Ontologies for finding journalistic angles

Andreas L. Opdahl[1] · Bjørnar Tessem[1]

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

Keywords Computational journalism · Data journalism · Journalistic knowledge platforms · News angles · Knowledge graphs · Ontology

1 Introduction

Journalism relies more and more on computers and the Internet [28]. Journalistic platforms such as NewsReader [51], SUMMA [13], and Reuters Tracer [25] are designed to continuously harvest potentially news-related information and make it useful for journalists. More general news platforms such as Event Registry [24] and the GDELT project[1] offer similar services to a wider audience.

Journalistic knowledge platforms must be able to cope with information that arrives from a wide variety of sources and in a wide variety of formats. *Knowledge graphs* [45] and related *semantic technologies* [3] appear well suited for this task because they have been developed specifically for integrating, enriching, and processing factual information. We envisage *journalistic knowledge graphs* that continuously

---

Communicated by Jelena Zdravkovic and Iris Reinhartz-Berger.

✉ Andreas L. Opdahl
Andreas.Opdahl@uib.no
http://i2s.uib.no

Bjørnar Tessem
Bjornar.Tessem@uib.no

[1] Department of Information Science and Media Studies, University of Bergen, 5020 Bergen, Norway

integrate new factual information from both news and pre-news sources; enrich it with reference and other contextual information; and prepare it for journalistic use.

This paper explores how journalistic knowledge graphs can be augmented with support for *news angles*, a concept that refers to how journalists make news events interesting for an audience [9], for example, by emphasising and including certain facts about an event over others. Although finding good news angles on unfolding events is a central skill for journalists, they remain a journalistic practice more than a theoretical concept. To our knowledge, news angles have not been studied at a deeper, structural level, for example, from a knowledge representation and reasoning perspective. We therefore seek to formalise news angles in order to develop an ICT platform that can help journalists with tasks such as: detecting newsworthy events quickly and precisely; identifying appropriate angles on those events; and contextualising those angles with suitable background information. Examples of angles are *local person*, *conflict*, *triumph over adversity*, and *fall from grace*. Some of them are more detailed versions of others, such as *David-versus-Goliath*, a subtype of *conflict*. We will encounter more examples soon.

Specifically, the paper will propose OWL ontologies [3] that can be used to organise knowledge graphs that support journalistic news angles. Our central aim is to identify the concepts and relations that such knowledge graphs must

capture. While there are many related ontologies available, we will argue that none of them satisfy our needs completely. We ask: *how can ontologies be used to organise journalistic knowledge graphs and augment them to support news angles?* We answer this research question by working through an archival example of a real news event. We present detailed knowledge graphs and ontologies that can be used to represent news items, events, and angles. We thereby also shed light on the development of *knowledge graph-driven software systems*, an increasingly important type of system that is driven by models both on the *operational (or run-time) level* where knowledge graphs represent individuals along with their type and their relations and on the *development (or software engineering) level* where ontologies represent relations between and constraints on resource types.

The rest of the paper is organised as follows: Sect. 2 reviews related work, and Sect. 3 outlines our research method. Section 4 discusses the concept of news angles, and Sect. 5 identifies central concepts and relations and presents ontologies to represent them. Section 6 discusses our proposal and compares our ontologies to existing ones, and Sect. 7 concludes the paper and outlines paths for further work.

2 Background

2.1 Computational journalism

Thurman [47] considers computational journalism to be "the advanced application of computing, algorithms, and automation to the gathering, evaluation, composition, presentation, and distribution of news". Computational journalism can aim either to support journalists or to automate journalism. Examples of automation include robot writing [23]. In contrast, this paper aims to support journalists by relieving them from much of the low-level work of collecting, checking, and organising facts.

In either variant, computational journalism relies increasingly on *machine learning*, *natural language processing (NLP)*, and other *artificial intelligence (AI)* techniques. Miroshnichenko [31] identifies four uses of AI for journalism: data mining, topic selection, commentary moderation, and news writing. This paper aims to support data mining and topic selection in particular, by applying knowledge graphs to represent the contents of news items and the events they describe. The knowledge graphs are extracted from news items using NLP techniques and ontologies.

2.2 News Hunter

In collaboration with Wolftech Broadcast Solutions, a developer of TV news production software for the international market, our research group is developing News Hunter, a knowledge graph-based *journalistic knowledge platform* [4,33]. News Hunter has been designed to continuously: *harvest* potentially news-related information from a variety of sources; *integrate* the information; *enrich* it with additional information from encyclopaedic and other reference sources; *organise* it for journalistic use; and *provide* potentially relevant information to journalists or the general audience, whether passively on demand or proactively through event detection. Section 5.1 will review the evolving News Hunter architecture in more depth.

