Towards Ontological Support for Journalistic Angles

Andreas L. Opdahl and Bjørnar Tessem

Dept. of Information Science and Media Studies,
University of Bergen, N-5020 Bergen, Norway
{Andreas.Opdahl,Bjornar.Tessem}@uib.no,
http://www.uib.no/en/rg/ssis

Abstract. Journalism relies more and more on information and communication technology (ICT). New journalistic ICT platforms continuously harvest potentially news-related information from the internet and try to make it useful for journalists. Because the information sources and formats vary widely, *knowledge graphs* are emerging as a preferred technology for integrating, enriching, and preparing journalistic information. The paper explores how journalistic knowledge graphs can be augmented with support for *news angles*, in order to help journalists detect newsworthy events and present them in ways that will interest the intended audience. We argue that finding newsworthy angles on news-related information is important as an example of a more general problem in information science: that of finding the most interesting events and situations in big data sets and presenting those events and situations in the most interesting ways.

Key words: Computational journalism, ICT tool for journalists, news platforms, newsroom systems, knowledge graphs, ontology

1 Introduction

Journalism relies more and more on computers and the internet [17]. News platforms such as Event Registry [14], Reuters Tracer [15], and Bloomberg's knowledge graph [29] continuously harvest potentially news-related information from the internet and try to make it useful for journalists. Because the information sources and formats vary widely, *knowledge graphs* and related semantic technologies [1] are emerging as preferred solutions for integrating, enriching, and preparing journalistic information. Semantic technologies support information *integration* because they offer a standard Resource Description Format (RDF) for representing and exchanging facts [1]. They support information *enrichment* because they represent resources — such as concepts and concrete objects (people, organisations, locations, works...) — using standard IRIs that provide access to further information about the resources. And they support reasoning techniques, for example using OWL DL or rule languages, that can be used for *preparing* information for journalists.

The paper explores how journalistic knowledge graphs — represented in RDF, a type of property graph — can be augmented with support for *news angles* that can be used to detect newsworthy events and present those events in interesting ways. Finding or inventing good news angles on unfolding events is a central journalistic skill, which we seek to formalise in order to help journalists with: responding quickly to newsworthy events; identifying appropriate angles on those events; and backing those angles up with relevant information. Examples of angles are *conflict*, *local person*, and *fall from grace*. Some angles are more detailed versions of others, such as *David-versus-Goliath*, a subtype of *conflict*. The paper proposes OWL ontologies that can be used to organise knowledge graphs that support news angles. We ask: *how can ontologies be used to organise journalistic knowledge graphs and augment them to support news angles.* To answer this question, the rest of the paper is organised as follows: Section 2 reviews existing work. Section 3 proposes suitable ontologies. Section 4 discusses our approach, before Section 5 concludes the paper and offers paths for further work.

2 Existing Work

2.1 Computational and data journalism

There are different ways to use computational resources in journalism [25, 3]: precision journalism, computer-assisted journalism, data journalism, database journalism, data-driven journalism, and finally computational journalism. The latter is characterised by its focus on computation and software as driving tools for creating journalistic content, whereas data journalism places the journalist in the driver's seat in the creation and presentation of content [25]. Following this typology, our paper presents a *computational journalism* approach, albeit one that aims at supporting journalists rather than automating journalism.

2.2 AI-support for journalism

Artificial intelligence (AI) in journalism can be diveded into four areas: data mining, topic selection, commentary moderation, and news writing [19]. Commercial companies such as Narrative Science and Automated Insights have already developed journalistic robots that automatically generate news reports in areas such as finance and sports [13]. In 2016 alone, Automated Insight's Wordsmith tool wrote and published 1.5 billion news reports, possibly more than all the human journalists in the world [19].

2.3 News platforms

Recent developments in AI have been driven in part by the availability of *big and open data sources* that are relevant for journalism. For example, researchers have investigated how news events can be extracted from big-data sources such as Tweets [12] and other texts [10].

We define a *news platform* [3] as an integrated system that continually *harvests* potentially news-related information from a variety of sources, *integrates* the information, *prepares* it for journalistic use, and *provides* potentially relevant information to journalists or the general audience, whether passively on demand or proactively through event detection. Reuters Tracer [15] is a news platform with similar goals to ours. It targets journalists, but does not use knowledge graphs and ontologies and does not support angles.

2.4 Knowledge graphs for journalists

Event Registry [14] is a news platform that collects news messages and lifts them into a semantic *knowledge graph* (in RDF) in order to detect and describe news events in real time. Bloomberg's knowledge graph [29] is a similar platform. They are both based on semantic technologies, but Event Registry targets a wider audience, and neither platform supports news angles.

