Article

Transitions in Journalism—Toward a Semantic-Oriented Technological Framework

Kosmas Panagiotidis and **Andreas Veglis** *

Media Informatics Lab, School of Journalism & Mass Communication, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; kpanagioti@jour.auth.gr
* Correspondence: veglis@jour.auth.gr

check for updates

Abstract: The technologies behind today's web services, tools, and applications are evolving continually. As a result, the workflows and methods of different business sectors are undergoing constant change. The news industry and journalism are heavily affected by these changes. New technological means for practicing journalism and producing news items are being incorporated in media workflows, challenging well established journalistic norms and practices. The perpetual technological evolution of the web creates a wide range of opportunities. For this reason, both technology companies and media organizations have begun to experiment with semantic web technologies. Our focus in this paper was to discover and define the ways that semantic technologies can contribute to the technological upgrade of everyday journalism. From this perspective, we introduced the term 'semantic journalism' and attempted to investigate the transition of journalism to a semantic-oriented technological framework.

Keywords: journalism; semantic web; semantic technologies; semantic journalism; semantic practices; media industry; news organizations

1. Introduction

The introduction of the Internet and the exploitation of its services resulted in a variety of consequences for the way the news industry operates (Deuze 2003) and the way journalism is produced (Van der Haak et al. 2012). The emergence of new tools and practices, along with the digitalization of the work process, provided new ways to produce information and a redefinition of professional journalism (Deuze and Witschge 2018). Currently, the audience has direct access to multiple digital sources, ranging from mainstream media to individual blogs and digital social platforms, from which they can create and distribute a wide range of information (Veglis and Maniou 2018). Journalists are expected to operate multiple services and platforms in order to be constantly informed of what is happening and deliver the news (Spyridou et al. 2013). In addition, the rapidly growing tendency of media organizations to use advanced web technologies in order to manage the everyday workload of news production and publication has attracted the interest of many academic researchers (Maass and Kowatsch 2012; Osman et al. 2014; Caswell 2019). These researchers conceptualized the transformation of journalism and explained the drastic changes under the rubric of convergence.

In Newman's recent survey (Newman 2019), *Digital News Report: Journalism, Media and Technology Trends and Predictions 2019*, the intentions of 200 publishers from 29 countries to invest in harnessing the potential of artificial intelligence (AI) and machine learning (ML) were clearly captured. Other studies in the field presented evidence that journalists understand the merits and the necessity to enhance their practices by exploiting new technological capabilities and foster innovation in the newsroom (Paulussen 2016). The emergence of more and more technologically advanced types of journalism leaves no space for doubts regarding the technological optimization of today's journalism.

Analyzed from this perspective, many journalistic practices are considered newborn due to their technological features. Namely, robot journalism (Latar 2018), immersive journalism (Rodríguez 2018), data journalism (Veglis and Bratsas 2017), multimedia journalism (Bull 2010), mobile journalism (known as mojo (Richardson 2012), drone journalism (Dalakas et al. 2017) and backpack journalism (Kennedy 2010)).

Using these findings as a starting point, this article suggests that there is a burgeoning need to further investigate how advanced web technologies and services like the semantic web (SW) influence journalism and media studies. Many scholars have attempted to present semantic-oriented solutions to intertemporal problems experienced by various companies, including media organisations, for example information retrieval data integration, prominence of news material, discovery, collection and verification of content, data management and many more (Janev and Vranes 2009). Although this information gives us important guidelines to understand the SW and its implications in journalism as well as insights for better use, this information has not been categorized, described and placed in a context that showcases journalism in the SW environment. For this reason, this study's principal purpose is to introduce the new-born practice of semantic journalism (SJ), which derives from the development of the semantic web (SW) and the embedment of semantic technologies in everyday journalism practices. Secondarily, within our intentions is to resolve all possible doubts regarding SJ by presenting all the relevant documentation and arguing that is already happening and evolving behind the scenes, fragmented and marginalized in an uncharted field. Overall, through the examination of various semantic-based solutions for journalism-related tasks, this research effort attempts to describe how the profession of journalism could be envisioned and realized in a technologically enhanced web environment, such as the SW.

The rest of the article is organized as follows. Section 2 provides a detailed literature review and is divided into four subsections. The first subsection deals with the problematic aspects of the current web and proposes the SW's philosophy as an alternative solution, followed by a brief presentation of the old and new web standards on which semantic solutions base their applications. The third subsection outlines the SW's technological framework in which journalism is contextualized. The fourth subsection lists the semantically oriented technological standards developed specifically for the news industry, featuring examples of real use cases. Similarly structured, Section 3 continues with the definition of the term semantic journalism (SJ), followed by its characteristics. Consequently, the classification between certain semantic journalism production processes and publication practices is documented and analyzed. After, Section 4 discusses the major limitations of this study, succeeded by our concluding remarks and future research extensions.

2. Literature Review

2.1. The Emergence of the Semantic Web

The web as an information space has been through various phases of development since it was established for public use. This progress was accomplished through the years, by incorporating modern innovative web technologies that redefined several aspects. This led to a mixture of spectacular successes and failures (Choudhury 2014). Compared to other media—like TV, radio, and newspapers—the simplicity, availability, reachability, and reduced exploitation costs of the web rendered it as the most popular and convenient platform for information publishing and dissemination (Panagiotidis et al. 2019). However, most of this information is published as unstructured text that is made available to a general audience only by means of web pages (Frasincar et al. 2011). In other words, although today's web services facilitate and underpin actions, such as interaction, participation, collaboration, content creation, and information sharing, it seems that the implementation means of these actions are flawed in terms of usability (Saridou et al. 2018). Over the years, the web's infrastructure proved insufficient and unable to manage the ever-growing unstructured amount of inherent information.

The key problem of today's web is located in the following paradox: while numerous technological developments enabled and promoted the production and circulation of web content, the developments required to organize, manage, and manipulate this content have been left behind. Unfortunately, the web grew primarily in terms of volume and not in structure. The media industry, and especially journalism, welcomed the abundance of information on the web because they understood the dynamics and the potential benefits of its exploitation. However, the latter is not fully feasible today because journalists must navigate in a network of unstructured interconnected forms of information (e.g., documents, images, and text) (Kuck 2004), which are designed only for human consumption (Shadbolt et al. 2006; Antoniou et al. 2012). On the one hand, this status quo strengthened journalist performances regarding information availability, but on the other hand, undermined it in terms of functionality and efficiency. In this technological era, where computers cannot read, understand, or manipulate information as humans do, certain journalistic tasks (e.g., discovery, collection, verification, analysis of information, etc.) are difficult to complete in a meaningful and effective way. At this point, SW technologies step in and offer suitable solutions.

According to Berners-Lee et al. (2001), "the 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". The need for this extension appeared in the foreground by the founders of the web themselves as a solution to the problems caused by its anarchic structure. The SW's added value is the structure it proposes. By semantically structuring web pages or documents, web content gains meaning (Poullet et al. 1997), and the web as a whole becomes more accessible to computers (Antoniou et al. 2012). As Choudhury (2014) argues, the "SW is driving the evolution of the current web by enabling machines to understand and respond to complex human requests". Such an understanding requires that the relevant information sources be semantically structured. The aim of the SW architecture (Horrocks et al. 2005) is to structure the web content in such a way to enable users to effectively discover, integrate, exchange, and interlink information (Ankolekar et al. 2007) and improve data management. In the SW framework, every piece of information will transmit its meaning, allowing computer programs to interpret information like humans and intelligently generate and distribute useful content tailored to human needs (Aghaei et al. 2012).