This paper builds on previous papers about News Hunter that: give an overview of the earlier News Hunter prototypes (which did not support angles) [4]; discuss the concept of news angles and outline a suitable big-data architecture [9]; investigate reasoning approaches for finding suitable news angles [46]; and discuss how angles can help formalising the concept of *newsworthiness* [2]. It is based on a paper presented at EMMSAD 2019 [34], which it extends in several ways by: reviewing related work more extensively; discussing annotation confidence, relevance, and strength in more detail; reviewing related ontologies more thoroughly; and showing how our proposal goes beyond existing ontologies.

2.3 Journalistic knowledge graphs

Knowledge graphs represent factual information as triples of subjects, predicates, and objects. Each subject is a material or conceptual resource, such as a person, an organisation, a place, a piece of information, or a concept. A predicate expresses a relation between the subject and the object, for example that an organisation employs a person, that a place has a name, or that a piece of information is about a concept. Hence, an object can be either a material or conceptual resource like a subject, or it can be a literal value, such as a string, number, time, or date. When a knowledge graph is represented in the Resource Description Format (RDF) [3], its subjects and predicates are represented as standardised Internationalised Resource Identifiers (IRIs), and its objects are represented either as IRIs or literals. Because the same IRI can be the subject and object of many triples, the facts form a directed graph with subjects and objects as nodes and predicates as arrows.

Example Figure 1 depicts a small knowledge graph of 6 resources (nodes) and 9 triples (arrows). Listing 1 shows the same graph in the more detailed Turtle notation we will use in the rest of the paper. The graph describes that president Mohamed Farmajo has appointed Hassan Ali Khayre as prime minister of Somalia. Khayre is a dual citizen of both Somalia and Norway. He has a geographical relation to Norway because he has resided in Vestre Slidre and worked
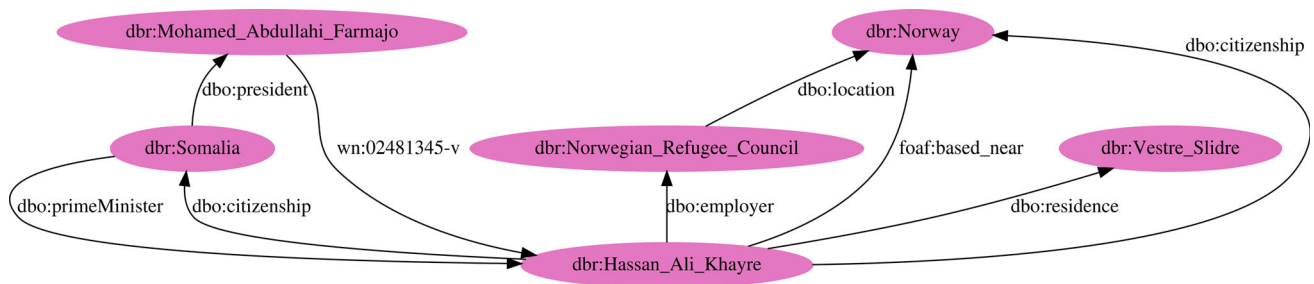
Fig. 1 A small knowledge graph describing that Hassan Ali Khayre, who has previously lived and worked in Norway, has been appointed prime minister of Somalia

Listing 1 The knowledge graph in Figure 2 represented using Turtle notation.

```
dbr:Mohamed_Abdullahi_Farmajo wn:02481345-v dbr:Hassan_Ali_Khayre .
dbr:Somalia dbo:president dbr:Mohamed_Abdullahi_Farmajo ;
           dbo:primeMinister dbr:Hassan_Ali_Khayre .
dbr:Hassan_Ali_Khayre dbo:citizenship dbr:Norway, dbr:Somalia ;
                     dbo:residence dbr:Vestre_Slidre ;
                     dbo:employer dbr:Norwegian_Refugee_Council ;
                     foaf:based_near dbr:Norway .
dbr:Norwegian_Refugee_Council dbo:location dbr:Norway .
```

for the Norwegian Refugee Council. Each resource in the graph is represented using a standard IRI defined by DBpedia (the `dbr:` prefix), and each predicate is represented using either a DBpedia IRI (`dbo:`), the Friend-of-a-Friend vocabulary (`foaf:based_near`), or the RDF version of WordNet (where `wn:02481345-v` represents a specific sense of the verb "appoint"). The graph can easily be expanded with related facts from DBpedia, Wikidata, and other data sets available in the Linked Open Data (LOD) cloud [5], a vast distributed repository of knowledge available as RDF graphs that use standard IRIs. More advanced knowledge graphs can also represent details such as the sources of facts and the time intervals during which they are valid.[2]