Beyond news platforms, researchers have used semantic technologies in other ways to make big and open data sources more readily available for journalists [27] and journalistic AI tools [7]. Fernandez et al. [4] propose an ontology for streamlining news production and distribution. Heravi et al. [9] advocate *social semantic journalism*, which uses natural-language processing (NLP) and semantic metadata together to: detect news events from socially-generated big data; verify information and its sources: identify eyewitnesses; and contextualise news events and their coverage.

2.5 News Hunter

In collaboration with Wolftech, a developer of news-production software for for the international market, our research group is developing News Hunter, a knowledge-graph based news platform for journalists [20, 2]. News Hunter is a proof-of-concept prototype that has been designed to continually harvest news items and social media messages from the web; analyse and represent them semantically in a knowledge graph; classify, cluster, and label them; enrich them with additional information from encyclopedic and other reference sources; and present them in real time to journalists as suggestions for new or updated reports.

This paper builds on previous papers that: give an overview of the previous News Hunter prototype (which did not support angles) [2]; discuss the concept of news angles and outline a suitable big-data architecture [5]; and investigate reasoning approaches for detecting news angles along with suitable ontologies [26]. Compared to a previous short paper [26], the present one: develops the ontologies further; discusses them in more detail; places them in an architectural context; and illustrates them using a real news event as a running example.

3 Ontologies

To prepare for a knowledge-graph based ICT platform for journalists, this section will present core ontologies for representing: potentially news-relevant information in semantic form; potentially newsworthy events detected and aggregated from that information; and possible news angles on those events. Building on our experiences from earlier News Hunter prototypes, this section will present the three corresponding ontologies and outline the roles they will play in the augmented News Hunter platform. For each ontology, we will first explain the *role* it plays in the News Hunter architecture; then the *ontology* itself and its central terms; the *processing* techniques that can be used to populate and analyse it; and finally an *example* graph in RDF, serialised using Turtle notation.

3.1 News items

News Hunter will continuously harvest potentially news-relevant information *items* from a variety of sources in different formats. So far, we have explored harvesting of: messages from social media like Facebook and Twitter; articles from newspapers on the web; and items from RSS. But relevant information items are available from a much wider range of sources that include: commercial news services like AP and Reuters; the home pages of commercial companies and public authorities; and the Internet of Things (IoT). We have so far focussed on textual items, but strive to develop a platform that is open to also include images, audio, and video in the future.

Architectural role: Harvested items are first filtered. The ones that are deemed potentially news-related are *lifted* into semantic form and represented as *item (sub-)graphs* of the central knowledge graph. A driving idea behind News Hunter is that the graph may facilitate reasoning that goes beyond standard text-based similarity searches: detecting and populating news angles is one example. Nevertheless, we also store each filtered item closer to its original form as a JSON object in a database, indexed from the knowledge graph.

Ontology: Figure 1 shows how a potentially news-relevant **Item** is represented semantically as an item graph.[1] Each item has an **originalTitle**, an **originalText**, and a **sourceIRL** among its attributes. It has a **Person** as its **contributor**, perhaps contributing through or on behalf of a **source Agent**. The agent can be, e.g., an organisation or web site, whereas the contributor can be a natural person or a social-media handle.

The item's semantics is represented by **Annotations**, each of which contains a single piece of semantic information about the item. In Figure 1, the **Annotation** class is therefore in turn related to an **Entity** in the knowledge graph, of which there are several subtypes:

[1] This and later OWL ontologies have been created using Protege-OWL and rendered using WebVOWL [16].

