towards social semantic journalism. Paper presented at the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, June 4–7.

Hermida, Alfred, Fred Fletcher, Darryl Korell, and Donna Logan. 2012. Share, like, recommend: Decoding the social media news consumer. *Journalism Studies* 13: 815–24. [CrossRef]

Hille, Sanne, and Piet Bakker. 2014. Engaging the social news user: Comments on news sites and facebook. *Journalism Practice* 8: 563–72. [CrossRef]

Kavitha, C. R., and Sunitha Mary Thomas. 2011. Requirement gathering for small projects using agile methods. *NCCSE IJCA* 122–28.

Koukis, Vangelis, Constantinos Venetsanopoulos, and Nectarios Koziris. 2013. okeanos: Building a cloud, cluster by cluster. *IEEE Internet Computing* 17: 67–71. [CrossRef]

Loper, Edward, and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv* arXiv:cs/0205028.

Maniou, Theodora, and Andreas Veglis. 2016. 'selfie journalism': Current practices in digital media. *Studies in Media and Communication* 4: 111–18.

Maniou, Theodora, Kosmas Panagiotidis, and Andreas Veglis. 2017. The politicization of selfie journalism: An empirical study to parliamentary elections. *International Journal of E-Politics (IJEP)* 8: 1–16. [CrossRef]

Manosevitch, Idit. 2011. User generated content in the israeli online journalism landscape. *Israel Affairs* 17: 422–44. [CrossRef]

Michailidou, Asimina, Pieter De Wilde, and Hans Jörg Trenz. 2014. *The Internet and European Integration: Pro-and Anti-EU Debates in Online News Media*. Leverkusen: Verlag Barbara Budrich.

Panagiotidis, Kosmas, Nikolaos Tsipas, Theodora Saridou, and Andreas Veglis. 2019. Semantic web services and applications in journalism. Paper Presented at 5th Annual International Conference on Communication and Management (ICCM2019), Athens, Greece, April 15–18.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and et al. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12: 2825–30.

Pomp, André, Alexander Paulus, Andreas Kirmse, Vadim Kraus, and Tobias Meisen. 2018. Applying semantics to reduce the time to analytics within complex heterogeneous infrastructures. *Technologies* 6: 86. [CrossRef]

Quandt, Thorsten. 2018. Dark participation. *Media and Communication* 6: 36–48. [CrossRef]

Saiedian, Hossein, and Robert Dale. 2000. Requirements engineering: Making the connection between the software developer and customer. *Information and Software Technology* 42: 419–28. [CrossRef]

Santana, Arthur D. 2014. Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice* 8: 18–33. [CrossRef]

Saridou, Theodora, Kosmas Panagiotidis, Nikolaos Tsipas, and Andreas Veglis. 2018. Semantic tools for participatory journalism. *Journal of Media Critiques [JMC]* 4. [CrossRef]

Saridou, Theodora, Nikolaos Tsipas, Kosmas Panagiotidis, and Andreas Veglis. 2019. Towards a semantic-oriented model of participatory journalism management: Perceptions of user-generated content. Paper presented at the 5th Annual International Conference on Communication and Management (ICCM2019), Athens, Greece, April 15–18.

Saridou, Theodora, and Andreas Veglis. 2016. Participatory journalism policies in newspapers' websites in greece. *Journal of Greek Media & Culture* 2: 85–101.

Shadbolt, Nigel, Tim Berners-Lee, and Wendy Hall. 2006. The semantic web revisited. *IEEE Intelligent Systems* 21: 96–101. [CrossRef]

Shanley, Paul. 2017. Producing a Tool That Will Automatically Verify Ugc. Available online: https://insights.ap.org/whats-new/producing-a-tool-that-will-automatically-verify-ugc (accessed on 24 August 2019).

Sheffield, Hazel. 2012. Huffpost Launches a Platform for User-Generated Reporting. Available online: https://archives.cjr.org/behind_the_news/huffpost_launches_a_platform_f.php (accessed on 24 August 2019).

Siapera, Eugenia, and Lia-Paschalia Spyridou. 2012. The field of online journalism: A bourdieusian analysis1. In *The Handbook of Global online Journalism*. Hoboken: John Wiley & Sons, pp. 77–97.

Singer, Jane B., David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich, and Marina Vujnovic. 2011. *Participatory Journalism: Guarding Open Gates at Online Newspapers*. Hoboken: John Wiley & Sons.

Sivakorn, Suphannee, Jason Polakis, and Angelos D Keromytis. 2016. I'm not a human: Breaking the google recaptcha. Paper presented at the Black Hat ASIA 2016, Singapore, 29 March–1 April.

Spettel, Stefan, and Dimitrios Vagianos. 2019. Twitter analyzer—How to use semantic analysis to retrieve an atmospheric image around political topics in twitter. *Big Data and Cognitive Computing* 3: 38. [CrossRef]

Spyridou, Lia-Paschalia. 2019. Analyzing the active audience: Reluctant, reactive, fearful, or lazy? forms and motives of participation in mainstream journalism. *Journalism* 20: 827–47. [CrossRef]

Spyridou, Paschalia-Lia, and Andreas Veglis. 2008. Exploring structural interactivity in online newspapers: A look at the greek web landscape. *First Monday* 13. [CrossRef]

Suau, Jaume, and Pere Masip. 2014. Exploring participatory journalism in mediterranean countries: Political systems and national differences. *Journalism Practice* 8: 670–87. [CrossRef]

TwitterAudit. 2019. Available online: https://www.twitteraudit.com/ (accessed on 24 August 2019).

Underwood, C. 2019. Automated Journalism–Ai Applications at New York Times, Reuters, and Other Media Giants. Available online: https://emerj.com/ai-sector-overviews/automated-journalism-applications/ (accessed on 24 August 2019).

Veglis, Andreas. 2013. Education of journalists on icts: Issues and opportunities. *Journal of Applied Journalism & Media Studies* 2: 265–79.

Veglis, Andreas. 2014. Moderation techniques for social media content. In SCSM 2014 Lecture Notes in Computer Science. Cham: Springer, vol. 8531.

Wilks-Harper, Ella. 2018. How Is User-Generated Content Transforming Storytelling? Available online: https://www.reuterscommunity.com/topics/newsroom-of-the-future/how-is-user-generated-content-transforming-storytelling/ (accessed on 24 August 2019).