Whereas a prime minister appointment in Somalia might not warrant prominent mention in national news outside Eastern Africa, the connection to the Norwegian Refugee Council and to Vestre Slidre means that the core facts represented in Fig. 1 are newsworthy in Norway too. But this connection may not be easy to detect for journalists who are not knowledgeable about both countries. The rest of the paper will therefore investigate how the event in Fig. 1 could potentially have been discovered through a journalistic knowledge graph augmented with support for a news angle such as *local person*.

---

[2] For example, the RDF version of Wikidata (`wikidata.org`) includes both *references* to the sources of facts and *qualifiers* such as their start and end times.

3 Research method

Because our research on News Hunter is exploratory and involves technology development, we have framed it as *design science* [17,18], investigating journalistic knowledge platforms by developing a series of proof-of-concept prototypes [4,33]. The gist of design science research is to advance theory and improve practice by incrementally developing and evaluating one or more research artefacts. We have focussed on two artefacts: an architecture (a high-level structure of system components) and a series of instantiations (a situated implementation in a specific environment) of that architecture in the form of prototypes [49]. Hence, the ontological constructs and models and the design principles behind our architecture form a "[n]ascent design theory— knowledge as operational principles/architecture" [15]—that can be explored in further research.

To investigate how journalistic knowledge graphs can be augmented with support for news angles, we build on our earlier discussions of the concept of news angles [9]. In this paper, we proceed to work through an archival example of a real news event, developing detailed ontologies and example knowledge graphs that can be used to represent a news item and a matching news angle. For this purpose, we must first identify the central concepts and relations to build our ontologies around. Having done this, we can proceed to link our central concepts to and extend our ontologies with concepts and relations from existing ontologies. We must also ensure that our ontologies support reasoning over knowledge

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

graphs. We will validate our proposal through the running example, by showing that the ontologies we propose indeed provide the concepts necessary to represent item sub-graphs that can be collated into event graphs that can in turn be matched with represented news angles. When work on the platform has proceeded further, we plan to perform more extensive evaluations with larger data sets of automatically lifted news items, automatically detected news events, and automatically matched news angles.

4 What is a news angle?

In the journalist's daily work, they need to find a way of presenting a newsworthy event that attracts readers or viewers. A *news angle* is thus defined as *how a journalist or other news worker makes an event (or situation) interesting for an audience* [9], for example when selecting which core facts to emphasise in a news report and which contextual facts to include. When the event itself has high public interest, it may be obvious from what angle is should be reported, but in many circumstances only a creative, original angle warrants a report on the event.

News angles and values are common journalistic ideas mentioned in text books, e.g. [44, p. 115], and in the research literature [16]. When selecting a news angle, the journalist chooses features of the event to focus on, like a particular person being involved, relationships between persons and other entities in the event, or unusual qualities of some of the features of the event. We have compiled a list of angles from academic textbooks [44] and web sites[3]. These include

- Conflict: the event accentuates a conflict among people or organisations.
- Human interest: the event involves an individual who is personally affected in some way.
- Impact: the event has an effect on society or nature.
- Influence: the event changes somebody's position or status in society.

- Milestone: the event is significant in the lifetime of someone or something.
- Proximity: the event has a particular relevance to a local place.
- Recency: the event has a particular relevance to current issues.

In addition to gaining the audience's attention, a news angle serves several additional purposes:

- it provides a *criterion* for selecting events that are worth reporting;
- it points towards *additional facts* to report;
- it suggests which *information sources* to use; and
- it serves as a *template* for how to present the event.

In practice, a journalist with sufficient background knowledge may observe a tweet like the one in Fig. 2 and start to think about news angles like the ones in Table 1.

5 Central concepts and relations

To prepare for a knowledge graph-based ICT platform for journalists, this section will propose thematic (sub-)ontologies for representing: potentially news-relevant information in semantic form (Sect. 5.2); potentially newsworthy events detected and aggregated from that information (Sect. 5.3); and possible news angles on those events (Sect. 5.4). While many ontologies with related purposes have already been presented for the news domain (Sect. 6), we are not aware of existing proposals with the same aim as ours: to develop a knowledge graph-driven journalistic knowledge platform that can support news angles.