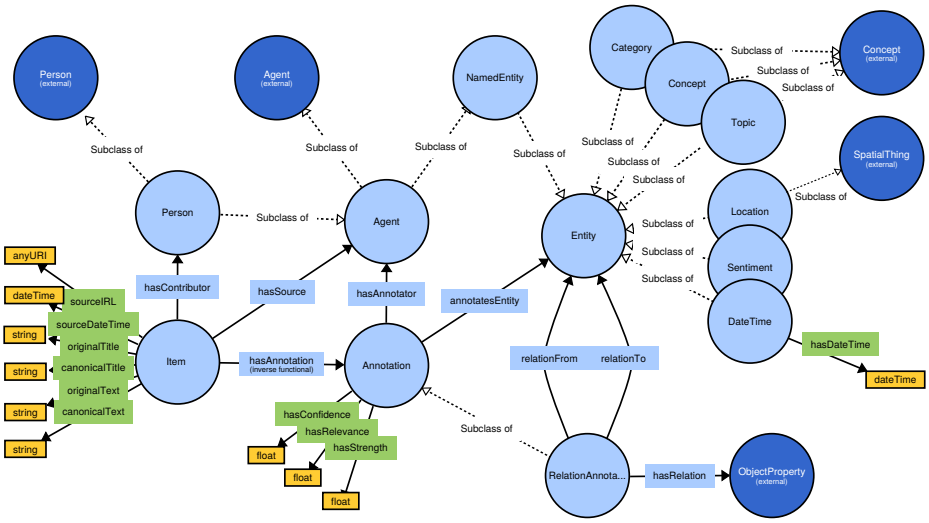

Fig. 1: Ontology for representing news items semantically as knowledge graphs.

– A **NamedEntity** mentioned in the text, possibly a named geolocation.
– A **Concept**, **Topic**, or **Category** reflected in the text, all of them subtypes of **skos:Concept**. The difference is that a concept must be a word or phrase used in the text, whereas topics and categories can be latent. Categories must also be taken from a restricted vocabulary, such as the IPTC media topics [11].
– A **Location** (**geo:SpatialThing**) or a **DateTime** (**xsd:dateTime**) associated with the text.
– A **Sentiment** reflected in the text.

Each instance of these classes (**NamedEntity**, **Concept**, **Topic**, **Category**, **Location**, **DateTime**, and **Sentiment**) has an IRI and can be extended with triples from the Linked Open Data (LOD) cloud and from proprietary resources. A **RelationAnnotation** can even represent a semantic relation between a pair of entities annotating the same item, using the **hasRelation** property to indicate the semantic relationship intended.

Each annotation has a **confidence** that expresses trust in its sources and a **relevance** that expresses how semantically central it is to the item. A relation annotation can also have a **strength** that expresses how forcefully the relation holds, if it is of a graded type. An annotation can also have a **foaf:Agent** as its **annotator**, which will usually be a piece of software or a service, such as a named entity linker or sentiment analyser. We expect that linking annotations to their annotators in this way will be useful whenever the semantic-lifting software is later improved or turns out to have been imprecise or faulty.

The news-item ontology in Figure 1 is also linked to common terms defined in other vocabularies, such as **foaf:Agent**, **skos:Concept**, and **geo:SpatialThing**. However, these are just examples. In further work, we want to align and enrich

Fig. 2: An example tweet written in Somali, announcing that President Farmajo has appointed Hassan Sheikh Ali as the new prime minister of Somalia.

Figure 1 with concepts from other semantic annotation ontologies, including the Meaning-of-a-Tag (MoaT [21]) ontology.

Processing: Lifting textual items into small knowledge graphs — or *item graphs* — shaped by Figure 1 requires natural-language processing (NLP) techniques. Our earlier prototypes [2] have explored using RAKE (Rapid Automatic Keyword Extraction) for lifting shorter messages, Textacy (a wrapper library for Spacy) for RSS feeds, and the Python-implementation of TextRank for longer texts. We are also considering more recent approaches that leverage distributed word representations (embeddings), such as the ones provided by word2vec [18], and we are reviewing tools such as FRED [6] that lift NL texts directly into knowledge graphs.

Example: Figure 2 shows a tweet posted by Universal Somali TV early in the morning on February 23rd 2017. Listing 1 shows an item graph that could result from lifting the text in this tweet, supported by the context provided by the news article it links to.[2] The tweet proclaims that President Mohamed Abdullahi Farmajo appoints Hassan Ali Khayre as the new Prime Minister of Somalia. Importantly, the lifting process has resolved the Somalian name Xasan Khayre Cali to its international counterpart: Hassan Ali Khayre. President Farmajo has been successfully resolved to a DBpedia IRI, whereas the new prime minister Khayre is not yet defined in DBpedia or Wikidata and is therefore given an internal News Hunter IRI that begins with `:unres/...` for *unresolved*.

Outside north-eastern Africa, the Somalian prime-minister appointment might not warrant prominent mention in the news. But a Norwegian newsroom could detect it as potentially newsworthy, because the newsroom's knowledge graph might already contain triples about the similar unresolved IRI `:unres/Hassan_Khaire`, shown in Listing 2, that was harvested and lifted from a YouTube video caption uploaded by The Royal House of Norway in 2010.

[2] For the purpose of the example, we have used Google Translate and adapted the outputs of IBM Watson's Natural Language Understanding service. We have enriched the resulting item graph with additional triples from DBpedia and Wikidata.

