Journalistic Knowledge Platform

Subjects: Computer Science, Interdisciplinary Applications | Computer Science, Information Systems | Computer Science, Artificial Intelligence
Contributor: Marc Gallofré Ocaña , Andreas Opdahl

A Journalistic Knowledge Platform (JKP) is an information system that employ artificial intelligence and big data techniques such as machine learning and knowledge graphs to manage and support the knowledge work needed in all stages of news production. JKPs automate the process of annotating metadata and support daily workflows like news production, archiving, monitoring, management and distribution. JKPs harvest and analyse news and social media information over the net in real time, leverage encyclopaedic sources, and provide journalists with both meaningful background knowledge and newsworthy information. JKPs can provide a digitalisation path towards reduced production costs and improved information quality while adapting the current workflows of newsrooms to new forms of journalism and readers' demands.

journalistic knowledge platform   artificial intelligence   knowledge graph   journalism

newsroom   information system

1.Extended definition

Innovation and digitalisation of newsrooms are needed to increase the quality and lower the cost of news production, changing how journalists and readers interact with news content and background information [1]. Newsrooms are therefore embracing big data and artificial intelligence (AI) techniques such as knowledge graphs and machine learning (ML) to manage and support the knowledge work needed in all stages of news production. The result is an emerging type of intelligent information system called the *Journalistic Knowledge Platform (JKP)*. JKPs can be described from a functional, an organisational and a technical perspective. From a functional point of view JKPs automate the process of annotating metadata and support daily workflows like news production [2][3], archiving [4][5], management [6][7] and distribution [8][9][10][11]. JKPs harvest and analyse news and social media information over the net in real time [12], leverage encyclopaedic sources [13], and provide journalists with both meaningful background knowledge [14] and newsworthy information [15]. From an organisational viewpoint: JKPs are deployed in newsrooms to manage the knowledge needed to support journalists with creativity and discovery tasks. These are tailored to the particular digital strategies and editorial lines to improve news broadcast. JKPs also follow media standards to facilitate communication with customers and providers, and are subject to legal regulations such as data privacy. From a technical perspective JKPs implement state-of-the-art AI technologies such as machine learning, natural language processing (NLP) and knowledge representation and reasoning. News-relevant information is represented in knowledge bases which are exploited with data analysis, reasoning

and information retrieval techniques to help journalists and readers dive more deeply into information, events and storylines.

2. State of Research on JKPs

1.1. Stakeholders

JKPs provide services to and interact with a large variety of stakeholders. **Figure 1** shows the identified stakeholders and their three top-level categories: general user, organisation and technical agent.
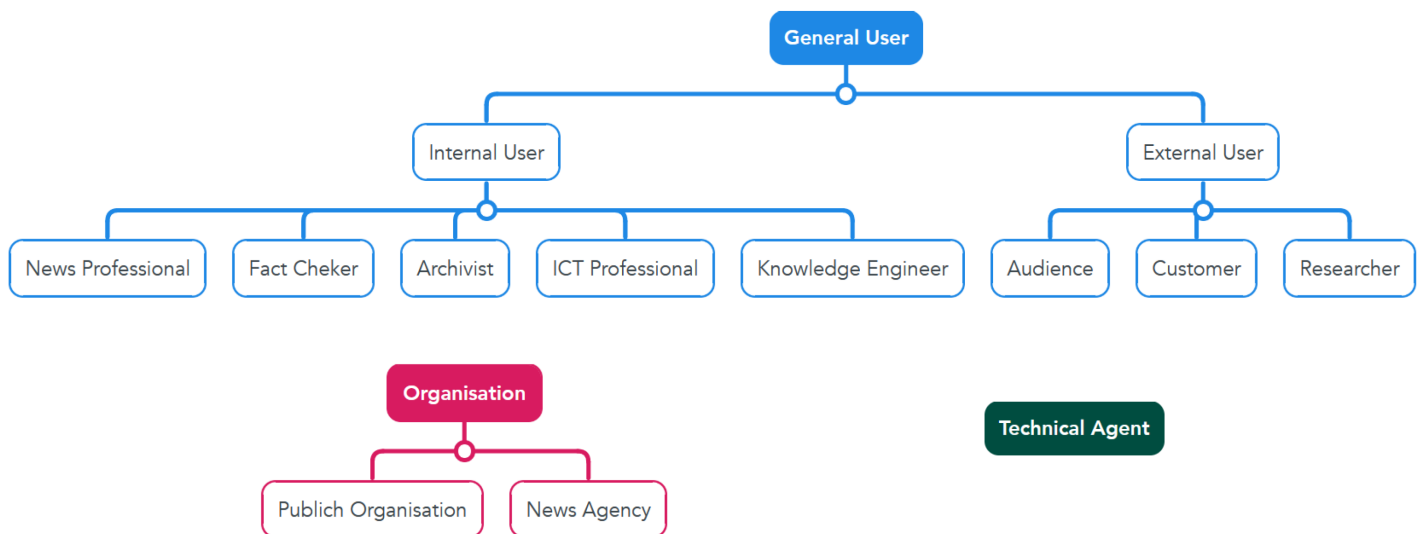

Figure 1. Stakeholder categories.