For each ontology, we will explain the *role* it plays in the News Hunter architecture; its central *concepts and relations*; its most closely *related ontologies*; the *reasoning* and other processing techniques used to populate and analyse it; and finally, an *example* graph in RDF, serialised using Turtle notation. Section 6 will review existing ontologies from the literature and show how our proposals go beyond them.

---

[3] Brad Phillips, 10 December 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 Fig. 2

Event:	President Farmajo has appointed Hassan Ali Khayre as the new prime minister of Somalia.
Human interest:	"Hassan was forced to leave his home country as a young man".
Proximity:	"Hassan Ali Khayre has lived as a refugee here in Vestre Slidre".
Actionability:	"Join our congratulations of the new Somalian prime minister!"
Influence:	"Khayre 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 Khayre's clans clashed during the southern unrest".
Recency:	"Khayre was not thought to be a contender for this position".

Fig. 3 An augmented News Hunter architecture that supports angles
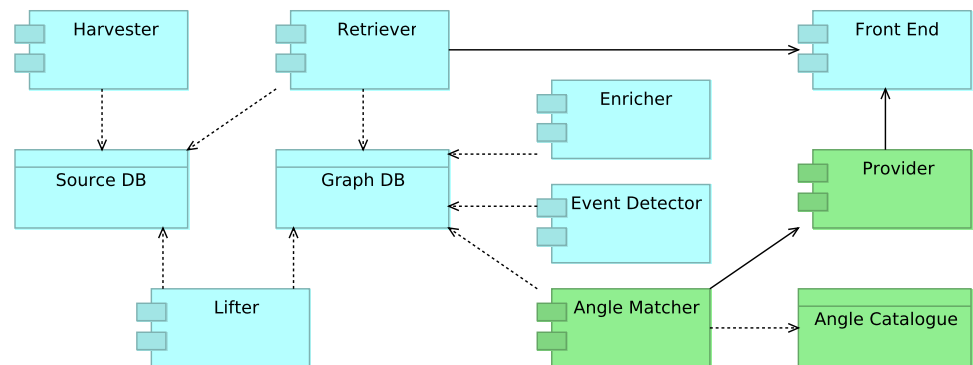
5.1 News Hunter architecture

To prepare for explaining the ontologies, however, we need to review the News Hunter architecture briefly. Figure 3 shows a simplified version of the architecture from [4] and suggests how it can be augmented with support for news angles.

The *Harvester* continuously downloads potentially newsworthy text items such as RSS messages and tweets from the net and inserts them into a *Source DB*. The *Lifter* in turn represents each text item as a small knowledge graph by invoking NLP services such as named-entity extraction, topic identification, and sentiment analysis. It uploads the resulting item graphs into a *Graph DB* (or triple store). The *Enricher* extends these item sub-graphs with additional triples retrieved from the LOD cloud. The *Front End* provides a text editor the journalists can type their stories into. The *Retriever* supplies the journalists with relevant background facts from the *Graph DB* and related text items from the *Source DB*, either on demand or proactively (by analysing the text in the editor using the *Lifter*).

At the same time, the *Event Detector* monitors the incoming item sub-graphs. When a sufficient number of similar or overlapping sub-graphs have arrived from sufficiently trustworthy sources, the event detector collates them into an event graph that is uploaded back into the *Graph DB*. The *Angle Matcher* in turn monitors the new event graphs to find ones that fit angles represented in the *Angle Catalogue*. When a fit is found, the event is considered potentially newsworthy. It is extended with appropriate background facts from the *Graph DB* according to its angle and submitted by the *Provider* to the *Front End* for consideration by the journalist. This simplified architecture has left out several components that were presented and discussed in [4] but the *Angle Matcher*, *Angle Catalogue*, and *Provider*, shown in green in Fig. 3, were not considered there.

5.2 News items

Hence, News Hunter will continuously harvest potentially news-relevant information *items* from a variety of sources and 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 potentially news-relevant texts 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; ideal and commercial news aggregators such as GDELT and WebHose[4]; and the Internet of Things (IoT). In addition to these real-time sources, it is also possible to populate our knowledge graphs with historical news items, for example taken from news archives or from encyclopaedia. We have so far focussed on textual items, but strive to develop an architecture that is open to also including images, audio, and video in the future.

