Towards a semantic-oriented model of participatory journalism management: Perceptions of user-generated content

Theodora Saridou
Kosmas Panagiotidis
Nikolaos Tsipas
Andreas Veglis

Abstract
During the last two decades, citizens’ participation in news production process has raised great academic and entrepreneurial interest for participatory journalism. Traditional procedures and concepts such as gatekeeping have been under discussion. News organizations redesign their websites in order to adopt tools and applications that make it possible for users to be active consumers or even co-producers of journalistic content, by liking, sharing, commenting and submitting material. At the same time, huge amounts of user-generated content are uploaded every minute on social media platforms. Subsequently, professionals have to deal with continually available information which requires management, classification and evaluation in order to keep high journalistic standards and to avoid problems, varying from plain grammar mistakes to serious situations of fake news, hostility or hate speech. Thus, there is the obvious need for a new model of managing participatory journalism, based on semantic technologies, which will support organized collection and moderation of content in an effective way and in short time. The main objective of this paper is to define the requirements and describe the characteristics that the model should have. For this purpose, two online surveys of journalists and users were conducted in Greece, in order to gain some insights concerning the development of the model. The paper presents the key findings from the surveys and identifies the views, the preferences and the experiences as expressed by the respondents, which lead to the tendency towards a collaborative, semantic-oriented way of submitting and receiving user-generated content.

Keywords: Participation, Semantics, User-generated content, Journalists, Model

1. Introduction
Although the idea of mediated participation is part of the long-standing debate between cyber-optimists and cyber-pessimists (Spyridou, 2018), it is accepted that the current new media era has seen a proliferation of user-generated content (UGC) and an increasing participation by Internet users in public communication, impelling newsrooms to integrate non-professional contributions into ordinary news practice (Tong, 2015). In the light of changing roles, users become prosumers

1PhD student, Aristotle University of Thessaloniki, Greece saridout@jour.auth.gr
2PhD student, Aristotle University of Thessaloniki, Greece kpanagioti@jour.auth.gr
3Post-doc researcher, Aristotle University of Thessaloniki, Greece nitsipas@auth.gr
4Professor, Aristotle University of Thessaloniki, Greece veglis@jour.auth.gr
(Deuze, 2003), while gatekeeping turns to gatewatching, as journalists do not define, but observe the information flow stemming from a variety of sources, receive UGC and construct a final product (Bruns, 2008). In a similar vein, Boczkowski (2005) refers to gateopening, when users actively co-develop website content and traditional work routines are questioned. Media organizations employ on their websites participatory formats that allow readers to actively consume or co-produce content. Classification of such formats has not led to a unique typology, since research has -among others- focused on the stages of news production (Singer et al., 2011), on the different propositions of communication process (Franquet et al., 2011) or on the concept of interactivity (Suau & Masip, 2014). Identifying the most prevalent and widespread participatory tools, Spyridou (2018) concluded to the following nine: 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.

Apart from the UGC that is produced on news websites, traditional news gathering process is also influenced by social media platforms, where users produce a wealth of data in the form of text, images and video, which must be processed, compiled and verified by journalists within a very short time span before being incorporated into a news story (Heravi et al., 2012). Journalists use social media to look for breaking news events, find ideas for stories, keep in touch with their audience and collect information (Weaver & Willnat, 2016, p. 853).

As a result, professionals are often challenged by a vast amount of content that has to be handled in tandem with other daily tasks (Boberg et al., 2018). Moreover, digital content can be processed, intentionally altered or falsified and redistributed relatively easy, posing threats to mass communication and journalistic processes that can be associated to unwanted content tampering, construction of fake evidences, sharing and propagation of untrue stories (Katsaounidou & Dimoulas, 2018). Journalists therefore face the challenge of carefully disentangling valid opinions from manipulated statements, with the obvious danger of either censoring genuine speech or letting noxious forms of user engagement slip through (Frischlich et al., 2019). As Quandt (2018) argues, the potential for dark participation is enormous, not only in comment sections controlled by the media themselves, but also on non-proprietary platforms, like Facebook or Twitter, where the negativity and toxic atmosphere can be present and the deliberative quality very low.