The general users can be divided between the internal users that belong to newsrooms and the external ones. The internal users are news professionals like journalists who use JKPs for creating histories [16][17]; fact-checkers who conduct an essential task in combating with fake news and misinformation [7]; archivists who maintain up-to-date the schemas and news archives [4]; ICT professionals and knowledge engineers who develop and maintain JKPs [2]. Whereas, the external users are the audience [11]; the customers to whom new agencies offer services and researchers who investigate JKPs or use JKP to analyse data.

JKPs support organisations in different ways: The most direct is in news agencies and news organisations where JKPs are deployed and adapted to particular digital strategies and purposes, but also to other news organisations that consume services from external JKPs. Moreover, JKPs provide services to both private and public organisations like governmental agencies that interact with or consume services from newsrooms. JKPs also interact indirectly with the organisations responsible for controlling news media standards, vocabulary and ontologies (e.g., the IPTC organisation). This impacts how JKPs are designed because the work of many news agencies depends on those standards, and JKPs often need to build on and comply with them. However, the media standards may not cover or fit the use cases of newsrooms. Hence, JKPs need to adapt or expand the media standards according to their needs.

Last but not least, the technical agent represents the JKPs and any system or technical infrastructure in newsrooms that support or interact with JKPs. A sub-type of the technical agent is the external system that communicates with newsroom services, like the customers' information systems [16].

1.2. Information

JKPs cover the whole news production pipeline from gathering information and news creation to knowledge exploitation and distribution. **Table 1** lists the identified categories of information.

Table 1. The most common types of information managed by JKPs.

Information	Explanation
News content	The reported story or event.
Textual data	Textual information.
Multimedia data	Images, videos and audio information.
Data format	The format in which the data is stored or structured.
Metadata	Data about or that describe the news content.
Linked Open Data (LOD)	Structured and open available data on the Internet (e.g., data from Wikidata and DBpedia) [18]
Events	Newsworthy happenings.
Information needs	Different information types and categories of interest.

News content is annotated and enriched with metadata using LOD, semantic vocabularies and ontologies. Metadata can describe different types of basic information like the authorship, language, creation time, ownership, media type, priority, status, version, keywords and categories; as well as inferred information like provenance, tone and sentiment, and the relevant persons, stories, locations, organisations and events [4][19][20].

Journalists and customers of newsrooms are highly interested in current events and their related information [2]. In addition, JKPs are designed to support additional information needs: General users want to have access to details about the stories (i.e., who, what, why, where and when), identify networks of actors and implications, search the events based on their type or place, obtain facts, and retrieve evidences [5][6][14]. News professionals need access to news archives and knowledge bases for documentation purposes, finding connections from past events, following histories and identifying emerging topics [4][16][21][22]. Additionally, customers have different information needs depending on their business or interests.

1.3. Functionalities

JKPs provide different functionalities to their users. **Table 2** lists the identified main functionalities.

Table 2. Most common type of functionalities and services provided in JKPs.

Functionality	Explanation
News creation	The process to create a news story.
Verification	The process of checking the facts and claims.
Source selection	The ability to select the information sources of interest.
Monitoring	The ability to continuously distil information from source.
Knowledge discovery	Functionalities for exploring relevant information.
Trends	The current newsworthy developments.
Alert	A notification.
Summarisation	Extracting and representing the key information from a larger text or group of text.
Clustering	Grouping similar stories or events.
Business support	Functionalities to support management workflows.
Content management	Functionalities oriented to store, organise and distribute information.
Personalisation	Providing information according to the user's interests.

News professionals use JKPs for news creation. This creative process involves different tasks such as discovering, collecting, organising, contextualising and publishing [23][24]. JKPs guide news professionals in writing up their stories [25], support them with contextual background knowledge [2][3][25], provide the means for comparing current events with other events [13] and facilitate access to previous work for creating similar content for a different audience, region or language [22]. JKPs also support news professionals with verification [26] tasks like fact-checking [9][27], provenance [5], rights and authorship management [16]. These are typically time-consuming tasks for journalists and fact-checkers that JKPs automate [7].

Source selection and monitoring functionalities are common across the studied JKPs that harvest and store content from internal and external sources and monitor them in real-time [9][11][21][22]. These functionalities allow journalists to automatically follow and distil news and social media of interest and relieve them from these time-consuming tasks.

Knowledge discovery [28] is one of the most attractive functionalities of JKPs. It allows users to obtain news insights, analysis and relevant information. Other interesting functionalities among the studied JKPs are the trends identification used to discover emerging topics, long-term developments and changes in events over time [11][20];

alerts to keep users up-to-date with the last incoming items [9][29][30]; summarisation [31] of news histories and events to provide additional insights [11]; clustering of story lines and events [13][22].

JKPs can be used as business support systems to manage and monitor internal newsrooms production, news coverage and broadcast decisions [22][29]. This helps managers and editors in allocating resources, avoiding duplicate work and detecting news that can be relevant to different audiences. JKPs are also used for content management that allows newsrooms to store, organise and distribute the daily produced content and metadata [4][6] [16].

Most of these functionalities should be personalised and tailored to the stakeholders' needs. Hence, JKPs allow the personalisation of their functionalities according to users' preferences and profiles [2][8][32].