2.2. Semantic Web Components

In order for the semantic solutions to be applied, certain old and new technological standards, for which the World Wide Web Consortium (W3C) is responsible, are used. They are presented briefly below, to introduce the reader into the SW era. It is not within the intentions of this paper to delve deeper into every standard. Specifically:

- The *Universal Code* (Unicode) is the global character encoding system with more than 137,929 characters (Unicode Consortium 2019) and it is used to represent and manipulate text in many languages.
- The *Uniform Resource Identifier* (URI) is a globally scoped character string that is used to identify digital resources or concepts in the web (Berners-Lee et al. 2005).
- The *Extensive Markup Language* (XML) is a markup language that models every piece of information. It is used to define a set of rules for encoding documents in a format that is both human and machine-readable (Bray et al. 2008). Additionally, XML schema is the vocabulary that is used to describe and verify elements inside an XML document and show the interrelationship between them and their complementary attributes (W3C World Wide Web Consortium).
- The *Resource Description Framework* (RDF) is a set of specifications used to create statements for conceptual descriptions of the information implemented in web resources. Moreover, an RDF schema is a data-modelling vocabulary for RDF data and these are used to describe groups of resources and the relationships between them through a hierarchical system of classes and properties (Decker et al. 2000).

- The *Web Ontology Language* (OWL) is a family of representation languages that provides terminological knowledge, bringing reasoning power to the SW. It provides infrastructure to RDF statements by describing concepts, characteristics of properties, constraints, and categories of things and relationships (W3C World Wide Web Consortium). An ontology is an explicit and formal specification of a conceptualization and is used for integrating heterogeneous databases, enabling interoperability among disparate systems and specifying interfaces to independent, knowledge-based services (Gruber 1995).

- Lastly, *SPARQLE Protocol and RDF Query Language* (SPARQL) is a semantic query language for databases and is used to retrieve data that follows the RDF specifications (W3C SPARQL Working Group 2013).

2.3. Contextualize Journalism in the Semantic Web

In today's exploding web landscape, where vast amounts of information are published from different sources across the world, journalists and media professionals often find it difficult to take advantage of all the available data. For this reason, and in order to deal with the everyday workload, professionals use more and more web services, tools, and applications (Maass and Kowatsch 2012). Unfortunately, no matter how many technological aids they use, these are often not sufficient to fully exploit the content of the web. The evolution of the SW will introduce a web of data (Polleres and Huynh 2009) in which the relations of every data item will be fully clarified and structured. As a result, humans and computers will be able to coexist and collaborate in an environment where the interconnection of concepts, rather than simply documents, will be feasible. In this way, journalists will acquire the power to recover targeted information and draw useful conclusions regarding topics that were previously difficult or impossible to research.

The SW promises an upgraded web environment with potential journalistic capabilities, powered by advanced technological features (Antoniou et al. 2012). The major contribution of these features is to define new methods of web information usage by computers, not only for display purposes but also for interoperability between systems (Cardoso et al. 2007). In the context of journalism, this type of technological evolution will unlock various prospects of data exploitation and information handling (Panagiotidis et al. 2020). As Saridou et al. (2018) argued, "the technologies responsible for upgrading the usability of the current web and its content are at the same time responsible for upgrading journalism and its research products". In theory, the applicability of SW's philosophy on contemporary investigative practices leads journalism to a higher functional level.

In practice, the use of semantic techniques for exploiting structured data (Gray et al. 2012) enables the development of new tools and methods not only for information processing (Heravi et al. 2012) but also for discovery, collection, analysis, integration, and reuse. This is why leading media organizations, in cooperation with technological giants and startups, are already beginning to embrace and incorporate SW features into their workflows. British Broadcasting Corporation (BBC) constitutes one of the most notable examples of semantic technologies adoption. The process has been revolutionary and has been in progress since 2004. In brief, it started with a general philosophy of assigning one web address (or Uniform Resource Locator—URL) for every BBC programme (i.e., TV, Radio, etc.) and afterwards converting these identifiers to URIs. Then, it rolled out incrementally with the consumption of Open Data, the adoption of Linked Data principles and the exploitation of RDF and ontologies in the organisation's websites. It ended up with the 2010 World Cup Website, the 2012 Olympic website and the BBC Programmes project, providing, to a so far traditional journalism outlet, significant strategic advantage in the market (Feigenbaum 2012). Although it may not be clearly perceived, low-level semantic technologies are also trending in a few other media pioneers like the New York Times (NYT), Thomson Reuters, Norwegian National Broadcaster (NRK), Agence France Press (AFP) and Associated Press (AP). These are some of the leading media organizations that are working towards the utilization of semantic services, technologies and tools in various ways; ranging from speeding up research to accumulating and cross referencing data (Engels and Tønnesen 2007; Underwood 2019).

Scenery has a wide variety of characteristics. Developers are using semantic technologies to augment the ways in which they create, link, and reuse content on social media sites (Bojars et al. 2008). Services, like Spectee and Trint, which provide the means to extract information from user-generated content (UGC) by semantically analyzing the aural, visual, and textual components, have become increasingly popular. Likewise, services, like Factmata and Fabula AI, and tools, like Twitteraudit, can be seen as detection mechanisms that give journalists the ability to better understand multimedia content and verify the authenticity of information (Saridou et al. 2018). Lastly, there are services, like Google Street View (Anguelov et al. 2010), that correlate content with geolocation data.

2.4. Semantic Web Standards for the Media Industry

Konstantinou et al. (2010) stressed that the great potential of the SW and its technologies become more evident through a number of semantically rich online applications. An indicative set of case studies presented by the W3C Semantic web Education and Outreach (SWEO) Interest Group (www.w3.org/2001/sw/sweo/public/UseCases/), demonstrated the potential carried by semantic technologies in several areas of interest, including the arts, automotive, broadcasting, cultural heritage, education, eGovernment, energy, eTourism, financial, geographic information systems, health care, the IT industry, and many more. Likewise there are semantically oriented technological standards developed specifically for the news industry. The International Press Telecommunications Council (IPCT; https://iptc.org/standards/) is responsible for them and their purpose is to improve the management and exchange of information between content providers, intermediaries, and consumers. Many other news industry giants are currently members of IPCT, like the Deutsche Presse-Agentur (dpa), BBC, Getty Images, Press Association (PA), and Reuters. Some of the most widely spread standards are the following:

- *News Industry Text Format* (NITF) is a solution for sharing news. Through the use of XML, the standard defines the content and structure of news articles in order to make them more searchable and useful. NITF is the means for identifying and describing numerous news characteristics, such as the who, what, when, where and why of an instance. Currently, many news agencies/companies publish NITF documents as Table 1 showcases.
- *Photo metadata* is a global standard that is used to provide general information regarding a picture (i.e., the format, copyrights, location, etc.), to be readable by both machines and humans.
- *Video Metadata Hub* is a newly introduced standard, and it serves as a recommendation, comprising widely used properties of a video (i.e., the name, data type, etc.), for shared video management.
- *Media Topics,* as recently updated, is a vocabulary with over 1200 terms used for the classification of media items based on text categorization.
- *NewsML-G2* is the most common standard for managing and exchanging individual and packages of news items in various formats (i.e., text, images, graphic, video, and audio) between various news outlets (i.e., newspapers, TV stations, and news aggregators). It provides rich metadata in XML format in order to facilitate the convergence with the SW, in terms of rich functionality, ease of use, compactness, and compatibility. Furthermore, the *EventsML-G2 and SportsML* standards are encountered under NewsML-G2, and these are used for the interchange of events and sports data (i.e., facts about an event, statistics, etc.), respectively. Some of the media organizations that are currently using NewsML-G2, EventsML-G2, and SportsML standards are listed in Table 2.
- *RightsML* standard is realized through the Rights Expression Languages. These are machine-readable languages determining the various restrictions/permissions for the use of a digital media item (i.e., an image) when it is made available within the media industry (i.e., photo agencies).
- *Ninjs* standardizes the representation of news in the JavaScript Object Notation (JSON) format. JSON is a data format with a wide range of applications. By expressing news items in JSON format,

the information becomes easily interpreted and interchanged between Application Programming Interfaces (APIs), mobile apps, and others.