Subsequently, it is necessary for journalists to establish consistent and transparent moderation strategies (Boberg et al., 2018), to develop their skills and use digital tools in the service of tracing information, forensic examination, UGC dissemination and verification (Johnston, 2016; Veglis, 2013). Since the Web evolution was built on structures which proved weak and unable to deal with the burden of managing huge amounts of information (Panagiotidis & Veglis, 2015), the application of automated, semi-automated and manual content annotation processes can significantly improve the management and retrieval of content at scale by enabling advanced categorization and search mechanisms (Lew et al., 2006; Saridou et al., 2018).

The main objective of this paper is to define the requirements and describe the characteristics that a new model of participatory journalism management based on semantic technologies should have in order to support collection and moderation of UGC effectively and in short time. For this purpose, two online surveys of journalists and users were conducted, in order to gain useful insights concerning the development of the model.

**Difficulties in managing UGC in journalism**
Although the concept of participatory journalism was initially accompanied by promises for democratization and community enhancement (Bowman & Willis, 2003; Gillmor, 2004), it is accepted that the incorporation of UGC in the journalistic process is not an easy task (Veglis, 2013), due to significant ethical, legal and responsibility issues (Saridou & Veglis, 2016, p. 88). Users’ contributions have raised questions about the actual practice and its implications for culture, society and politics (Manosevitch, 2011), while journalists often see comments as a necessary evil, which holds intolerable standards of expression and poses problems stemming from low quality (Singer et al., 2011). Specifically, many professionals underline the excessive use of inappropriate language, flaming, stereotyping and superficial discourse that impede constructive public discourse (Manosevitch, 2011, p. 425). Research also shows that incivility is a common feature of public discussions. Defining the term, Coe et al. (2014) refer to features of discussion that convey an unnecessarily disrespectful tone towards the discussion forum, its participants or its topics, mainly in the forms of name-calling, aspersion, lying, vulgarity and pejorative for speech. Incivility is not only associated with key contextual factors, such as the topic of the article and the sources quoted within it, but also with the frequency users comment. For example, “hard news” topics about politics or economy generate a higher percentage of uncivil commentary in contrast to articles about health or lifestyle and at the same time frequent commenters are more civil than occasional ones (Coe et al., 2014, p. 669).

Furthermore, the examples that could be labelled as dark participation range from misinformation and hate campaigns to individual trolling and cyberbullying (Quandt, 2018), while the spreading of fake news, disinformation and conspiracy theories in UGC are forms of deviances as well (Frischlich et al., 2019). In addition, reports about malicious “pseudo-users”, troll armies and social bots have jarred confidence in even the fundamental assumption that user interaction is interaction with users (Frischlich et al., 2019, p. 3). Issues about UGC lie also in the field of responsibility as traditional legal liability models are questioned in the light of changing author, editor and publisher’s role (Valcke & Lenaerts, 2010). When users produce content on a professional website, the debate on liability concerns them, the journalist, the website owner and the Internet service provider as well.

Trying to ensure the quality of their job and to avoid problems like the abovementioned, professionals have to manage and control users’ content. When pre-moderation is used, every piece of UGC is checked before publication and high security is achieved. However, this method is labor-intensive and costly (Santana, 2014; Singer et al., 2011). On the other hand, post-moderation policies lead to simpler and more open comment systems, but can lower the quality (Hille & Baker, 2014).