1.4. Techniques

JKPs implement and combine different IT techniques to fulfil their functionalities. **Table 3** lists the IT techniques that identified.

Table 3. The most common IT techniques used in JKPs.

Technique	Explanation
Semantic technologies	Set of technologies designed to work with LOD and semantic data [33].
Fact extraction	The techniques used to identify factual claims.
Conceptual model	A representations of the world or a part of.
Reasoning	The techniques used to infer knowledge.
Network analysis	The techniques used to analyse networks of things.
Event analysis	The techniques used to analyse events.
Natural Language Processing (NLP)	A set of techniques intended to work and process language.
AI training	The process of creating and tuning an AI model to perform on a given dataset or scenario.

Semantic technologies [33] and similar semantic representation techniques are widely utilised in all the studied JKPs. They use semantic technologies for automating annotation, disambiguating, enriching and leveraging news items with information from external knowledge bases [2][4][9][20]. The semantic representations provide neutral language, explicit relations and facilitate structural matching and lingual independence. They are used for clustering news items and events [13] and detecting trends and story lines [5]. These semantic representations

together with fact extraction techniques are used to obtain factual claims from news items and link them to their sources and facts in external knowledge bases (e.g., Wikidata, Wikipedia) [5][9][22].

Conceptual models provide vocabularies, schemas and ontologies. These are often implemented using semantic technologies and represent news stories, events and related information. In addition, conceptual models can define users' interests and preferences [8][10][16], and provide shared resources and formats to facilitate content management and semantic interoperability [4][6][14][20].

Conceptual models and semantic technologies are also used for reasoning, network analysis and event analysis. Reasoning techniques abstract and infer new knowledge from news items, events and temporal aspects [20]. Network analysis is used to find networks of actors, organisations and their implications [5]. Event analysis is applied to detect, identify, cluster and annotate the events described in the news [11][13][16].

The aforementioned techniques are supported by NLP tasks such as named entity recognition, relation extraction and temporal expression normalisation [9][10][11][20][34]. These NLP tasks, among others, are used in many of the components and functionalities of JKPs. In order to obtain optimal results from the NLP tasks, near-continuous training on extensive news corpora [13] is needed to always keep the machine learning models up-to-date.

1.5. Components

JKPs rely on different components to fulfil their functionalities and support users (see **Figure 2**).
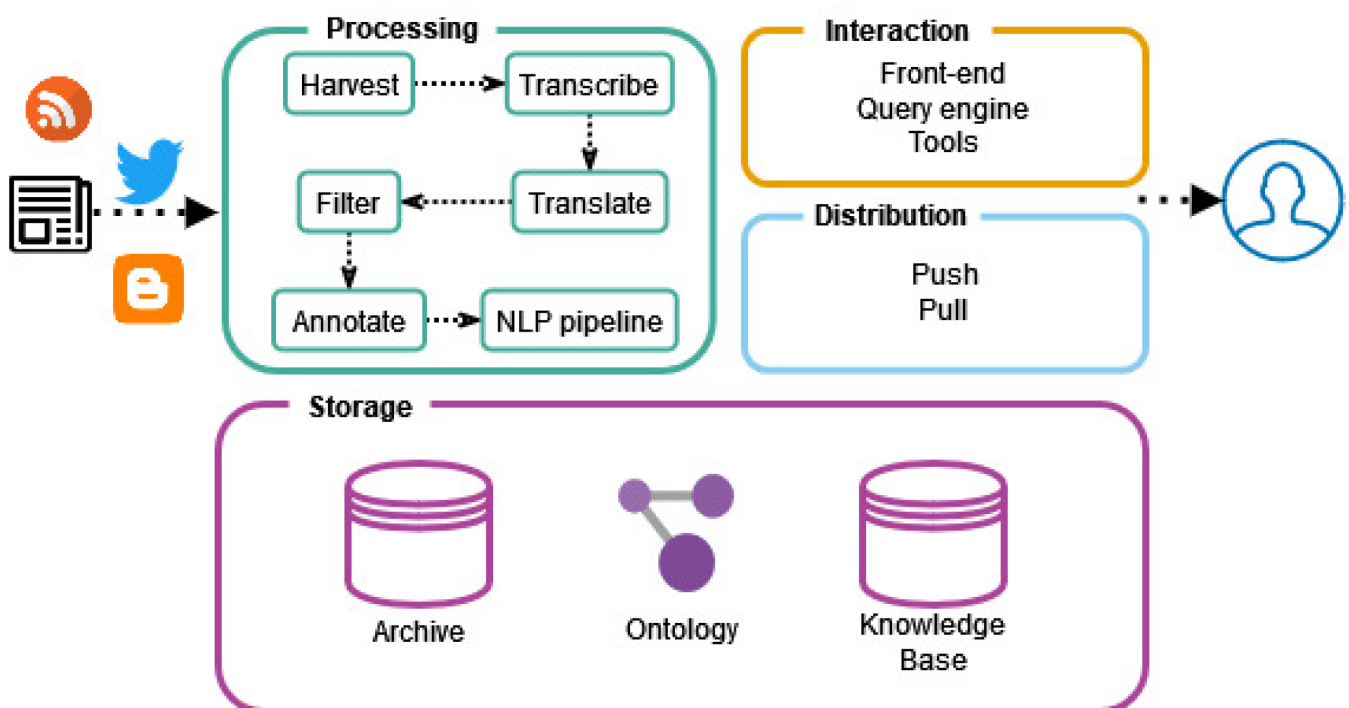

Figure 2. JKP components.