- *rNews* is a standard for embedding machine-readable metadata in online news. As news publishers are using their web pages to disseminate information, by using semantic markup to annotate news-specific metadata they render their web page documents and the content of their databases more recoverable by machines and more visible to the search, social, and aggregation ecosystem. The New York Times (https://www.nytimes.com/) and the Austria Presse Agentur (APA) (https://www.ots.at/) are currently using this standard.

Table 1. News agencies/companies that publish News Industry Text Format (NITF) documents.

#	Agency/Company	Description	URL
1	Agence France Presse (AFP)	International News Agency	https://www.afp.com/
2	Agenzia Nazionale Stampa Associata (ANSA)	Italy's largest newswire	https://www.ansa.it/
3	Associated Press (AP) Digital	Division of The Associated Press	http://www.apdigitalnews.com/
4	The Business Journals	Division of American City Business Journals	https://www.bizjournals.com/
5	Stibo DX (formerly known as CCI Europe)	Software company	https://www.stibodx.com/
6	Deutsche Presse-Agentur (dpa)	German News Agency	https://www.dpa.com/de/
7	Inquirer Group of Companies (INQ7 Interactive)	Mass media conglomerate	http://www.inq7.net/
8	AB Kvällstidningen Expressen	Swedish portal	https://www.expressen.se/
9	LexisNexis	Corporation for computer-assisted solutions (i.e., archive database for news)	https://www.lexisnexis.com/en-us/gateway.page
10	netPR	Polish Press Agency	https://netpr.pl/
11	The New York Times	American newspaper	https://www.nytimes.com/
12	Newsmax Medien GmbH	American news media organization	https://www.newsmax.com/
13	Norsk Telegrambyrå (NTB)	Norwegian News Agency	https://www.ntb.no/
14	pressetext Nachrichtenagentur (pte)	Multimedia news agency	https://www.pressetext.com/
15	Tidningarnas Telgrambyrå (TT)	Swedish newswire	https://tt.se/
16	Tiscali GmbH	Munich-based portal	https://www.tiscali.com/

Table 2. Media organizations using NewsML-G2, EventsML-G2, and SportsML standards.

| #	Agency/Company	Description	URL
1	All Headline News (AHN)	Online news wire service	https://allheadlinenews.com/
2	Associated Press (AP)	News agency	https://www.ap.org/en-gb/
3	European Broadcasting Union (EBU)	Alliance of public service media organizations	https://www.ebu.ch/home
4	Fourth Estate	Cooperative Public-benefit corporation	https://www.fourthestate.org/
5	Deutsche Presse-Agentur GmbH (dpa)	German news agency	https://www.dpa.com/de/
6	TT Nyhetsbyrån	News Agency	https://tt.se/
7	YourStory	Digital-media platform	https://yourstory.com
8	AP mobile	Multimedia News portal	https://apnews.com/Mobile
9	Austria Presse Agentur: Startseite (APA)	National News Agency	https://www.apa.at/Site/index.de.html
10	Entertainment and Sports Programming Network (ESPN)	Broadcast Network	https://www.espn.com/
11	PA Media	Multimedia News Agency	https://pa.media/
12	Univision	Television Network	https://www.univision.com/
13	Yahoo! Sports	News website	https://sports.yahoo.com/

The aforementioned commonly used standards were presented briefly with the aim to showcase the added value of semantic technologies. Inarguably, there are more standards used for various media-related purposes; however, it is not within the intentions of the authors to proceed into further analysis.

3. Semantic Journalism Identity

3.1. Definition of Semantic Journalism

The transition from the current web to its extended version—the SW—will mark, likewise, the transition from journalism to semantic journalism, as illustrated in Figure 1. By adopting and

implementing the conception of the SW's evolution in the field of journalism, we introduce the newborn practice of semantic journalism (SJ).
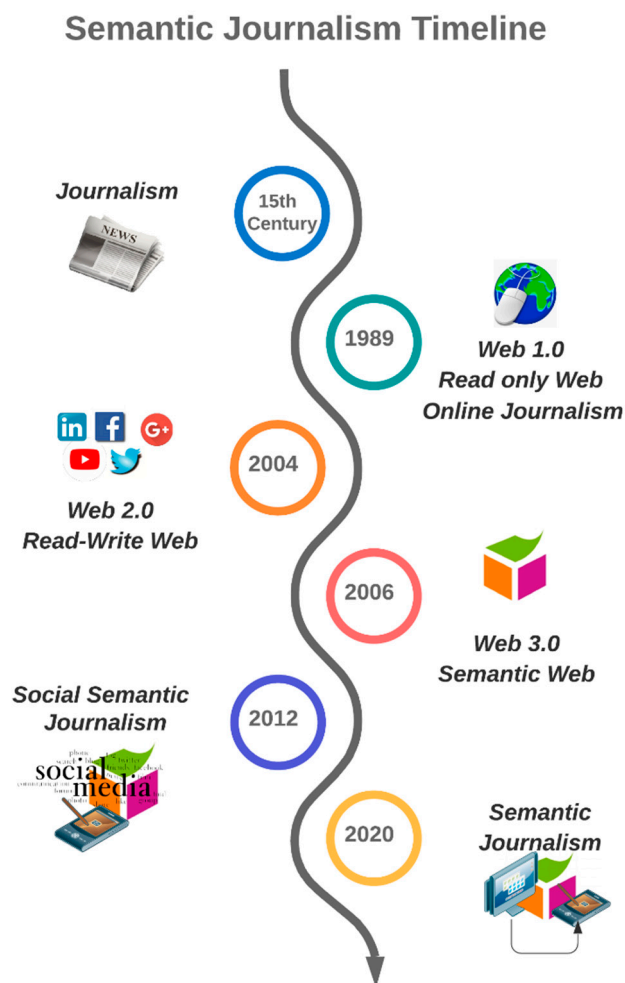
Figure 1. The timeline from journalism to semantic journalism.

The term semantic journalism (SJ) derives from the convergence of the words semantic web and journalism. The key points of SJ theory are the following: It represents a new trend in journalism. It aims at the modernization of existing workflows by adopting the fundamental principles of the SW. It embodies all those use cases of applications with semantic characteristics that can serve journalism. It fulfills the technological enrichment of research methods, such as searching, collecting, editing, categorizing, and verifying information. It leverages specific aspects of semantic technology through a journalistic perspective. It marks the use of new tools that improve the management, analysis, and dissemination of web content for greater efficiency and impact of journalistic outputs. Lastly, it maps a new field of application for semantic technologies.

In detail, SJ is a technologically advanced type of journalism that integrates characteristics from the SW in order to provide journalists with modern alternatives. The use of semantically structured data and tools with semantic features are a few of the first indications that depict the transition from journalism to SJ. This new breed of journalism concerns all journalists who understand the power of the Internet and its services. In recent years, the development of the SW, as an extended version of today's most popular service of the Internet—the World Wide Web—has attracted significant attention due to its aggregation features (Heravi et al. 2012). Some of the media organisations mentioned before have already devoted serious amounts of resources to explore and take advantage of semantic technologies.