When distributed moderation is chosen, users are involved in the control process, often earning some kind of privileges (Lampe & Resnick, 2004). News media can let users themselves moderate content by using buttons to rank comments or to report abuse (Hille & Baker, 2014, p. 565). All strategies can be accompanied by purely technical methods such as automated moderation under predefined filters that detect and replace banned words, phrases, IP addresses and paragraphs (Veglis, 2014) or Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHA), which are employed in the websites more rarely (Saridou & Veglis, 2016), aiming to block computer-derived mass entries. The reCAPTCHA service offered by Google is the most widely used CAPTCHA service and has been adopted by many websites for preventing automated bots from conducting nefarious activities (Sivakorn et al., 2016). Except for concerns about human, financial and time resources, the individual moderation decision is affected by newsroom routines,
media organizations for which journalists’ work, the societal institutions and social system in which they operate, their personal experiences or even gut feelings (Boberg et al., 2018, p. 60).

**Semantic Web and tools in journalism**

Aiming at the exploitation of UGC, many media organizations have built platforms, such as CNN iReport, BBC’s UGC Hub and Guardian Witness, where users can submit and journalists can elaborate content. During the past few years, news media outlets have also started using artificial intelligence technology in new ways, from speeding up research to accumulating and cross-referencing data (Underwood, 2019). The New York Times, for example, has implemented artificial intelligence to moderate reader comments, encourage constructive discussion and eliminate harassment and abuse. Comments are organized interactively so that viewers can quickly see which can be considered ‘toxic’ and which may be more illuminating, by sliding a bar across the top of the page. The closer the bar gets to the right, the more toxic the comments become (Underwood, 2019).

A coarse classification of the tools and services used by media organizations in UGC-driven platforms can lead to the categories of content discovery and monitoring in social media, distributed multimedia content capturing, semantic analysis of multimedia content, audience involvement in news making, UGC verification and UGC sharing (Saridou et al., 2018). It is, however, noticed that the aforementioned examples focus on building heterogeneous platforms and rely in most cases on the integration with services provided by third parties, while professional journalists are still involved in many stages of the process (Saridou et al., 2018, p. 286).

Semantic Web technologies can help the evolution of journalism due to their aggregation features, regarding the advanced ability of collecting information (Heravi et al., 2012). Through the utilization of new tools and methods, journalists can search, find and process any piece of information more quickly and more easily, thanks to the exploitation of structured data (Gray et al., 2012). Nevertheless, such semantic techniques have not been widely adopted by media organizations (Brandtzaeg et al., 2016). One of the main reasons is the immature semantic technology and the lack of a unified platform in which journalists can concentrate in the utilization of UGC. Thus, a semantic-oriented model of participatory journalism management can support organized collection and moderation of content in an effective way and in short time.

**2. Methodology**

In order to define the requirements and describe the characteristics that such a model should have, the following research questions were posed:

RQ1: How do journalists manage UGC in their work routine?
RQ2a: How do journalists perceive the importance of education in UGC management?
RQ2b: How do journalists perceive users’ role in a semantic-oriented participatory journalism management model?
RQ3: How do users participate in content production in media organizations?
RQ4: How do users perceive their role in a semantic-oriented participatory journalism management model?

In order to answer these questions, two online surveys of journalists and Internet users were conducted. The first one was conducted among journalists of the Journalists’ Union of Macedonia-Thrace Daily Newspapers, of the Journalists’ Union of Thessaly-StereaEllada-Evia Daily Newspapers and among professionals of online media (magazines, newspapers and news portals). A questionnaire consisting of 22 questions was distributed to the potential participants via email.
total, 52 people responded. Of these, 51.1 percent were female and 48.9 percent were male. The majority (48.9 percent) was between 31 and 40 years old and 35.6 percent belonged to the 41 to 50 age group. The second survey was conducted among undergraduate, postgraduate and PhD students of the School of Journalism and Mass Media Communication of the Aristotle University of Thessaloniki. A hyperlink which led to a questionnaire consisting of 11 questions was uploaded on the relevant Facebook group. Overall, 52 users responded.

The data from both surveys were collected in a one-month period, between October, 9 2018 and November 9, 2018, using Google Forms. All responses were anonymous. An introductory text informed the participants about the anonymity, the use of the obtained data and the scope of the survey.