The processing components cover tasks from data gathering to transforming input sources into knowledge representations. The textual and multimedia sources are continuously harvested. However, not all contents receive the same interest from news professionals. Thus, the harvested content is also translated [22] and filtered according with the different stakeholders' interests and needs. In the studied JKPs, spoken content is transcribed [22] and images are textually described [2] to be able further process them.

The harvested content is automatically annotated with metadata (e.g., authorship, categories and topics) to support functionalities like business support, content management and personalisation [4][16][29][32]. The annotated content is often processed by a NLP pipeline using state-of-the-art NLP and natural language understanding modules to perform linguistic tasks such as co-reference resolution, named entity recognition, relation extraction and sentiment analysis [5][9][35]. Both the results of the NLP pipeline and the annotated content are represented semantically following a predefined schema or ontology. These representations link the annotations to a knowledge base (e.g., an RDF-based knowledge graph) [10][20] and enrich the news items with facts from external knowledge bases (e.g., the LOD cloud, DBpedia and Wikidata) [5][13].

The storage infrastructure of a JKP can be composed of an archive, an ontology and a knowledge base. The archive can store millions of historical news articles, biographies, reports [4][20] and other relevant textual and multimedia items. The knowledge base is where the annotated semantic representations of news items are stored and enriched with external information [4][5][14]. The ontology is used to represent the structure of the news items, leveraged information, metadata and vocabulary [4][14][16][29]. Most recent JKPs also include dedicated storage for real-time news-related feeds [22].

1.6. Concerns

Stakeholders, information, functionalities, techniques and components are influenced or affected by additional concerns of various types. **Table 4** lists the identified concerns.

Table 4. Concerns related to JKPs.

Aspect	Explanation
Customers heterogeneity	The diversity of newsroom customers.
Standards	Standards like IPTC topics or RDF.
Ownership	Copyrights, authorship and licensing information.
Multilingual content	Content produced in various languages.
Timeliness	The temporal aspect of news, when they are published and when the stories happen.
Human factors	Human-related aspects that affect newsroom and JKPs.

Aspect	Explanation
Quality	The information and data quality.
Big data	Aspects related to the large volume of data, variety of data and velocity in which data is produced.
Performance	The ability to provide results with the expected quality and on time.
Legacy	Old systems or repositories.
Software architecture	The structure and components of a software system [36].
Maintenance	The ability to reuse, fix and update existing systems.

companies and institutions, and use different systems to interact with JKPs [16][22]. To improve the interoperability between news agencies and stakeholders, JKPs utilise standards like the IPTC news codes, media topics, semantic vocabularies and RDF [4][16], and keep track of information related to ownership, such as authorship, copyrights, privacy and sources [2][37]. JKPs can also use the ownership information to control the information provenance and reliability [5] by, for example, tracking back the information to its original source and identifying trustworthy providers.

JKPs attempt to address different human factors in newsrooms. JKPs automate error-prone and time-consuming processes that were performed manually like news tagging, source monitoring, information filtering, verification, fact-checking and finding related articles and relevant information [4][7][9][11][16]. Hence, JKPs free journalists from these tedious tasks and improve their results. As a result, JKPs facilitate high-quality information to meet the standards of their stakeholders [2].

On the technical side, JKPs deal with big data requirements like volume, velocity, variety. Hence, the components of JKPs are designed considering their performance to minimise the processing and distribution times [2][5]. JKPs also integrate legacy components and facilitate interoperability with other systems and external services [6][14][16][20][30]. All these factors make the software architecture of JKPs complex and difficult to maintain without guidance.

2. Future Directions for Research on JKPs

2.1. Implications for Research

2.1.1. Stakeholders

Studies on understanding how journalists embrace digital tools can aid in better adapting JKPs to the way journalists work. Such studies should consider the journalists' perceptions on using intelligent systems for creating news, how journalists process and use background information and the journalists' experiences working with AI, etc. Along these lines, related studies have been proposed, but not limited to, the journalists' usage of social media for gathering and verifying information [38][39] and the relation of the journalism practices and AI [40][41]. Similar user-oriented studies should be conducted on readers and younger and future generations of news consumers to

identify what new forms of interaction and consumption are more appealing to them. These studies could consider, for example, the readers' perceptions of automated journalism [42][43] and young people's engagement with news recommendations [44].

2.1.2. Information

To date, the knowledge extraction and recognition of entities from images and videos remain limited. Due to that, JKPs are not able to capture enough information from multimedia news. Promising directions for extracting knowledge from multimedia sources are multimodal machine learning approaches [45] that combine different types of data such as visual and text representations [46][47] and spoken language understanding tasks that analyse and detect audio speech [48]. Another limitation for knowledge extraction is the dark entities (i.e., those entities that do not exist yet in the knowledge base) [49][50]. Fresh stories about newer facts are the most attractive news, therefore, the chances of finding entity representations for those newer facts in knowledge bases are low. Therefore, research on knowledge extraction from multimedia news and dark entities can improve news representation in JKPs.