This is because the operational and managerial difficulties of today's web leave them with no choice but to search for more efficient methods of data exploitation. The prevailing circumstances are the ones that cause the development and application of semantic technologies in the media. These, in turn, generate fruitful conversations regarding the future of journalism.

3.2. Characteristics of Semantic Journalism

The SW package of theory and practice gives everyday journalistic activities a technological boost. Actions that are able to accommodate semantic features lead to the realization of journalism in the SW context. Through this process, the extension of journalism and the construction of SJ are achieved. The concept of SJ is controversial, however, as although it is easily conceived, it is undocumented. Over the years, initiatives, like CNN's iReport (Hellmueller and Li 2015) and BBC's News Lab BBC Programmes (Dowman et al. 2005) discussed the different ways journalists can benefit from the use of semantic technologies.

One of the recent efforts to investigate and identify this field within the specific context of user-generated content (UGC) and social media, is social semantic journalism (SSJ) (Heravi et al. 2012). The SSJ and its framework deals with the vast amount of UGC across social media platforms and the limited amount of time that journalists have to spare in order to extract potential news stories from these mostly unstructured, unfiltered, and unverified data. For this reason, it proposes a semantic-based solution that can formalise and link unstructured UGC to other semantically-enriched data sets for integration, verification, and fact-checking purposes (Heravi and Mcginnis 2015). The researchers behind the SSJ effort focused on assisting journalists with content creation by providing them with contextual information, first for aggregating and verifying user-generated news and second, for making judgments related to a social media item or a media item from an unknown source.

Another initiative of high significance that embodies semantic features is structured journalism or database journalism (Caswell 2019). This is an emerging form of journalism that publishes news content as entries in a database, enabling users to explore the content in ways that reveal trends and patterns and create new stories and visualizations. The database itself is the story, which enables researchers to explore and sort the content guided by their own personal curiosities, mix and match bits and pieces of information and see tallies of published work (Gourarie 2015).

However, the introduction of SJ serves a wider purpose. SJ is an attempt to address pre-publication and post-publication issues and catalogue every aspect of journalism practice with semantic orientation. This is the reason we believe that SJ is an umbrella term that incorporates not only news production processes, but also publication practices that exploit semantic technologies.

3.3. Semantic Journalism Production Processes

The classification regarding the semantic-oriented news production processes, as depicted in Figure 2, could have been made in various ways as most of them are not completely isolated and independent, but instead have overlaps and are related to each other. As the analysis below will show, although some production processes are self-sufficient and can stand on their own, they can also be identified in one or more publication practices. We choose to proceed with this segmentation in order to investigate them from a specific point of view, that of SJ.

Semantic annotation is defined as the action of describing (part of) an electronic resource (e.g., text, image, video, etc.) using metadata whose meaning is formally specified in an ontology (Fernández 2010). This is the process that creates semantic labels of documents for the SW aiming to establish relations between pieces of annotated data. The semantic annotation plays an important role in a variety of semantic applications, such as the generation of linked data, extraction of information, advanced searching (based on concepts), and reasoning regarding web sources (Pech et al. 2017). Currently, there are a few semantic annotation use cases that are potentially useful for journalists, such as the Concretely Annotated New York Times schema, which provides layers of annotation for over 1.8 million of its published articles (Ferraro Francis et al. 2014). Additionally,

real time semantic annotation use cases that are potentially useful for journalists are the Slovenian Press Agency paradigm (Event Registry 2017) and the SciHi blog (http://scihi.org/) powered by the company YOVISTO.
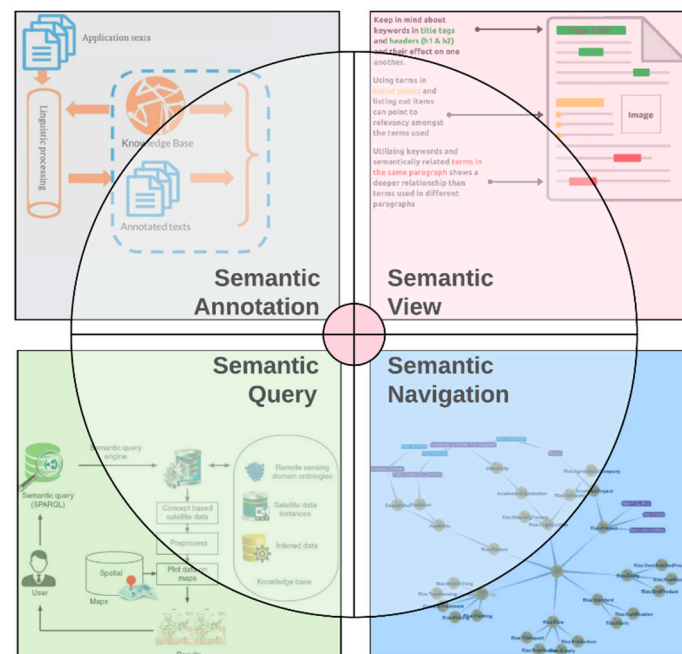
Figure 2. Semantic-oriented news production processes.

The rest fall into the blogosphere's realm. Through the use of semantically enriched blog metadata, the following processes take place (Cayzer and Shabajee 2003; Cayzer 2004a, 2004b). Namely:

The *Semantic View* enables context-sensitive, schema-driven views of the blog content. For example, bibliographic items within a blog are viewed as small explanatory cards arranged in tables, grouped in clusters, and based on a scoring system.

The *Semantic Navigation* facilitates new navigation modalities for finding items of interest by using links such as "agrees with", "part of", and "find similar entries".

The *Semantic Query* empowers richer query and discovery mechanisms over and above free text search. Semantic metadata can help build rich queries for accessing a community's collective knowledge using links, such as "Find all blog entries about a paper written by this author" and "Find all blog items about my friends").

3.4. Semantic Journalism Publication Practices

Despite their contribution so far, the last three processes also play a significant role as parts of the semantic blogging practice. This is one of the four semantic-oriented news production practices demonstrated in Figure 3. *Semantic blogging* is defined by the use of rich metadata, as emphasized above, in order to transform blogs from simple online diaries to full participants of an information-sharing ecosystem. This vision proposes the materialization of informal decentralized knowledge management. The semantic blogging demonstrator is a simple prototype application that was produced around bibliography management and includes the abovementioned semantic capabilities (Reynolds et al. 2004). Additionally, the BBC Programmes is considered a successful paradigm of semantic blogging technology embodiment (Raimond et al. 2010).
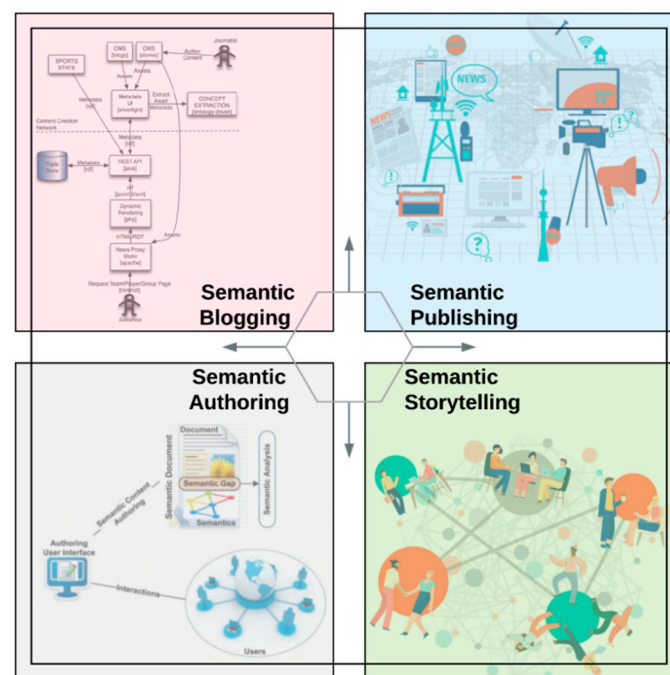
Figure 3. Semantic-oriented news publication practices.