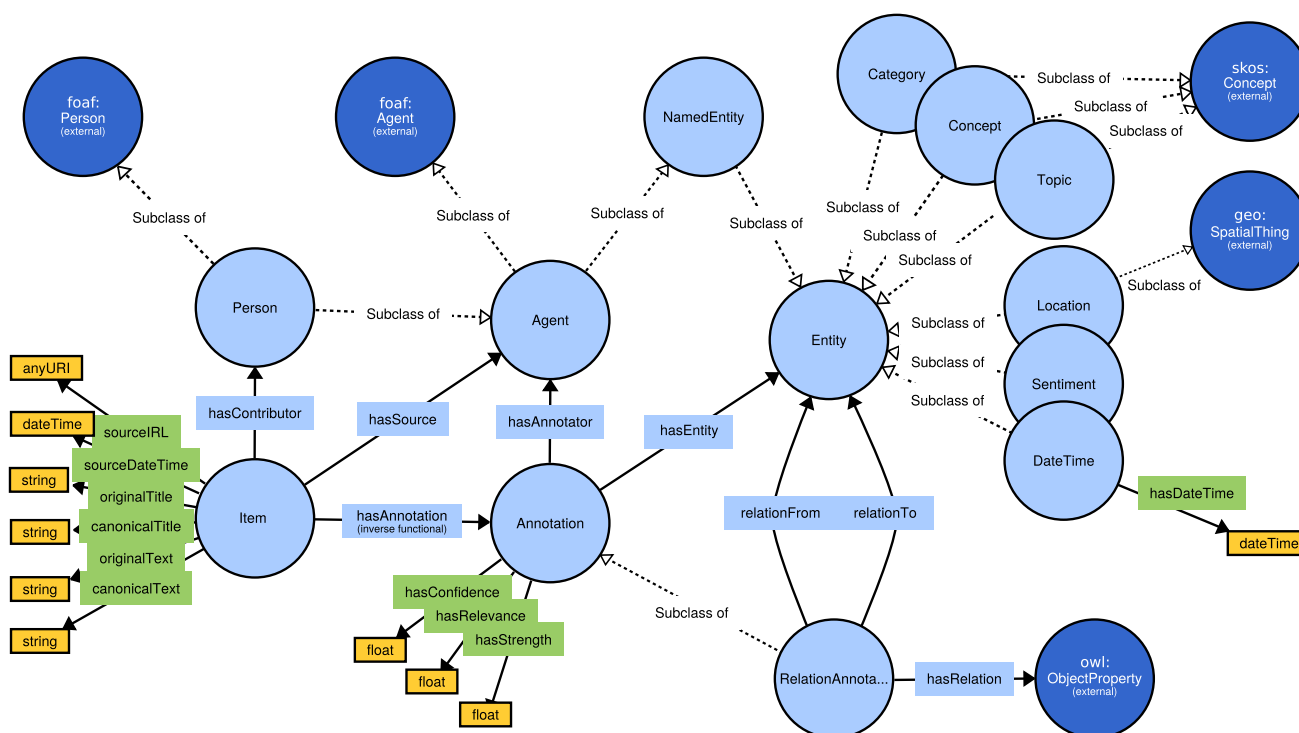
[4] http://webhose.io.

Fig. 4 Ontology for representing news items semantically in a knowledge graph

Role in the architecture Harvested items are first filtered. The ones that are deemed potentially news-relevant are then *lifted* into semantic form and represented as *item (sub-)graphs* of the central knowledge graph. Whereas standard text-based similarity searches are restricted to topics and named entities, it is a driving idea behind News Hunter to leverage the structure of this graph to facilitate more precise reasoning: the structural matching of events with news angles in this paper is one example. Nevertheless, we also store each filtered item closer to its original form as a JSON object in the source database, indexed from the knowledge graph.

Concepts and relations Figure 4 shows how a potentially news-relevant **Item** is represented semantically as an *item graph*.[5] Each item has an **originalTitle**, an **originalText**, and a **sourceIRL** among its datatype properties. It has a **Person** as its **contributor**, perhaps contributing through or on behalf of a **source Agent**. The agent can be, for example, an organisation or web site, whereas the contributor can be a natural person or a social-media handle. Although not shown in the figure, an **Agent** has got a **confidence** score normalised to the unit interval [0 : 1], representing how much the agent's items are trusted. Also not shown are the **confidence** scores of **Items**, which must be smaller than or equal to the confidence scores of their contributor.