2.1.3. Functionalities

Non-technical users find it difficult to perform complex searches in knowledge bases, archives and background information due to their lack of expertise. The usage of chatbots can aid user interaction using natural language [22][51]. Additional solutions that can support journalists' interaction with knowledge and information, and automate news production are text summarisation [31], automated reporting or story generation [52][53] and automatic data visualisation [54]. Augmented reality may also bring new possibilities for assisting the exploration of information using knowledge representations and LOD [55].

2.1.4. Techniques

Due to the increase in misinformation and propaganda, it is crucial for journalists and readers to detect and distinguish trustworthy information from fake and biased news. Hence, research on JKPs should include automating the detection of fake news, political bias and rumours across social media platforms and news sources [26][56]. Techniques for such purposes can benefit from research on automating fact-checking [7][27], detecting derived or copied works [11], and media and audio forensics to identify manipulated or tempered multimedia files [57][58]. In addition, identifying misinformation items before they are stored in the knowledge base can improve the data quality of JKPs. Another promising direction is the inclusion of neural-symbolic AI [59] techniques as part of the different components of JKPs. Neural-symbolic AI combines neural networks with reasoning and logic. This can facilitate the inference and deductive reasoning over the data in the JKPs and reduce the computational cost of reasoning over knowledge graphs [60].

2.1.5. Components

In addition to automatic techniques for verification and fact-checking, promising collaborative tools for news and social media verification that involve journalists and readers [61] should be considered. Some of these tools such as WeVerify employ blockchain and knowledge graphs services for recording debunked claims and news. These collaborative repositories could be considered as additional information sources from which JKPs can obtain checked claims and provenance information but also contribute with verified information. Apart from this, the current JKPs are focused on in-house platforms that are typically accessed through a computer and oriented to print journalism. However, there is limited research on components that can facilitate access to the services offered by JKPs for mobile journalism [62] (i.e., journalism edited and published through smartphones and oriented towards audio-visual storytelling).

2.1.6. Concerns

There are no gold standards or methodologies to evaluate JKPs. Accordingly, research needs to include the design and study of evaluation methods for JKPs. Moreover, readers and journalists may perceive results from JKPs as less transparent and difficult to understand [63] as they are driven by AI. To improve their perception of trustworthiness and transparency, research on JKPs should consider explainable AI methods [64].

2.2. Implications for Practice

2.2.1. Stakeholders

To date, there have not been any studies on the implementation of JKPs in newsrooms. Such studies should evaluate the effectiveness, adoption and demand of JKPs. The experiences in implementing JKPs can help to draw a digitalisation path for newsrooms by providing best practices and identifying the main obstacles and solutions. This can support newsrooms with the definition of their roadmaps towards the adoption of JKPs, as it facilitates the identification of the most relevant aspects of JKPs and particular needs according to their current stage. Related studies have considered and provided guidelines for the utilisation of AI in news creation processes in a broader sense [65].

2.2.2. Information

The literature is unclear on how JKPs should best represent events and there is no general agreement on what constitutes an event [11]. Events can range from fine-grained actions like a shot, injury or a handshake between two actors [5] to bigger and broader events like the Spanish Civil War and the COVID-19 pandemic [13] or events in between like a trial process. Therefore, research on JKPs needs to define and discuss how different types of events at different granularity can co-exist in a JKP and what conceptualisations of the event are useful for specific use cases.

2.2.3. Functionalities

A better understanding of how to represent events and news items can bring new possibilities for JKPs, for example, on data analysis like measuring the popularity of people and companies [5], finding cause and effect

relations [11], and identifying newsworthy events for specific audiences and particular user' interests [8][32][66].

2.2.4. Techniques

One of the main limitations of the studied JKPs is the extraction of enough and precise information from text and multimedia to represent news stories in high detail [9][29]. JKPs use relation extraction models to extract the textual relations between the entities in news text [5][35]. However, these models are in an early research stage and the extracted relations are basic and limited for representing news [67]. Therefore, the functionalities that are based on these models must be considered for the longer term.

2.2.5. Components

Current open-source large triple-stores are not scalable and their reasoning services are time-consuming and use too many computing resources. This limits the possibilities for JKPs to exploit reasoning capabilities and analyse large knowledge graphs. Hence, scalable triple-stores and mechanisms for better reasoning over large knowledge graphs can ease the incorporation of such solutions and bring new possibilities for JKPs. A promising approach is the inclusion of entity spaces [68]. These are vector spaces that represent the different entities of a knowledge graph and also capture their semantic information. They can be used to speed up processes that require complex graph explorations like inferring and disambiguating knowledge for unseen entities. Another promising approach for integrating and managing information from different types of databases is the usage of virtual knowledge graph [69]. Virtual knowledge graphs represent the schema of the different databases and provide mechanisms for querying the databases using SPARQL, hence, it integrates databases on the schema level and reduces data replication.

2.2.6. Concerns

Only the most recent projects proposed systems to deal with big data [17][20][22]. Their architectures must also keep the machine learning models up-to-date and replace them for future best-of-breed, facilitate the schema evolution of knowledge bases and ease the expansion, distribution and independence of services [70]. Research on software reference architectures [71] for JKPs can assist in better designing and implementing them, as well as establishing a vocabulary and a framework to compare JKPs.