Semantic publishing refers to the process of publishing content on the web accompanied by metadata (Shadbolt et al. 2006) and well-defined, machine-readable information, which enables computers to understand its structure and even its meaning (Shotton et al. 2009). As Peroni (2017) points out, "Semantic publishing involves the use of web and SW technologies and standards for the semantic enhancement, for example of a published journal article, so as to improve its discoverability, interactivity, openness and (re-)usability for both humans and machines". Complementary bibliography states that the term is used to describe a set of technologies that set up the conditions for the semantic enrichment of content, aiming to augment its meaning, facilitate its linking to semantically related articles, provide access to data within the article in actionable form, or permit the integration of data between papers (Kuhn and Dumontier 2017). Representative examples of semantic publishing web sites and solutions, such as the ones of BBC and Ontotext, displayed in Table 3, are beginning to come to the foreground (Georgiev et al. 2013; Gonzalez 2014).

Table 3. Examples of semantic publishing web sites.

#	Agency/Company	Description/Website	URL
1	2010 World Cup	BBC Future Media	http://news.bbc.co.uk/sport2/hi/football/world_cup_2010/default.stm
2	London's Olympic 2012	BBC Future Media	https://www.bbc.com/sport/olympics/2012
3	Sport BBC	BBC Future Media	https://www.bbc.com/sport
4	News on the Web (NOW)	Ontotext Platform	https://now.ontotext.com/#channel?uri=Business&type=channel
5	The Next Web (TNW)	Publisher	https://thenextweb.com/about/
6	Audible	News Agency	https://www.audible.com/
7	Orbis Terra Media	Digital-media platform	https://www.orbisterramedia.com/
8	Press Association (PA)	Multimedia News Agency	https://pa.media/
9	Newz	Dutch tabloid-sized newspaper	https://www.newz.nl/
10	EuroMoney	Magazine	https://www.euromoney.com/
11	Publicis Pixelpark	Advertising Service Provider	https://www.digitaspixelpark.com/
12	Unidad Editorial	Spanish Media Group	http://www.unidadeditorial.com/

Semantic authoring of content is to combine semantically structured pieces of information based on an ontology (Hasida 2007). In detail, as Khalili and Auer (2013) indicated, "semantic authoring is the tool-supported manual composition process of creating semantic documents, resulting either from the use of semantic knowledge representation formalisms (i.e., RDF, OWL, etc.) or from the

use of non-semantic representation forms (i.e., text or hypertext), which are enriched with semantic representations during the authoring process". The digital agency 3pc GmbH provides a prime environment of semantic authoring that can semantically process collections of information in order to enable the efficient authoring of professional, visually appealing, and engaging content products (Rehm et al. 2017). Over the years, multiple ontologies, as presented in Table 4, have been developed and used in the news domain for authoring purposes (Fernández et al. 2010).

Table 4. List of the ontologies used in the news domain.

#	Agency/Company	Ontology Name	URL
1	Agencia EFE - Spanish news agency	News Engine Web Services (NEWS) ontology	https://www.efe.com/efe/espana/1
2	Agenzia Nazionale Stampa Associata (ANSA)—Italian News Agency	News Engine Web Services (NEWS) ontology	https://www.ansa.it/
3	The New York Times	OpenLink New York Times ontology	https://bit.ly/398mU0F & https://www.openlinksw.com/customer-snapshot/
4	British Broadcasting Corporation (BBC)	BBC Ontologies	https://www.bbc.co.uk/ontologies

Semantic storytelling can be conceptualised as the automatic (or semi-automatic) generation of different storylines based on the information extracted, classified, and annotated within extensive textual datasets (Rishes et al. 2013). Semantic storytelling can also be defined as the identification of interesting story paths (Schneider et al. 2017a) or the recommendation of interesting nuggets of information based on a certain set of content using a concrete narrative style or voice (Rehm et al. 2017). Semantic storytelling is a storytelling approach that bundles a flexible set of semantic services, like the semantic analysis of document collections, in order to point out stimulating correlations between the different entities mentioned in the collections (Schneider et al. 2017b; Rehm et al. 2017). Today, there are different semantic storytelling systems, as shown in Table 5. Within their basic functions are, first, to take large amount of documents and extract the entities and relations between them as well as the temporal information and relationships and, second, to automatically produce a hypertext view of the document cluster in order to support journalists to quickly and efficiently familiarise themselves with the content of the processed documents (Rehm et al. 2019).

Table 5. List of news agencies using semantic storytelling systems.

#	Agency/Company	Storytelling System/Service	URL
1	Australian Associated Press (AAP)		
2	Belga		
3	Agenzia Nazionale Stampa Associata (ANSA) – Italian News Agency	Superdesk	https://www.superdesk.org/
4	Suomen Tietotoimisto (STT) - Finnish News Agency		
5	Norsk Telegrambyrå (NTB) – Norwegian News Agency		
6	Zweites Deutsches Fernsehen (ZDF) - German public-service television broadcaster	Condat AG	https://en.condat.de/
7	rbb Fernsehen (RBB) - German free-to-air television channel		

4. Research Challenges and Elements of Discussion

Each of the aforementioned semantic-oriented news production and publication activities contributes to the construction of a sophisticated journalism reality. In this reality, such activities, along with their fundamental operational principles, become integral aspects of journalism. This is where the difference between SJ and today's journalism is located. SJ differs in the culture of web usage. The web as a service, along with its technologies and content, will not have a secondary role in practising journalism, as is happening today in several cases, but likely a primary role instead. As a result, in the future, more sophisticated web content management technologies will be used, providing more complete journalistic products in terms of knowledge acquisition, as well as automated investigative methods powered by new tools. Against this background, a few key limitations appear. The authors of this research focused on three interrelated topics that are open to debate. These are the evolutionary course of the web, the adoption of semantic technologies, and the establishment of SJ.

Specifically, the first and foremost issue that must be discussed concerns the integration of the SW and the usage options. As already explained, the SW effort provides smart solutions and alternatives to the flaws of the current web (Domingue et al. 2011). However, since its first conception in the 1960s, and after the endorsement of the term by Tim Berners Lee and his team in 2001 (Berners-Lee et al. 2001), the SW vision remains incomplete. Admittedly, the fact that it has been under development for a long time (Bengtson 2015) raises many questions regarding the feasibility of fulfillment (Marshall and Shipman 2003), censorship and privacy, the scalability and visualization of content, the availability of ontologies, and the multilinguality and stability of SW languages (Benjamins et al. 2002). The Incubator Group for Uncertainty Reasoning of the World Wide Web (URW3-XG) placed all the challenges that SW has to face under the single heading of 'uncertainty' (Kenneth et al. 2008). The question is, thus, when will the SW develop and become mainstream?

Following this skeptical point of view, another issue arises. Many journalists, media scholars, and professionals question the design, development, and implementation prospects of semantic technologies and debate whether and how they will affect the current status quo in their respected fields (Pellegrini 2012; Halevy and McGregor 2012; Dörr 2016). As presented earlier, there are many recorded cases of semantic technology embodiment. Despite their number, researchers observed that such technologies have not been widely adopted by media organizations (Pomp et al. 2018).