The item's semantics is represented by **Annotations**, each of which contains a single piece of semantic information about the item or a part of it.[6] In Fig. 4, each **Annotation** is in turn related to an **Entity** in the knowledge graph, of which there are several subtypes:

– 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 are taken from a restricted vocabulary, such as the IPTC Media Topics.
– A **Location** (**geo:SpatialThing**) or a **DateTime** (**xsd:dateTime**) associated with the text.
– A **Sentiment** reflected in the text.

Each instance of these subclasses (**NamedEntity**, **Concept**, **Topic**, **Category**, **Location**, **DateTime**, and **Sentiment**) has an IRI and can be extended with facts from the Linked Open Data (LOD) cloud and from proprietary data sets. The final subclass, **RelationAnnotation**, represents semantic relations between pairs of other **Entities** that annotate

5 This and later OWL ontologies have been created using Protege-OWL and rendered using WebVOWL [27].

6 Although not shown in the figure, the EARMARK ontology [36] can be used to indicate precisely which part of the item text an annotation refers to.

Springer

the same item. Each **RelationAnnotation** is related to an **owl:ObjectProperty** that describes the type of relationship.

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. Linking annotations to their annotators in this way is needed whenever the semantic-lifting software is later improved or turns out to have been imprecise or faulty.

Furthermore, an annotation has a **confidence**, a **strength** and a **relevance**, each normalised to the unit interval $[0 : 1]$. The *confidence* describes how much trust is placed in the annotation. It is typically returned by the **annotator Agent**. When assessing the overall confidence in an annotation, both annotation confidence and item confidence must be taken into account.[7]

The *strength* describes how strongly a graded annotation applies to a news item. For example, for a sentiment annotation like **anger**, it gauges the degree of anger expressed in the text whereas, for a relation such as **likes**, it represents how strongly one entity (a prospective informant) likes another (a person in the news). Hence, **strength** is important for the meaning of graded entities and relations. For non-graded entities and relations, such as the **AngelaMerkel** individual or a **marriedTo** relation, the strength is always one.

The *relevance* describes how important a role the entity or relation plays in the item. For example, in the sentence "Blast at Rally for Afghan President Kills at Least 24", the entities "Blast", "Kills", and "24" should most likely be ranked as more relevant than "President", for example to avoid misinterpretations such as "the president was killed" or "the president killed at least 24". Hence, **relevance** captures an aspect of annotations that is important for downstream analysis. It is orthogonal to strength. For example, it may be important that one news item conveys a weak emotional reaction (high relevance, but low strength).

Related ontologies The item annotation ontology in Fig. 4 has already been linked to common terms defined in other vocabularies, such as **foaf:Agent**, **foaf:Person**, **skos:Concept**, and **geo:SpatialThing**. However, these are just examples. In further work, we want to align and enrich Fig. 4 with concepts from related ontologies, in particular from the IPTC's NewsML G2 and the BBC Ontologies. Section 6 reviews and compares Fig. 4 to these and other existing ontologies for annotating news items and other texts.

Reasoning Lifting textual items into item (sub-)graphs—small knowledge graphs shaped by Fig. 4—requires natural-language processing (NLP) techniques. Earlier prototypes [4] have explored 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 currently identifying named entities using DBpedia Spotlight [29] and Spacy-NEL[8]. Integrated tools like FRED [12] and PIKES [8] are already able to automatically lift NL texts to small knowledge graphs such as ours, and lifting techniques that use word embeddings (e.g. [30]) and deep learning (e.g. [22]) keep improving. We have presented a survey of recent named-entity extraction techniques [1] that can be used in combination with techniques for topic identification and relation extraction to represent the contents of news texts increasingly precisely as item graphs.

Example Figure 2 shows a tweet posted by Universal Somali TV early in the morning on 23 February 2017. Listing 2 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.[9] The tweet proclaims that President Mohamed Abdullahi Farmajo appoints Hassan Ali Khayre as the new Prime Minister of Somalia. Importantly, we assume that the translation and lifting steps have 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 prefix `unres:...` for *unresolved*.

Although Hassan Ali Khayre might not be a well-known person outside of Somalian politics, a knowledge graph populated over time with social-media content might already contain the triples in Listing 3, which have been harvested and lifted from the caption of a YouTube video uploaded by The Royal House of Norway in 2010. Although the IRIs are not identical, the `foaf:names` in Listings 2 and 3 are sufficiently similar for a person name resolver to make the connection, perhaps supported by other triples not shown in the listings. The knowledge graph in a Norwegian newsroom might thereby contain the information necessary to detect the prime minister appointment as potentially newsworthy due to the *local-person* connection.