References

1. Beckett, C. New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence; Technical Report; Polis, London School of Economics and Political Science: London, UK, 2019.

2. Fernández, N.; Blázquez, J.M.; Fisteus, J.A.; Sánchez, L.; Sintek, M.; Bernardi, A.; Fuentes, M.; Marrara, A.; Ben-Asher, Z. NEWS: Bringing Semantic Web Technologies into News Agencies. In

Proceedings of the Semantic Web—ISWC 2006, Athens, GA, USA, 5–9 November 2006; pp. 778–791.

3. Maiden, N.; Zachos, K.; Brown, A.; Brock, G.; Nyre, L.; Nygård Tonheim, A.; Apsotolou, D.; Evans, J. Making the News: Digital Creativity Support for Journalists. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–11.

4. Castells, P.; Perdrix, F.; Pulido, E.; Rico, M.; Benjamins, R.; Contreras, J.; Lorés, J. Neptuno: Semantic Web Technologies for a Digital Newspaper Archive. In Proceedings of the Semantic Web: Research and Applications, ESWS 2004, Heraklion, Crete, Greece, 10–12 May 2004.

5. Rospocher, M.; van Erp, M.; Vossen, P.; Fokkens, A.; Aldabe, I.; Rigau, G.; Soroa, A.; Ploeger, T.; Bogaard, T. Building Event-Centric Knowledge Graphs from News. J. Web Semant. 2016, 37–38, 132–151.

6. Raimond, Y.; Scott, T.; Oliver, S.; Sinclair, P.; Smethurst, M. Use of Semantic Web technologies on the BBC Web Sites. In Linking Enterprise Data; Springer: New York, NY, USA, 2010.

7. Miranda, S.A.; Nogueira, D.; Mendes, A.; Vlachos, A.; Secker, A.; Garrett, R.; Mitchel, J.; Marinho, Z. Automated Fact Checking in the News Room. In Proceedings of the World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA; pp. 3579–3583.

8. Kalfoglou, Y.; Domingue, J.; Motta, E.; Vargas-Vera, M.; Buckingham Shum, S. myPlanet: An ontology driven Web based personalised news service. In Proceedings of the International Joint Conference on Artificial Intelligence, Washington, DC, USA, 4–10 August 2001; Volume 2001, pp. 44–52.

9. Java, A.; Finin, T.; Nirenburg, S. SemNews: A Semantic News Framework. In Proceedings of the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, Boston, MA, USA, 16–20 July 2006.

10. Borsje, J.; Levering, L.; Frasincar, F. Hermes: A Semantic Web-Based News Decision Support System. In Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08, Fortaleza, Brazil, 16–20 March 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 2415–2420.

11. Leban, G.; Fortuna, B.; Brank, J.; Grobelnik, M. Event Registry: Learning about World Events from News. In Proceedings of the 23rd International Conference on World Wide Web, WWW'14 Companion, Seoul, Korea, 7–11 April 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 107–110.

12. Liu, X.; Nourbakhsh, A.; Li, Q.; Shah, S.; Martin, R.; Duprey, J. Reuters tracer: Toward automated news production using large scale social media data. In Proceedings of the 2017 IEEE

International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 1483–1493.

13. Rudnik, C.; Ehrhart, T.; Ferret, O.; Teyssou, D.; Troncy, R.; Tannier, X. Searching News Articles Using an Event Knowledge Graph Leveraged by Wikidata. In Proceedings of the Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1232–1239.

14. Ramagem, D.B.; Margerin, B.; Kendall, J. AnnoTerra: Building an integrated earth science resource using semantic Web technologies. IEEE Intell. Syst. 2004, 19, 48–57.

15. Al-Moslmi, T.; Gallofré Ocaña, M.; Opdahl, A.L.; Tessem, B. Detecting Newsworthy Events in a Journalistic Platform. In Proceedings of the 3rd European Data and Computational Journalism Conference, Malaga, Spain, 1–2 July 2019; pp. 3–5.

16. Fernández, N.; Fuentes, D.; Sánchez, L.; Fisteus, J.A. The NEWS ontology: Design and applications. Expert Syst. Appl. 2010, 37, 8694–8704.

17. Liu, X.; Li, Q.; Nourbakhsh, A.; Fang, R.; Thomas, M.; Anderson, K.; Kociuba, R.; Vedder, M.; Pomerville, S.; Wudali, R.; et al. Reuters Tracer: A Large Scale System of Detecting & Verifying Real-Time News Events from Twitter. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, Indianapolis, IN, USA, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 207–216.

18. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: The story so far. In Semantic Services, Interoperability and Web Applications: Emerging Concepts; IGI Global: Hershey, PA, USA, 2011; pp. 205–227.

19. Kobilarov, G.; Scott, T.; Raimond, Y.; Oliver, S.; Sizemore, C.; Smethurst, M.; Bizer, C.; Lee, R. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In The Semantic Web: Research and Applications; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5554.

20. Vossen, P.; Agerri, R.; Aldabe, I.; Cybulska, A.; van Erp, M.; Fokkens, A.; Laparra, E.; Minard, A.L.; Aprosio, A.P.; Rigau, G.; et al. NewsReader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. Spec. Issue Knowl.-Based Syst. Elsevier 2016, 110, 60–85.