3. Data analysis

In order to analyze how they manage UGC in their everyday work routine (RQ1), journalists were firstly asked whether they had been part of a journalistic team that managed UGC. More than half (53.8 percent) of the participants gave a negative response, with the rest responding positively. However, the rates were reversed when participants were asked whether they had managed UGC as part of their tasks on their own (not as part of a team), since 53.8 percent responded positively. When asked which participatory tools are mostly preferred by users, journalists highlighted tools that allow sharing to social networking sites, tools for content rating and comment sections. On the contrary, participants considered that discussion forums, citizen blogs and tools for submission of textual and audio-visual material are much less or hardly ever used by audience members.

Regarding comment management systems used by media organizations, Facebook by far (86.5 percent) and services like the Disqus comment system (25 percent) are the most used, according to journalists.

Focusing on the problems that occur in participatory contributions, respondents identified spelling and syntactical mistakes along with intellectual property violations as the most frequently encountered problems, while fake news, personality insults, hate speech, defamation, spamming and trolling are very frequently faced as well. Asked how they handle UGC, half of the respondents said that they do not perform any kind of classification. Furthermore, 59.6 percent of the journalists stated that they face problems with quality management and exploitation of UGC, mainly due to high workload and different content types (e.g. text, photo, and video). Quite often problems were also ascribed to low content standardization (e.g. various image file formats) and mass production.

Afterwards participants were asked which content moderation method they use and indicated pre-moderation as the most employed (57.7 percent). This was followed by post-moderation, moderation based on users’ reactions and automated moderation. It is, however, noticeable that 15.4 percent of the participants do not moderate at all (Figure 1). When asked which method they use to verify UGC authenticity, more than half of the respondents (57.7 percent) mentioned user registration, 42.3 percent mentioned check of multimedia content, 34.6 percent check of publication time, 26.9 CAPTCHA and 21.2 check of the location.
Trying to identify journalists’ perceptions of the role of education in UGC management (RQ2a), it was obvious that most respondents (51.9 percent) are not satisfied with the functionality of the UGC management tools they use, with a large group of respondents (88.5 percent) considering that journalists do not use the best management and control practices. In a similar context, the overwhelming majority (96.2 percent) believes in the need for further journalists’ education in this field, while a vast majority (90.4 percent) would spend time to get trained as well.

RQ2b tried to identify how journalists perceive users’ role in a semantic-oriented participatory journalism management model. Among respondents, 80.7 percent believe that users’ engagement in UGC management processes would be very or quite effective and 17.3 percent believe that users’ engagement would have little effectiveness. At the same time, only 1.9 percent of the journalists see no effectiveness at all. More than half of the respondents (61.5 percent) find that cooperation between citizens and journalists would work more efficiently if it took place in the website of the media organization, while the rest consider an autonomous platform outside the main organization as a better option. Although most of the respondents seem willing to cooperate with users in UGC management, 63.5 percent of them are not aware of environments that support the collaboration between citizens and journalists in the submission, management and publication of UGC. It is finally clear that journalists (92.3 percent) believe in the need for creation of an integrated UGC management tool with specific features, such as collection, word processing, classification and content verification.

RQ3 was related to the way users participate in content production in media organizations. As indicated by the participants of the second survey, they produce content mainly for business and less for entertainment or social interest reasons. Photos are the type of content mostly submitted by users (75 percent), with text (69.2 percent) and video (30.8 percent) following (Figure 2). When asked through which participatory tools they submit content, users -in line with the journalists’ responses in the first survey- indicated tools that allow sharing to social networks, tools for content rating and comment sections. On the contrary, as journalists noticed too, users do not use discussion forums, citizen blogs and tools for submission of textual and audio-visual material. Among participants, 69.2 percent stated satisfaction with the functionality of the content submission services provided by news websites.
As asked about their possible role in a participatory journalism management model (RQ4), 86.5 percent of the users claimed that they have never been invited by a news medium to manage content posted by other users. Nevertheless, 94.2 percent believe that their participation in UGC management in a website would be very or quite effective and 84.6 percent would like services that support collaboration between citizens and journalists in a participatory context. Specifically, more than two-thirds (71.2 percent) of the respondents would prefer such an environment to work in the website of the media organization and not in a separate, autonomous platform. It is also clear that users (82.7 percent) believe in the need for creation of an integrated UGC management tool with specific features, such as collection, word processing, classification and content verification. Asked, finally, how they would choose to access a content submission service, the most popular ways seem to be using a web browser from a desktop/laptop (63.5 percent) and using an application from a mobile phone (51.9 percent).