Listing 1: The tweet in Figure 2 represented semantically as an item graph.

```
:twitter834619575509594114 a :Item;
    :source dbp:dbpedia.org/resource/Universal_Television_(Somalia);
    :sourceIRL "https://twitter.com/ ... /834619575509594114";
    :originalText "WAR DEG DEG: Madaxweyne Farmaajo oo ..."@so;
    :canonicalText "PRESS RELEASE: President Farmajo appointed
                    Hassan Sheikh Ali as new prime minister";
    :hasAnnotation [    a :Annotation;
                        :hasEntity dbp:Mohamed_Abdullahi_Farmajo;
                        :hasConfidence 0.9;
                        :hasRelevance 0.33;
                        :hasStrength 1.0                        ];
    :hasAnnotation [    a :Annotation;
                        :hasEntity :unres/Hassan_Ali_Khayre     ];
    :hasAnnotation [    a :RelationAnnotation;
                        :hasRelation wn:appoint;
                        :relationFrom dbp:Mohamed_Abdullahi_Farmajo;
                        :relationTo :unres/Hassan_Ali_Khayre    ];
    :hasAnnotation [    a :RelationAnnotation;
                        :hasRelation wde:Q14212;
                        :relationFrom :unres/Hassan_Ali_Khayre;
                        :relationTo dbp:Prime_minister          ]
```

3.2 News events

To represent potentially newsworthy *events*, the individual item graphs must be clustered, merged, and enriched to form *event (sub-)graphs* of the central knowledge graph. Because they are aggregated, event graphs provide more complete and precise information than individual item graphs, many of which may only describe a small part or aspect of an event. Event graphs are also corroborated by more sources, which is particularly important for social-media messages that originate from less known contributors and whose annotations may have low confidence.

Architectural role: Items are clustered into event graphs according to their annotations, such as their named entities, concepts/topics/categories, locations and date-times, most of which will be shared by many item sub-graphs. Clustering can take into account item annotations that are identical as well as related: either semantically, for example through taxonomical or mereological relations, or lexically, for example using Levenshtein distance or similar measures to detect different spellings of the same name. To the extent possible, cluster detection should also identify how larger events are composed of sub-events with temporal, causal, and other relations between them. Annotation entities and relations from item graphs in the same cluster are then merged to form the event graph, whose entities can be enriched with further triples taken from the Linked Open Data

Listing 2: An existing item graph with facts about Hassan Khaire harvested from YouTube's video descriptions dated 2010.

```
:youtubetMkyoqpM4Pc a :Item;
    :source dbp:Norwegian_royal_family;
    :sourceIRL "https://www.youtube.com/watch?v=tMkyoqpM4Pc";
    :originalText "Intervju med Hassan Khaire ..."@no;
    :canonicalText "Interview with Hassan Khaire ...";
    :hasAnnotation [    a :Annotation;
                        :hasEntity :unres/Hassan_Khaire            ];
    :hasAnnotation [    a :Annotation;
                        :hasEntity dbp:Norwegian_Refugee_Council   ];
    :hasAnnotation [    a :RelationAnnotation;
                        :hasRelation wdp:P39;
                        :relationFrom :unres/Hassan_Khaire;
                        :relationTo dbp:Norwegian_Refugee_Council  ];
    :hasAnnotation [    a :RelationAnnotation;
                        :hasRelation dbo:location;
                        :relationFrom dbp:Norwegian_Refugee_Council;
                        :relationTo dbr:Norway                     ]
```

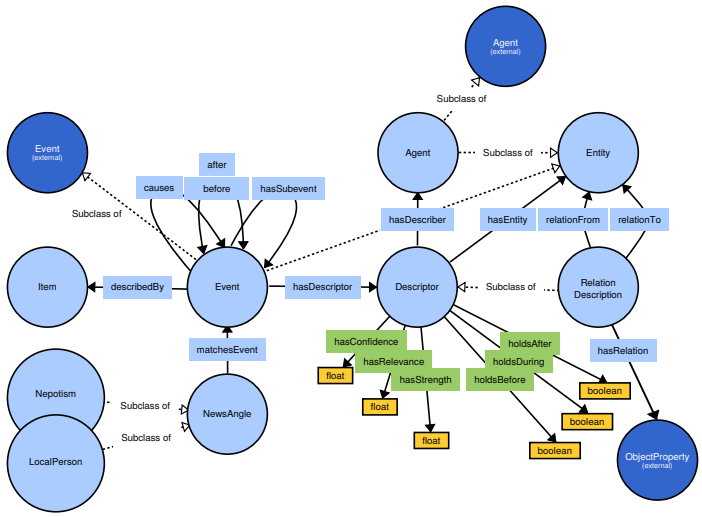

Fig. 3: Ontology for representing news events semantically as knowledge graphs.