21. Kattenberg, M.; Beloki, Z.; Soroa, A.; Artola, X.; Fokkens, A.; Huygen, P.; Verstoep, K. Two architectures for parallel processing for huge amounts of text. In Proceedings of the Language Resources and Evaluation Conference (LREC). European Language Resources Association (ELRA), Portorož, Slovenia, 23–28 May 2016; pp. 4513–4519.

22. Germann, U.; Liepins, R.; Barzdins, G.; Gosko, D.; Miranda, S.; Nogueira, D. The SUMMA Platform: A Scalable Infrastructure for Multi-lingual Multi-media Monitoring. In Proceedings of the

ACL 2018, System Demonstrations, Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 99–104.

23. Gutierrez Lopez, M.; Makri, S.; MacFarlane, A.; Porlezza, C.; Cooper, G.; Missaoui, S. Making newsworthy news: The integral role of creativity and verification in the human information behavior that drives news story creation. J. Assoc. Inf. Sci. Technol. 2022; online version of record.

24. Deuze, M. On creativity. Journalism 2019, 20, 130–134.

25. Berven, A.; Christensen, O.A.; Moldeklev, S.; Opdahl, A.L.; Villanger, K.J. A knowledge-graph platform for newsrooms. Comput. Ind. 2020, 123, 103321.

26. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst. Appl. 2020, 153, 112986.

27. Guo, Z.; Schlichtkrull, M.; Vlachos, A. A Survey on Automated Fact-Checking. Trans. Assoc. Comput. Linguist. 2022, 10, 178–206.

28. Diakopoulos, N. Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. Digit. J. 2020, 8, 945–967.

29. Domingue, J.; Motta, E. PlanetOnto: From news publishing to integrated knowledge management support. IEEE Intell. Syst. Their Appl. 2000, 15, 26–32.

30. Germann, U.; Liepins, R.; Gosko, D.; Barzdins, G. Integrating Multiple NLP Technologies into an Open-source Platform for Multilingual Media Monitoring. In Proceedings of the Workshop for NLP Open Source Software (NLP-OSS), Melbourne, Australia, 19–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 47–51.

31. El-Kassas, W.S.; Salama, C.R.; Rafea, A.A.; Mohamed, H.K. Automatic text summarization: A comprehensive survey. Expert Syst. Appl. 2021, 165, 113679.

32. Schouten, K.; Ruijgrok, P.; Borsje, J.; Frasincar, F.; Levering, L.; Hogenboom, F. A semantic web-based approach for personalizing news. In Proceedings of the 2010 ACM Symposium on Applied Computing—SAC '10, Sierre, Switzerland, 22–26 March 2010; ACM Press: Sierre, Switzerland, 2010; p. 854.

33. Shadbolt, N.; Berners-Lee, T.; Hall, W. The Semantic Web Revisited. IEEE Intell. Syst. 2006, 21, 96–101.

34. Paikens, P.; Barzdins, G.; Mendes, A.; Ferreira, D.C.; Broscheit, S.; Almeida, M.S.; Miranda, S.; Nogueira, D.; Balage, P.; Martins, A.F. SUMMA at TAC Knowledge Base Population Task 2016. In Proceedings of the Ninth Text Analysis Conference (TAC), Gaithersburg, MA, USA, 14–15 November 2016.

35. Al-Moslmi, T.; Gallofré Ocaña, M. Lifting News into a Journalistic Knowledge Platform. In Proceedings of the CIKM 2020 Workshops, Galway, Ireland, 19–23 October 2020.

36. Garlan, D. Software Architecture. In Encyclopedia of Software Engineering; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008.

37. Gallofré Ocaña, M.; Al-Moslmi, T.; Opdahl, A.L. Data Privacy in Journalistic Knowledge Platforms. In Proceedings of the CIKM 2020 Workshops, Galway, Ireland, 19–23 October 2020.

38. Neuberger, C.; Nuernbergk, C.; Langenohl, S. Journalism as Multichannel Communication. J. Stud. 2019, 20, 1260–1280.

39. Zhang, X.; Li, W. From Social Media with News: Journalists' Social Media Use for Sourcing and Verification. J. Pract. 2020, 14, 1193–1210.

40. Stray, J. Making Artificial Intelligence Work for Investigative Journalism. Digit. J. 2019, 7, 1076–1097.

41. Broussard, M.; Diakopoulos, N.; Guzman, A.L.; Abebe, R.; Dupagne, M.; Chuan, C.H. Artificial Intelligence and Journalism. J. Mass Commun. Q. 2019, 96, 673–695.

42. Graefe, A.; Bohlken, N. Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. Media Commun. 2020, 8, 50–59.

43. Tandoc, E.C., Jr.; Yao, L.J.; Wu, S. Man vs. Machine? The Impact of Algorithm Authorship on News Credibility. Digit. J. 2020, 8, 548–562.

44. Swart, J. Experiencing Algorithms: How Young People Understand, Feel About, and Engage with Algorithmic News Selection on Social Media. Soc. Media Soc. 2021, 7, 20563051211008828.