**Discussion**

The findings of the first survey indicate that half of the journalists do not classify the content submitted by users, while the biggest obstacles to management and exploitation mainly derive from heavy workload, high content heterogeneity and a small degree of content standardization. While managing users’ contributions, professionals detect problems that not only diminish quality, such as grammar mistakes, but also may give rise to legal consequences, such as offenses of personality and hate speech.

The second survey, on the other hand, shows that users submit photos, texts and videos, mostly for business purposes. When participating, they share content to social networking sites, rank news items or comment on published stories, but they do not create their own content. Their participation is mainly limited to interaction with already produced, professional content. This remark ties in with the findings of previous studies about participation in the news production process, which indicate that media organizations are reluctant to allow readers to set the news agenda or take part in the gatekeeping process, while there are only a few options for citizen stories, always subject to strict editorial control (Singer et al., 2011).
As a key result, both surveys show a strong agreement with a new, collaborative way of managing UGC in a semantic context, as user involvement in such a model is deemed effective both from journalists and from users themselves. It is obvious that a new participatory journalism model should be 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, in order to serve verification purposes it should support features for data export related to the place and time that a photo or video was taken. It would preferably be embedded in the organization's website, designed for working both through computer browsers and mobile applications. Additionally, it should have some characteristics inspired by social networking platforms, such as ‘like’ and ‘comments’, since not only users, but also journalists seem familiar with the relative social media functions (Table 1).

With professionals, finally, declaring willingness to spend time for their training on UGC platforms, model could be presented to journalists and evaluated by them for further improvements and adjustments.

<table>
<thead>
<tr>
<th>Relationship journalists - users</th>
</tr>
</thead>
<tbody>
<tr>
<td>collaborative</td>
</tr>
</tbody>
</table>

**Structure – Design**
- fast to use
- part of the workflow
- on the organization’s website
- computer browsers and mobile applications
- characteristics from social networking platforms

**Moderation**
- moderation by journalists
- problem flagging by users

**Available services**
- automatic categorization of UGC
- detection of malicious content
- data export for multimedia content

However, limitations of the study should be noted. First, only Greek participants were surveyed and the results may not be valid for other countries. Moreover, the second survey was conducted among students of journalism and communication, who are assumed to have more knowledge about news media than the general public. Therefore, a more varied sample may had generated different findings.

4. Conclusion
Huge amounts of constantly produced user-generated information necessitate a thorough handling by professionals. This study examined perceptions of UGC in order to define the requirements and characteristics that a semantic-oriented model of participatory journalism management should have. Journalists and users agree that cooperation can lead to a more efficient way of managing
participation in news production process. In order to design the model, we should take into account the needs of the directly involved parties. Specifically, journalistic working reality, difficulties in UGC exploitation and users’ activity should constitute the basic framework. Additional studies can include qualitative interviews to gain deeper understanding of the journalists’ attitude towards UGC. Further, it would be worth focusing on the modelling and designing of such a system. Future extension of this work can also focus on the actual implementation of the features and the evaluation of the model.

5. Acknowledgments
This research is carried out/funded in the context of the project “Semantic services and models of participatory journalism management” (MIS 5005184) under the call for proposals “Supporting researchers with emphasis on new researchers” (EDULLL 34). The project is co-financed by Greece and the European Union (European Social Fund- ESF) by the Operational Programme Human Resources Development, Education and Lifelong Learning 2014-2020.

6. References