(LOD) cloud and other sources, either by linking to external knowledge graphs or by downloading and inserting RDF triples into the event graph.

Ontology: Figure 3 shows how a potentially newsworthy **Event** is represented semantically as an event graph. Each **Event** is **describedBy** one or more **Items** that it has been derived from. It can come *before* or *after* and it can *cause* other

events, and it can have *subevents*. The semantics of an **Event** is represented in further detail by **Descriptors**, each of which contains a single piece of semantic information about the event. Analogously to item annotations, each **Descriptor** is further related to an **Entity** with subtypes as in Figure 1. There are also **RelationDescriptors** that represent semantic relationships between pairs of entities in the same event graph.

Figure 3 also shows how event **Descriptors** have *confidence*, *strength*, and *relevance* values in the same way as item annotations. In addition, **Descriptors** can hold *before*, *during*, and/or *after* the **Event**.

Pointing forward to the next section, an event can *match* one or more **NewsAngles**, of which two subtypes are shown: **LocalPerson** and **Nepotism**. They will be explained in Section 3.3. In further work, we want to align and enrich Figure 3 with concepts from other event ontologies and frameworks, such as the Event Ontology [3], the ACE framework [4], and the Simple Event Model (SEM [28]), and Eso [23].

Processing: Simple clustering of item graphs by annotation similarity is straightforward. An earlier prototype used Scikit-learn's DBSCAN algorithm, which offers scalability and focus on neighbourhood size at the expense of uneven cluster sizes [2]. Other researchers have investigated detection of events in knowledge graphs [22], as well as relations between events. Merging entities and relations from item graphs belonging to the same event is also straightforward. An earlier prototype enriched the knowledge graph with DBpedia triples [2]. Wikidata, the triple-oriented fact database behind Wikipedia and its sister projects, is a more recent alternative. Compared to DBpedia, it offers a more uniform ontology and property-level provenance.

Example: In the example from Listing 1, Universal Somali TV could be a trusted source whose news item might suggest a new event without further corroboration. But confidence in the new event would increase as news items from other sources independently reported the same information. Listing 3 shows an event graph that might result from enriching the facts in Listing 1 with facts from external sources like DBpedia and Wikidata and from the related item graph shown in Listing 2, assuming that the similar-looking IRIs for Hassan Ali Khayre have been resolved to the same individual.

3.3 News angles

Some exceptional events are newsworthy in themselves. But most events have to be made newsworthy by presenting them according to a *news angle*. Matching event graphs with news angles is a bi-directional process, in which the core facts of the event suggest candidate news angles and the candidate news angles in turn encourage additional facts to be sought, whether manually or by automated means.

[3] http://motools.sourceforge.net/event/event.html
[4] https://www.ldc.upenn.edu/collaborations/past-projects/ace

Listing 3: An event graph that represents the appointment of the new prime minister of Somalia, along with his relation to Norway.

```
[   a :Event;
    :describedBy :twitter834619575509594114;
    :hasDescriptor [ a :Descriptor;
                     :hasEntity dbp:Mohamed_Abdullahi_Farmajo   ];
    :hasDescriptor [ a :Descriptor;
                     :hasEntity :unres/Hassan_Ali_Khayre        ];
    :hasDescriptor [ a :Descriptor;
                     :hasRelation wn:appoint;
                     :relationFrom dbp:Mohamed_Abdullahi_Farmajo;
                     :relationTo :unres/Hassan_Ali_Khayre       ];
    :hasDescriptor [ a :RelationDescriptor;
                     :hasRelation wde:Q14212;
                     :relationFrom :unres/Hassan_Ali_Khayre;
                     :relationTo dbp:Prime_minister;
                     :onlyAfter xsd:true                        ];
    :hasDescriptor [ a :Descriptor;
                     :hasEntity dbp:Norwegian_Refugee_Council   ];
    :hasDescriptor [ a :RelationDescriptor;
                     :hasRelation wdp:P39;
                     :relationFrom :unres/Hassan_Ali_Khayre;
                     :relationTo dbp:Norwegian_Refugee_Council  ];
    :hasDescriptor [ a :RelationDescriptor;
                     :hasRelation dbo:location;
                     :relationFrom dbp:Norwegian_Refugee_Council;
                     :relationTo dbr:Norway                     ];
    :hasDescriptor [ a :RelationDescriptor;
                     :hasRelation :basedNear;
                     :relationFrom :unres/Hassan_Ali_Khayre;
                     :relationTo dbr:Norway                     ]
]
```