45. Guo, W.; Wang, J.; Wang, S. Deep Multimodal Representation Learning: A Survey. IEEE Access 2019, 7, 63373–63394.

46. Mogadala, A.; Kalimuthu, M.; Klakow, D. Trends in integration of vision and language research: A survey of tasks, datasets, and methods. J. Artif. Intell. Res. 2021, 71, 1183–1317.

47. Chen, S.; Aguilar, G.; Neves, L.; Solorio, T. Can images help recognize entities? A study of the role of images for Multimodal NER. In Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021), Online, 11 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 87–96.

48. Shon, S.; Pasad, A.; Wu, F.; Brusco, P.; Artzi, Y.; Livescu, K.; Han, K.J. SLUE: New Benchmark Tasks For Spoken Language Understanding Evaluation on Natural Speech. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 7927–7931.

49. van Erp, M.; Ilievski, F.; Rospocher, M.; Vossen, P. Missing Mr. Brown and buying an Abraham Lincoln—Dark entities and DBpedia. In Proceedings of the Third NLP & DBpedia Workshop, Bethlehem, PA, USA, 11 October 2015; pp. 81–86.

50. Al-Moslmi, T.; Gallofré Ocaña, M.; Opdahl, A.L.; Veres, C. Named entity extraction for knowledge graphs: A literature overview. IEEE Access 2020, 8, 32862–32881.

51. Luo, B.; Lau, R.Y.; Li, C.; Si, Y.W. A critical review of state-of-the-art chatbot designs and applications. WIREs Data Min. Knowl. Discov. 2022, 12, e1434.

52. Miroshnichenko, A. AI to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is "Yes"). Information 2018, 9, 183.

53. Alhussain, A.I.; Azmi, A.M. Automatic Story Generation: A Survey of Approaches. ACM Comput. Surv. 2021, 54, 1–38.

54. Zhu, S.; Sun, G.; Jiang, Q.; Zha, M.; Liang, R. A survey on automatic infographics and visualization recommendations. Vis. Inform. 2020, 4, 24–40.

55. Lampropoulos, G.; Keramopoulos, E.; Diamantaras, K. Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: A review. Vis. Inform. 2020, 4, 32–42.

56. Zhou, X.; Zafarani, R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Comput. Surv. 2020, 53, 1–40.

57. Pasquini, C.; Amerini, I.; Boato, G. Media forensics on social media platforms: A survey. EURASIP J. Inf. Secur. 2021, 2021, 1–19.

58. Bhagtani, K.; Yadav, A.K.S.; Bartusiak, E.R.; Xiang, Z.; Shao, R.; Baireddy, S.; Delp, E.J. An Overview of Recent Work in Media Forensics: Methods and Threats. arXiv 2022, arXiv:2204.12067.

59. Hitzler, P.; Bianchi, F.; Ebrahimi, M.; Sarker, M.K. Neural-symbolic integration and the Semantic Web. Semant. Web 2020, 11, 3–11.

60. Hitzler, P.; Krotzsch, M.; Rudolph, S. Foundations of Semantic Web Technologies; CRC Press: Boca Raton, FL, USA, 2010.

61. Thomson, T.; Angus, D.; Dootson, P.; Hurcombe, E.; Smith, A. Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities. J. Pract. 2020, 16, 1–25.

62. Salzmann, A.; Guribye, F.; Gynnild, A. "We in the Mojo Community"—Exploring a Global Network of Mobile Journalists. J. Pract. 2021, 15, 620–637.

63. Shin, D. Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism. J. Stud. 2021, 22, 1047–1065.

64. Kaur, D.; Uslu, S.; Rittichier, K.J.; Durresi, A. Trustworthy Artificial Intelligence: A Review. ACM Comput. Surv. 2022, 55, 1–38.

65. Lopez, M.G.; Porlezza, C.; Cooper, G.; Makri, S.; MacFarlane, A.; Missaoui, S. A Question of Design: Strategies for Embedding AI-Driven Tools into Journalistic Work Routines. Digit. J. 2022, 10, 1–20.

66. Motta, E.; Daga, E.; Opdahl, A.L.; Tessem, B. Analysis and Design of Computational News Angles. IEEE Access 2020, 8, 120613–120626.

67. Yan, Y.; Sun, H.; Liu, J. A Review and Outlook for Relation Extraction. In Proceedings of the 5th International Conference on Computer Science and Application Engineering, CSAE 2021, Sanya, China, 19–21 October 2021; Association for Computing Machinery: New York, NY, USA, 2021.

68. van Erp, M.; Groth, P. Towards Entity Spaces. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 2129–2137.

69. Xiao, G.; Ding, L.; Cogrel, B.; Calvanese, D. Virtual Knowledge Graphs: An Overview of Systems and Use Cases. Data Intell. 2019, 1, 201–223.

70. Gallofré Ocaña, M.; Opdahl, A.L. Developing a Software Reference Architecture forJournalistic Knowledge Platforms. In Proceedings of the ECSA2021 Companion Volume, Växjö, Sweden, 13–17 September 2021.

71. Martínez-Fernández, S.; Ayala, C.P.; Franch, X.; Marques, H.M. Benefits and drawbacks of software reference architectures: A case study. Inf. Softw. Technol. 2017, 88, 37–52.