From this perspective, the third issue comes in the foreground and poses a question regarding the evolution and establishment of SJ. The above-mentioned journalistic trends rely on contemporary technologies (software and hardware), matured enough to be used for both live and on-demand news reporting. On the contrary, the build-up of SJ depends on developing techniques and technologies, such as the semantic ones. Their immaturity raises doubts regarding the foundations of SJ and its formulation in today's media landscape. In addition, its dual purpose concerning the enhancement of current reporting methods and the revitalisation of journalism's credibility is undermined by several corporate-related factors, such as problems in incorporating new techniques (Paulussen et al. 2011), disregard for multiple initiatives in the field (i.e., apps and web platforms), investment stagnation in conducting educational workshops and utilizing specialized tools, and, most importantly, the low levels of information literacy (Association of College & Research Libraries, Education & Behavioral Sciences Section Communication Studies Committee A. C. R. L) among employees.

5. Concluding Remarks and Future Research

The researchers of this paper believe that prospective study of the SW philosophy can bring to the foreground adoptable, contemporary approaches regarding journalism. We attempted to introduce an enhanced version of today's journalism in which every activity throughout the reporting process acquires a semantic orientation. Following the rapid proliferation of interest around semantic technologies, it became clear that they have much to offer in response to media needs (Evain 2014). For them to be partially incorporated in journalistic practices and contribute to the rise of a new tendency, the so called 'semantic journalism', was inevitable.

As analysed above, the development and consolidation of the SW and its technologies is the key precondition for the evolution of SJ. Without this, the forms of journalistic practice that use semantic technologies will remain in the background and remain criticized as a science fiction scenario. The apparent superiority of semantic technologies does not seem sufficient to overcome the current web technologies, which remain a cherished mainstay of journalism. Fortunately, there are a few indications in the opposite direction. The adoption of SW philosophy by some media pioneers can motivate other organizations and companies to act similarly and, together, launch the extension of the web. As a result of this extension, several long-established journalistic norms will be re-evaluated (i.e., the use of semantically structured information as raw material for reporting purposes), workflows and practices will be challenged as dysfunctional (i.e., information retrieval, interconnectivity between data sources, conclusion extraction, and the automation of processes) and research methodologies will be upgraded (i.e., the formulation and integration of semantic technologies for the exploitation of

online information). In total, this mobility, at a theoretical and practical level, could help journalism to upgrade its products and regain its prestigious place in society.

Undoubtedly, there are several important aspects in the area of 'semantic journalism' that require further research. As this paper is only the first attempt, future work could include a more detailed literature review on the role of semantic technologies in defining new approaches of practicing journalism, the discovery of more media initiatives that use semantic technologies, the impact of SW in the media industry, the presentation and analysis of SW applications and tools for journalists, the best practices for implementation, research on the correlation between new types of journalism and semantic-oriented web applications, and, lastly, the introduction of a complete SJ practising model. Moreover, the investigation of various theoretical frameworks (for example the diffusion of innovation theory) that can help us understand how new technologies spread and were adopted through time, as well as the hypotheses about potential for the future of information and society as more and more outlets are embracing these technologies, can be very valuable contributions to future extensions of our work.

References

Association of College & Research Libraries, Education & Behavioral Sciences Section Communication Studies Committee (A. C. R. L). 2012. Information literacy competency standards for journalism students and professionals. *College and Research Libraries News* 73: 274–85. [CrossRef]

Aghaei, Sareh, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani. 2012. Evolution of the World Wide Web: From WEB 1.0 TO WEB 4.0. *International Journal of Web & Semantic Technology* 3: 1–10. [CrossRef]

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]

Ankolekar, Anupriya, Markus Krötzsch, Duc Thanh Tran, and Denny Vrandecic. 2007. The Two Cultures: Mashing up Web 2.0 and the Semantic Web. Paper presented at the 16th International Conference on World Wide Web, Alberta, Canada, May 8–12; pp. 825–34.

Antoniou, Grigoris, Paul Groth, Frank van Harmelen, and Rinke Hoekstra. 2012. *A Semantic Web Primer Hardcover*, 3rd ed. Cambridge: The MIT Press.

Bengtson, Jason. 2015. The Semantic Revolution. *Journal of Electronic Resources in Medical Libraries* 12: 72–82. [CrossRef]

Benjamins, Victor Richard, Jesús Contreras, Oscar Corcho, and Asunción Gómez-Pérez. 2002. The six challenges of the Semantic Web. Available online: http://oa.upm.es/5668/1/Workshop06.KRR2002.pdf (accessed on 9 April 2020).

Berners-Lee, Tim, Hendler James, and Lassila Ora. 2001. The semantic web. *Scientific American* 284: 28–37. [CrossRef]

Berners-Lee, Tim, Roy Fielding, and Larry Masinter. 2005. Uniform Resource Identifiers (URI): Generic Syntax. *Internet Engineering Task Force*. [CrossRef]

Bojars, Uldris, Alexandre Passant, John Breslin, and Stefan Decker. 2008. Social Networks and Data Portability using Semantic Web technologies. Paper presented at BIS 2008 Workshop on Social Aspects of the Web in Conjunction with 11th International Conference on Business Information Systems, Innsbruck, Austria, May 5–7; Available online: http://hdl.handle.net/10379/398 (accessed on 9 April 2020).

Bray, Tim, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yergeau. 2008. Extensible Markup Language (XML) 1.0 (Fifth Edition). *W3C Recommendation*. Available online: https://www.w3.org/TR/REC-xml/ (accessed on 8 April 2020).

Bull, Andy. 2010. *Multimedia Journalism: A Practical Guide*. Abingdon-on-Thames: Routledge.

Cardoso, Onelio Jorge, Martin Hepp, and Miltiadis Lytras. 2007. *Real-World Applications of Semantic Web Technology and Ontologies*. Berlin and Heidelberg: Springer, ISBN 978-0-387-48530-0.

Caswell, David. 2019. Structured journalism and the semantic units of news. *Digital Journalism*, 1134–56. [CrossRef]

Cayzer, Steve. 2004a. Semantic Blogging: Spreading the Semantic Web Meme. *XML Europe*. Available online: https://purehost.bath.ac.uk/ws/files/129237838/HPL_2004_48.pdf (accessed on 9 April 2020).

Cayzer, Steve. 2004b. Semantic Blogging and decentralized knowledge management. *Communication of the ACM* 47: 47–52. [CrossRef]

Cayzer, Steve, and Paul Shabajee. 2003. Semantic Blogging and Bibliography Management. Paper presented at BlogTalk Proc, Vienna, Austria, May 23–24; pp. 101–8.

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 (IJCSIT)* 5: 8096–100.

Dalakas, Andreas, Dimoulas Charalampos, Kalliris Giorgos, and Veglis Andreas. 2017. Drone journalism: Generating immersive experiences. *Journal of Media Critiques* 3: 187–99. [CrossRef]

Decker, Stefan, Sergey Melnik, Frank Harmelen, Dieter Fensel, Michel Klein, Michael Erdmann, and Ian Horrocks. 2000. The semantic web: the roles of XML and RDF. *IEEE Internet Computing*. [CrossRef]

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

Deuze, Mark, and Tamara Witschge. 2018. Beyond Journalism: Theorizing the Transformation of Journalism. *Journalism* 19: 165–81. [CrossRef] [PubMed]

Domingue, John, Dieter Fensel, and James A. Hendler. 2011. *Handbook of Semantic Web Technologies*. Berlin and Heidelberg: Springer Science & Business Media.

Dörr, Konstantin Nicholas. 2016. Mapping the field of Algorithmic Journalism. *Digital Journalism* 4: 700–22. [CrossRef]

Dowman, Mike, Valentin Tablan, Cristian Ursu, Hamish Cunningham, and Borislav Popov. 2005. Semantically enhanced television news through web and video integration. Paper presented at Second European Semantic Web Conference (ESWC'2005), Heraklion, Crete, Greece, May 29–June 1.