Although news angles and values are common journalistic ideas mentioned in many text books, e.g., [24, p. 115], they have not yet, to our knowledge, been analysed in depth from a knowledge representation and reasoning perspective. As a starting point, we have compiled a list of angles from academic textbooks [24] and web sites[5]. Table 1 lists examples of potential angles on the tweet from Figure 2.

Architectural role: News angles are important both for detecting newsworthy events and for presenting them in ways that may interest the intended audience.

[5] Brad Phillips, December 10th 2014: `https://www.prdaily.com/Main/Articles/16_story_angles_that_reporters_relish_17748.aspx` ; Wesley Upchurch, September 1st 2018: `http://www.streetdirectory.com/etoday/ten-common-news-angles-for-media-releases-uuofou.html`

Table 1: Alternative news angles on the tweet from Figure 2.

Event:	***President Farmajo has appointed Hassan Ali Khaire as the new prime minister of Somalia.***
Human interest:	"Sheikh Ali was forced to leave his home country as a young man."
Proximity:	"Hassan Ali Khaire has lived as a refugee here in Vestre Slidre."
Actionability:	"Join our congratulations of the new Somalian prime minister!"
Influence:	"Khaire inherits a decade-long destabilising conflict with Ethiopia."
Milestone:	"Next year marks the 30th anniversary of the first peace treaty between Somalia and Ethiopia."
Conflict:	"Farmajo and Khaire's clans clashed during the southern unrest."
Recency:	"Khaire was not thought to be a contender for this position."

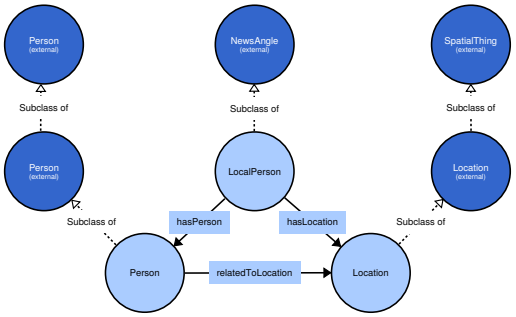

Fig. 4: Ontological representation of the local-person news angle.

A news angle can be understood as a pattern with which an event graph can potentially be matched and which offers one or more extended patterns according to which the event graph can be enriched in interesting ways. The part of an event graph that matches a news angle becomes a *fabula (sub-)graph*. The term *fabula* adopted from literary theory [8] to denote *the facts that a story is about* in contrast to the *narrative*, which denotes the presentation of those facts as a *story*. Although our representations of news angles and fabulas might support automatic narration as well, our work in News Hunter is currently limited to proposing angled events as fabulas to journalists as an aid.

Ontology: Figure 4 shows how the **LocalPerson** news angle can represented in OWL. It is a particularly simple angle, matched whenever a central **Person** in an event graph is **relatedToLocation** to a particular **Location** that is of importance to the journalist's intended audience. Figure 3 already showed how such an **Event** can be **matched** by a **NewsAngle** to form a fabula. It is possible that Figure 4 and our other news-angle ontologies will need to be supplemented with additional rules and constraints in further work, perhaps using domain-specific modelling notations on top of our ontological approach.

Figure 5 illustrates a more complex news angle, that of **Nepotism** [26], in which a **PowerfulPerson controls** a **Value** which a **GainingPerson** achieves **ac-**

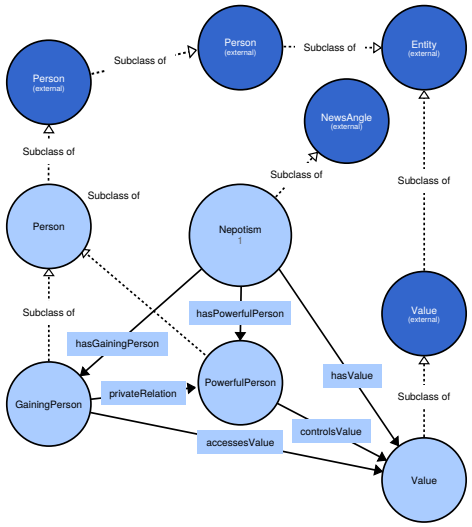

Fig. 5: Ontological representation of the nepotism news angle.

cess to because of her/his **privateRelation**, typically a **familyRelation**, to the **PowerfulPerson**. Because neoptism proper requires causality, the angle in Figure 5 represents a weaker *potential nepotism* that mandates further investigation by journalists.