5.3 News events

To represent potentially newsworthy *events* with higher confidence and in more detail, the individual item graphs must be clustered, merged, and enriched to form *event (sub-)graphs* of the central knowledge graph. Because they are aggregated,

---

[7] If the item has no confidence, the confidence of its contributor is used instead.

[8] https://spacy.io/api/entitylinker.

[9] For the purpose of the example, we have adapted the outputs of Google Translate and IBM Watson's Natural Language Understanding service. We have enriched the resulting item graph with additional facts from DBpedia and Wikidata and added examples of confidence, strength, and relevance.

Listing 2 The tweet in Figure 2 represented semantically as an item graph using Turtle notation.

```
:twitter834619575509594114 a :Item;
   :source dbr:Universal_Television_(Somalia);
   :sourceIRL "https://twitter.com/ ... /834619575509594114";
   :originalText "WAR DEG DEG: Madaxweyne Farmaajo oo ..."@so;
   :canonicalText "PRESS RELEASE: President Farmajo appointed
                   Hassan Ali Khayre as new prime minister"@en;
   :hasAnnotation [    a :Annotation;
                       :hasEntity dbr:Mohamed_Abdullahi_Farmajo;
                       :hasConfidence 0.9;
                       :hasRelevance 0.95;
                       :hasStrength 1.0                        ],
                  [    a :Annotation;
                       :hasEntity unres:Hassan_Ali_Khayre      ],
                  [    a :RelationAnnotation;
                       :hasRelation wn:02481345-v;
                       :relationFrom dbr:Mohamed_Abdullahi_Farmajo;
                       :relationTo unres:Hassan_Ali_Khayre     ],
                  [    a :RelationAnnotation;
                       :hasRelation dbo:office;
                       :relationFrom unres:Hassan_Ali_Khayre;
                       :relationTo dbr:Prime_minister          ],
                  [    a :RelationAnnotation;
                       :hasRelation foaf:name;
                       :relationFrom unres:Hassan_Ali_Khayre;
                       :relationTo "Hassan Ali Khayre"@en      ].
```

Listing 3 An existing item graph with facts about Hassan Khaire harvested from YouTube's video descriptions dated 2010.

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

event graphs provide more complete and precise information than individual item graphs, each of which may only describe a small part or aspect of an event. For the same reason, event graphs are 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.

Role in the architecture 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. 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 facts taken from the Linked Open Data
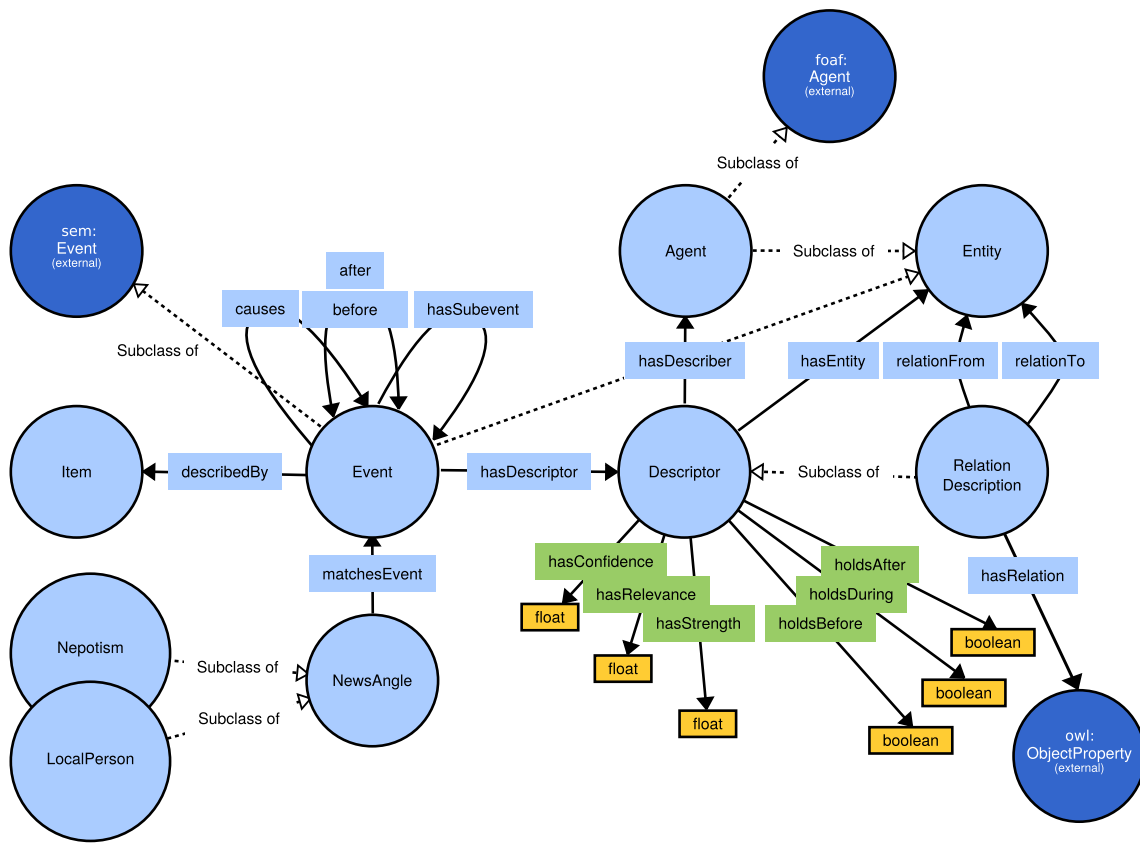
Fig. 5 Ontology for representing news events semantically in a knowledge graph