Engels, Robert H. P., and Jon Roar Tønnesen. 2007. A Digital Music Archive (DMA) for the Norwegian National Broadcaster (NRK) using Semantic. *World Wide Web Consortium. Semantic Web Use Cases and Case Studies*. Available online: https://www.w3.org/2001/sw/sweo/public/UseCases/NRK/NRK.pdf (accessed on 8 April 2020).

Evain, Jean-Pierre. 2014. Semantic Technologies in Broadcasting Production. Paper presented at 10th International Conference on Semantics, Knowledge and Grids, Beijing, China, August 27–29.

Event Registry. 2017. Use of Semantic Annotation in News Monitoring Explained. *Medium*. Available online: https://medium.com/event-registry/use-of-semantic-annotation-in-news-monitoring-explained-ead7445f2b6f (accessed on 9 April 2020).

Feigenbaum, Lee. 2012. BBC's Adoption of Semantic Web Technologies: An Interview. *CMS Wire*. Available online: https://www.cmswire.com/cms/information-management/bbcs-adoption-of-semantic-web-technologies-an-interview-017981.php (accessed on 30 June 2020).

Fernández, Norberto. 2010. Semantic Annotation Introduction. In Liao Yongxin, Lezoche Mario, Panetto Hervé and Boudjlida Nacer. 2011. Why, Where and How to use Semantic Annotation for Systems Interoperability. Proceedings of the 1st UNITE Doctoral Symposium. Romania. Available online: https://hal.archives-ouvertes.fr/hal-00597903/document (accessed on 9 April 2020).

Fernández, Norberto, Damaris Fuentes Luis Sánchez, and Jesús A. Fisteus. 2010. The NEWS ontology: Design and applications. *Expert Systems with Applications* 37: 8694–704. [CrossRef]

Ferraro Francis, Max Thomas, Matthew Gormley, Travis Wolfe, Craig Harman, and Benjamin Van Durme. 2014. Concretely Annotated Corpora. Paper presented at NIPS Workshop on Automated Knowledge Base Construction (AKBC), NIPS Workshop, Long Beach, CA, USA, June 22–24. Available online: https://www.akbc.ws/2014/submissions/akbc2014_submission_18.pdf (accessed on 9 April 2020).

Frasincar, Flavius, Jethro Borsje, and Frederik Hogenboom. 2011. Personalizing news services using semantic web technologies. In *E-Business Applications for Product Development and Competitive Growth: Emerging Technologies*. Hershey: IGI Global, pp. 261–89. [CrossRef]

Georgiev, Georgi, Borislav Popov, Petya Osenova, and Marin Dimitrov. 2013. Adaptive Semantic Publishing. Ontotext AD, Bulgaria. In *WaSABi'13: Proceedings of the 2013th International Conference on Semantic Web Enterprise Adoption and Best Practice*. vol. 1106, pp. 35–44. Available online: http://ceur-ws.org/Vol-1106/paper4.pdf (accessed on 9 April 2020).

Gonzalez, Jose. 2014. Semantic Publishing: A Case Study for the Media Industry. *Meaning Cloud*. Available online: https://www.meaningcloud.com/blog/semantic-publishing-case-study-media-industry (accessed on 9 April 2020).

Gourarie, Chava. 2015. Structured journalism offers readers a different kind of story experience. *Columbia Journalism Review*. Available online: https://www.cjr.org/innovations/structured_journalism.php (accessed on 9 April 2020).

Gray, Jonathan, Lucy Chambers, and Liliana Bounegru. 2012. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Sebastopol: O'Reilly Media, Inc.

Gruber, Thomas. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* 43: 907–28. [CrossRef]

Halevy, Alon Y., and Susan McGregor. 2012. Data Management for Journalism. *IEEE Data Engineering Bulletin* 35: 7–15.

Hasida, Kôiti. 2007. Semantic Authoring and Semantic Computing. In *New Frontiers in Artificial Intelligence*. Edited by A. Sakurai, K. Hasida and K. Nitta. JSAI 2003, JSAI 2004. Lecture Notes in Computer Science. Berlin and Heidelberg: Springer, vol. 3609. [CrossRef]

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]

Heravi, Bahareh Rahmanzadeh, and Jarred Mcginnis. 2015. Introducing Social Semantic Journalism. *The Journal of Media Innovations* 2. [CrossRef]

Heravi, Bahareh Rahmanzadeh, Marie Boran, and John Breslin. 2012. Towards social semantic journalism. In *Sixth International AAAI Conference on Weblogs and Social Media*. Dublin: AAAI Press, pp. 14–17.

Horrocks, Ian, Bijan Parsia, Patel-Schneider Peter, and Hendler James. 2005. Semantic web architecture: Stack or two towers? In *International Workshop on Principles and Practice of Semantic Web Reasoning*. Berlin and Heidelberg: Springer, pp. 37–41.

Janev, V., and S. Vranes. 2009. Semantic web technologies: Ready for adoption? *IT Professional* 11: 8–16. [CrossRef]

Kennedy, Tom. 2010. American University School of Communication Whitepaper on Backpack Journalism. *Studylib*. Available online: https://studylib.net/doc/13886164/american-university-school-of-communication-whitepaper-on (accessed on 8 April 2020).

Kenneth, J. Laskey, Kathryn B. Laskey, Paulo C. G. Costa, Mieczyslaw M. Kokar, Trevor Martin, and Thomas Lukasiewicz. 2008. Uncertainty Reasoning for the World Wide Web. *World Wide Web Consortium. Incubator Group Report*. Available online: https://www.w3.org/2005/Incubator/urw3/XGR-urw3/ (accessed on 9 April 2020).

Khalili, Ali, and Sören Auer. 2013. User interfaces for semantic authoring of textual content: A systematic literature review. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 22: 1–18. [CrossRef]

Konstantinou, Nikolaos, Dimitrios-Emmanuel Spanos, Periklis Stavrou, and Nikolas Mitrou. 2010. Technically Approaching the Semantic Web Bottleneck. *International Journal of Web Engineering and Technology* 6: 83–111. [CrossRef]

Kuck, G. 2004. Tim Berners-Lee's Semantic Web. *SA Journal of Information Management* 6. [CrossRef]

Kuhn, Tobias, and Michel Dumontier. 2017. Genuine semantic publishing. *Data Science* 1: 139–54. [CrossRef]

Latar, Noam Lemelshtrich. 2018. Robot journalism: Can Human Journalism Survive? In *Robot Journalism*. Singapore: World Scientific, pp. 29–40. [CrossRef]

Maass, Wolfgang, and Tobias Kowatsch, eds. 2012. *Semantic Technologies in Content Management Systems: Trends, Applications and Evaluations*. Berlin and Heidelberg: Springer Science & Business Media.

Marshall, Catherine C., and Frank M. Shipman. 2003. Which semantic web? Paper presented at ACM Conference on Hypertext and Hypermedia, Nottingham, UK, August 26–30; pp. 57–66.

Newman, Nic. 2019. Journalism, Media and Technology Trends and Predictions 2019. *Reuters Institute for the Study of Journalism*. Available online: http://www.digitalnewsreport.org/publications/2019/journalism-media-technology-trends-predictions-2019/#2-3-ai-and-the-newsroom (accessed on 6 May 2020).

Osman, Taha, Dhavalkumar Thakker, and Gerald Schaefer. 2014. Utilising semantic technologies for intelligent indexing and retrieval of digital images. *Computing* 96: 651–68. [CrossRef]

Panagiotidis, Kosmas, Tsipas Nikolaos, Saridou Theodora, and Veglis Andreas. 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.