Processing: Because they may involve identical or taxonomically related classes and relations, the library of news angles will form a more or less connected news-angle ontology. The central classes and relations in this slowly evolving ontology will suggest what to look for in order to identify angles in the much more rapidly changing event sub-graphs. Selecting the best or most promising matches between events and news angles remains an issue for further work. Most likely, it will depend on a combination of factors that include recency and reliability of the event and originality and fit of the angle.

We envisage a News Hunter architecture in which many collaborating agents specialise in maintaining and leveraging specific classes and relations in the connected news-angle ontology, continuously looking for changes that contribute towards enabling or disabling particular angles in response to unfolding events. For example, a *local-person* agent would specialise in deriving **Person-relatedToLocation-Location** triples from semantically related facts in the knowledge graph, including **basedNear** and **location** facts.

Example: Listing 4 shows the *fabula graph* that results from matching the facts in Listing 3 with the news angle in Figure 4. This graph comprises only a single triple, possibly derived by a *local-person agent* from triples stating that Khayre has worked for the Refugee Council located in Norway. Although the graph is simple, its triple is important as it forms the core fabula of the angled news report, to which interesting related facts from the LOD cloud could be added.

Listing 4: The small fabula graph that results from matching the facts from the Somalian prime minister tweet with the local-person news angle.

```
[   a :LocalPerson;
    :hasPerson unres/Hassan_Ali_Khayre;
    :hasLocation dbr:Norway;
] .
:unres/Hassan_Ali_Khayre :relationToLocation dbp:Norway
```

4 Discussion

We have proposed OWL ontologies that can be used to organise journalistic knowledge graphs and augment them with support for news angles. To the best of our knowledge, this is the first attempt to analyse and represent news angles as OWL ontologies. We also think that developing a news platform that supports news angles is new, and we suggest for the first time how ontologies for annotations, events and news angles can be combined in a journalistic platform.

We expect the proposed ontologies to evolve as we develop the News Hunter proof-of-concept architecture and prototype. Additional ontologies will also be needed, for example to: organise different types of input items; represent available analysis techniques and tools; propagate information about provenance/confidence and terms-of-use; reason about privacy; describe editorial and journalistic preferences; etc. Although we have presented them separately in this paper, we see the ontologies merely as alternative thematic windows into a single logically contiguous, but perhaps physically distributed, knowledge graph.

Our paper has focussed on ontology — and thus knowledge-graph — structure, leaving architectural and algorithmic issues to parallel [5, 26] and future work. Our ontology structure can potentially also shape News Hunter's processing structure, so that different sub-systems, perhaps implemented as collaborating agents, can take responsibility for different ontologies. For example, one group of agents would lift data into item graphs, another would collate item graphs into events, a third would match event graphs with news angles, etc. In such an architecture, the knowledge graph would be split by ontology into contiguous and sometimes overlapping *smart graphs* that comprise both ontology definitions (TBox) and corresponding RDF triples (ABox), and associated software agents with responsibility for maintaining and leveraging the triples.

5 Conclusion

In a world of ever-increasing information, journalists are not the only ones facing the challenge of: finding the most interesting events and situations in big data sets and presenting those events and situations in the most interesting ways. We therefore hope our results will be useful for and inspire practice and research in other information systems areas beyond journalism and the news.

Our work on News Hunter and its parent project News Angler is just starting, and interesting paths for further work include: developing the architecture further and populating it with live and test data; collecting libraries of news angles, both manually and automatically; adapting and extending suitable analysis techniques for analysing news items, detecting and aggregating events, finding suitable news angles, and identifying causal and other relations between events; and selecting the most suitable and appropriate empirical research goals and evaluation approaches for the different parts of our project.

Acknowledgement

The News Angler project is funded by the Norwegian Research Council's IKT-PLUSS programme as project 275872.

References

1. Dean Allemang and James Hendler. *Semantic web for the working ontologist: Effective modeling in RDFS and OWL*. Elsevier, 2011.
2. Arne Berven, Ole A. Christensen, Sindre Moldeklev, Andreas L. Opdahl, and Kjetil J. Villanger. News Hunter: Building and mining knowledge graphs for newsroom systems. *NOKOBIT*, 26, 2018.
3. Nicholas Diakopoulos. Computational journalism and the emergence of news platforms. *The Routledge Companion to Digital Journalism Studies, London: Routledge, Taylor and Francis group*, 2016.
4. Norberto Fernández, Damaris Fuentes, Luis Sánchez, and Jesús A Fisteus. The NEWS ontology: Design and applications. *Expert Systems with Applications*, 37(12):8694–8704, 2010.
5. Marc Gallofré, Lars Nyre, Andreas L. Opdahl, Bjørnar Tessem, Christoph Trattner, and Csaba Veres. Towards a big data platform for news angles. In *4th Norwegian Big Data Symposium — NOBIDS 2018.*, November 2018.
6. Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. Semantic web machine reading with FRED. *Semantic Web*, 8(6):873–893, 2017.
7. Roberto García, Ferran Perdrix, Rosa Gil, and Marta Oliva. The semantic web as a newspaper media convergence facilitator. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(2):151–161, 2008.
8. Pablo Gervas. Computational approaches to storytelling and creativity. *AI Magazine*, 30(3):49–62, July 2009.
9. Bahareh Rahmanzadeh Heravi and Jarred McGinnis. Introducing social semantic journalism. *The Journal of Media Innovations*, 2(1):131–140, 2015.
10. Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska De Jong. An overview of event extraction from text. In *W'shop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011) at 10th ISWC 2011*, volume 779, pages 48–57. Citeseer, 2011.
11. International Press Telecommunications Council. Media Topics: Subject taxonomy for the media — the successor of the Subject Codes.

12. Alan Jackoway, Hanan Samet, and Jagan Sankaranarayanan. Identification of live news events using Twitter. In *Proc. 3rd ACM SIGSPATIAL Int. Workshop on Location-Based Social Networks*, pages 25–32. ACM, 2011.
13. Noam Lemelshtrich Latar. The robot journalist in the age of social physics: The end of human journalism? In *The new world of transitioned media*, pages 65–80. Springer, 2015.
14. Gregor Leban, Blaz Fortuna, Janez Brank, and Marko Grobelnik. Event Registry: Learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 107–110. ACM, 2014.
15. Xiaomo Liu, Quanzhi Li, Armineh Nourbakhsh, Rui Fang, Merine Thomas, Kajsa Anderson, Russ Kociuba, Mark Vedder, Steven Pomerville, Ramdev Wudali, 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*, pages 207–216. ACM, 2016.
16. Steffen Lohmann, Vincent Link, Eduard Marbach, and Stefan Negru. Webvowl: Web-based visualization of ontologies. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 154–158. Springer, 2014.
17. Marcel Machill and Markus Beiler. The importance of the internet for journalistic research: A multi-method study.... *Journalism Studies*, 10(2):178–203, 2009.
18. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
19. Andrey Miroshnichenko. AI to bypass creativity. Will robots replace journalists? (The answer is "yes"). *Information*, 9(7), July 2018.
20. Andreas L. Opdahl, Arne Berven, Kamal Alipour, Ole A. Christensen, and Kjetil J. Villanger. Knowledge graphs for newsroom systems. *NOKOBIT*, 24, 2016.
21. Alexandre Passant and Philippe Laublet. Meaning of a Tag: A collaborative approach to bridge the gap between tagging and linked data. *LDOW*, 369, 2008.
22. Marco Rospocher, Marieke van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37:132–151, 2016.
23. Roxane Segers, Piek Vossen, Marco Rospocher, Luciano Serafini, Egoitz Laparra, and German Rigau. Eso: A frame based ontology for events and implied situations. *Proceedings of MAPLEX*, 2015, 2015.
24. Pamela J Shoemaker and Stephen D Reese. Mediating the message: Theories of influences on mass media content, 1995.
25. Eirik Stavelin. *Computational Journalism. When journalism meets programming.* Phd Thesis, University of Bergen, 2013.
26. Bjørnar Tessem and Andreas L. Opdahl. Supporting journalistic news angles with models and analogies, 2019. Proc. IEEE RCIS 2019, Brussels.
27. Raphaël Troncy. Bringing the IPTC news architecture into the semantic web. In *International Semantic Web Conference*, pages 483–498. Springer, 2008.
28. Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136, 2011.
29. Nikos Voskarides, Edgar Meij, Ridho Reinanda, Abhinav Khaitan, Miles Osborne, Giorgio Stefanoni, Prabhanjan Kambadur, and Maarten de Rijke. Weakly-supervised contextualization of knowledge graph facts. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 765–774. ACM, 2018.