Panagiotidis, Kosmas, Tsipas Nikolaos, Saridou Theodora, and Veglis Andreas. 2020. A Participatory Journalism Management Platform: Design, Implementation and Evaluation. *Social Sciences* 9: 21. [CrossRef]

Paulussen, Steve. 2016. Innovation in the Newsroom. In *The SAGE Handbook of Digital Journalism*. Edited by Tamara Witschge, C. W. Anderson, David Domingo and Alfred Hermida. London: Sage, pp. 113–27.

Paulussen, Steve, Davy Geens, and Kristel Vandenbrande. 2011. Fostering a Culture of Collaboration: Organizational Challenges of Newsroom Innovation. In *Making Online News*. Edited by David Domingo and Chris Paterson. Newsroom Ethnographies in the Second Decade of Internet Journalism. New York: Peter Lang, vol. 2, pp. 3–14.

Pech, Fernando, Alicia Martinez, Hugo Estrada, and Yasmin Hernandez. 2017. Semantic Annotation of Unstructured Documents Using Concepts Similarity. *Scientific Programming Techniques and Algorithms for Data-Intensive Engineering Environments* 2017: 7831897. [CrossRef]

Pellegrini, Tassilo. 2012. Semantic metadata in the news production process: Achievements and challenges. Paper presented at 16th International Academic Mindtrek Conference, Tampere, Finland, October 3–5; pp. 125–33.

Peroni, Silvio. 2017. Automating semantic publishing. *Data Science* 1: 155–73. Available online: https://content.iospress.com/articles/data-science/ds012 (accessed on 4 May 2020). [CrossRef]

Polleres, Axel, and David Huynh Huynh. 2009. Special issue: The web of data. In *Journal of Web Semantics First Look 7_3_1*. Amsterdam: Elsevier.

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]

Poullet, Line, Pinon Jean-Marie, and Calabretto Sylvie. 1997. Semantic structuring of documents. Paper presented at Third Basque International Workshop on Information Technology—BIWIT'97—Data Management Systems, Biarritz, France, July 2–4; pp. 118–24. [CrossRef]

Raimond, Yves, Scott Tom, Oliver Silver, Sinclair Patrick, and Smethurst Michael. 2010. Use of Semantic Web technologies on the BBC Web Sites. In *Linking Enterprise Data*. Boston: Springer, pp. 263–83. [CrossRef]

Rehm, Georg, Julián Moreno-Schneider, Peter Bourgonje, Ankit Srivastava, Rolf Fricke, Jan Thomsen, Jing He, Joachim Quantz, Armin Berger, Luca König, and et al. 2017. Different Types of Automated and Semi-automated Semantic Storytelling: Curation Technologies for Different Sectors. In *Language Technologies for the Challenges of the Digital Age*. Edited by G. Rehm and T. Declerck. Lecture Notes in Computer Science. Cham: Springer, vol. 10713.

Rehm, Georg, Karolina Zaczynska, and Julian Moreno-Schneider. 2019. Semantic Storytelling: Towards Identifying Storylines in Large Amounts of Text Content. Paper presented at the Text2StoryIR '19 Workshop, Cologne, Germany, April 14. Edited by A. Jorge, R. Campos, A. Jatowt and S. Bhatia. Available online: http://ceur-ws.org.Availableonline:http://ceur-ws.org/Vol-2342/paper8.pdf (accessed on 9 April 2020).

Reynolds, Dave, Cayzer Steve, Shabajee Paul, and Steer Damian. 2004. SWAD-Europe deliverable 12.1.8: SWAD-E Demonstrators—Lessons Learnt. *World Wide Web Consortium*. Available online: http://www.w3.org/2001/SW/Europe/reports/demo-lessons-report/ (accessed on 9 April 2020).

Richardson, Allissa. 2012. Mobile journalism: A model for the future. *Diverse Issues in Higher Education* 29: 24.

Rishes, Elena, Stephanie M. Lukin, David K. Elson, and Marilyn A. Walker. 2013. Generating Different Story Tellings from Semantic Representations of Narrative. In *ICIDS. Lecture Notes in Computer Science*. Edited by H. Koenitz, T. I. Sezen, G. Ferri, M. Haahr, D. Sezen and G. Catak. Interactive Storytelling. Cham: Springer, vol. 8230. [CrossRef]

Rodríguez, Nohemí Lugo. 2018. Immersive journalism design within a transmedia space. In *Exploring Transmedia Journalism in the Digital Age*. Hershey: IGI Global, pp. 67–82. [CrossRef]

Saridou, Theodora, Panagiotidis Kosmas, Tsipas Nikolaos, and Veglis Andreas. 2018. Semantic Tools for Participatory Journalism. *Journal of Media Critiques* 4: 281–94. [CrossRef]

Schneider, Moreno Julian, Bourgonje Peter, and Rehm Georg. 2017a. Towards User Interfaces for Semantic Storytelling. In *HIMI 2017. Lecture Notes in Computer Science*. Edited by S. Yamamoto. Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration. Cham: Springer, vol. 10274. [CrossRef]

Schneider, Moreno Julian, Ankit Srivastava, Peter Bourgonje, David Wabnitz, and Georg Rehm. 2017b. Semantic storytelling, cross-lingual event detection and other semantic services for a newsroom content curation dashboard. In *Proceedings of the Second Workshop on Natural Language Processing meets Journalism–EMNLP 2017 Workshop (NLPMJ 2017)*. Edited by O. Popescu and C. Strapparava. Copenhagen: Association for Computational Linguistics, pp. 68–73.

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

Shotton, David, Katie Portwin, Graham Klyne, and Alistair Miles. 2009. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Computational Biology* 5: e1000361. [CrossRef] [PubMed]

Spyridou, Lia-Paschalia, Matsiola Maria, Veglis Andreas, Kalliris Giorgos, and Charalampos Dimoulas. 2013. Journalism in a state of flux: Journalists as agents of technology innovation and emerging news practices. *International Communication Gazette* 75: 76–98. [CrossRef]

Underwood, Corina. 2019. Automated Journalism—AI Applications at New York Times, Reuters, and Other Media Giants. *Emerji. The AI Research and Advisory Company*. Available online: https://emerj.com/ai-sector-overviews/automated-journalism-applications/ (accessed on 8 April 2020).

Unicode Consortium. 2019. Unicode Version 12.0.0. *Unicode*. Available online: http://www.unicode.org/versions/Unicode12.0.0/ (accessed on 8 April 2020).

Van der Haak, Bregtje, Michael Parks, and Manuel Castells. 2012. The future of journalism: Networked journalism. *International Journal of Communication* 6: 16.

Veglis, Andreas, and Charalampos Bratsas. 2017. Reporters in the age of data journalism. *Journal of Applied Journalism & Media Studies* 6: 225–44. [CrossRef]

Veglis, Andreas, and Theodora A. Maniou. 2018. The Mediated Data Model of Communication Flow: Big Data and Data Journalism. *KOME: An International Journal of Pure Communication Inquiry* 6: 32–43. [CrossRef]

W3C (World Wide Web Consortium). 2012. OWL 2 Web Ontology Language Document Overview (Second Edition). *W3C Recommendation*. Available online: https://www.w3.org/TR/owl-overview/ (accessed on 8 April 2020).

W3C (World Wide Web Consortium). 2015. XML Technology. Schema. *w3.org*. Available online: https://www.w3.org/standards/xml/schema (accessed on 8 April 2020).

W3C SPARQL Working Group. 2013. SPARQL 1.1 Overview. *W3C Recommendation*. Available online: https://www.w3.org/TR/sparql11-overview/ (accessed on 8 April 2020).