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

Concepts and relations Figure 5 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 similar to those in Fig. 4. **RelationDescriptors** represent semantic relationships between pairs of entities in the same event graph.

Figure 5 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 Sect. 5.4.

Related ontologies Particularly relevant are again the IPTC's EventsML G2 vocabulary and the BBC Ontologies. Section 6

will review and compare our contribution to these and other existing event-related ontologies.

Reasoning Simple clustering of item graphs by annotation similarity is straightforward. 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 subevents with temporal, causal, and other relations between them.

An earlier prototype clustered items using Scikit-learn's DBSCAN algorithm, which offers scalability and focus on neighbourhood size at the expense of uneven cluster sizes [4]. Other researchers have investigated detection of events in knowledge graphs [24,38], as well as relations between events [43].

Merging entities and relations from item graphs that belong to the same event is also straightforward, as long as standard identifiers (IRIs) are used during lifting. We have so far used DBpedia IRIs where available, and an earlier prototype enriched the knowledge graph with DBpedia facts [4]. But many other sources of standard IRIs and related facts are

Listing 4 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 dbr:Mohamed_Abdullahi_Farmajo   ],
                   [ a :Descriptor;
                     :hasEntity dbr:Hassan_Ali_Khayre          ],
                   [ a :Descriptor;
                     :hasRelation wn:02481345-v;
                     :relationFrom dbr:Mohamed_Abdullahi_Farmajo;
                     :relationTo dbr:Hassan_Ali_Khayre         ],
                   [ a :RelationDescriptor;
                     :hasRelation dbo:primeMinister;
                     :relationFrom dbr:Hassan_Ali_Khayre;
                     :relationTo dbr:Somalia;
                     :onlyAfter xsd:true                        ],
                   [ a :Descriptor;
                     :hasEntity dbr:Norwegian_Refugee_Council   ],
                   [ a :RelationDescriptor;
                     :hasRelation dbo:employer;
                     :relationFrom dbr:Hassan_Ali_Khayre;
                     :relationTo dbr:Norwegian_Refugee_Council  ],
                   [ a :RelationDescriptor;
                     :hasRelation dbo:location;
                     :relationFrom dbr:Norwegian_Refugee_Council;
                     :relationTo dbr:Norway                     ],
                   [ a :RelationDescriptor;
                     :hasRelation foaf:based_near;
                     :relationFrom dbr:Hassan_Ali_Khayre;
                     :relationTo dbr:Norway                     ].
```

available in the LOD cloud, like Wikidata and GeoNames[10], a freely available geographical database of more than 25 million geographical names (toponyms) that refer to over 11 million unique features.

Example In the example from Listing 2, Universal Somali TV might be treated as a trusted source whose news item is considered a new event without further corroboration. But if news items from other sources would report the same information independently, confidence in the new event would increase, perhaps along with completeness and precision. Listing 4 shows an event graph that could result from enriching the facts in Listing 2 with facts from external sources like DBpedia and Wikidata and from the related item graph shown in Listing 3, assuming that the similar-looking IRIs for Hassan Ali Khayre have been resolved.

5.4 News angles

As noted in Sect. 4, some exceptional events are newsworthy in themselves, but most events have to be made newsworthy by reporting them from a news angle. We represent news angles as: *core patterns* that can be matched with events to
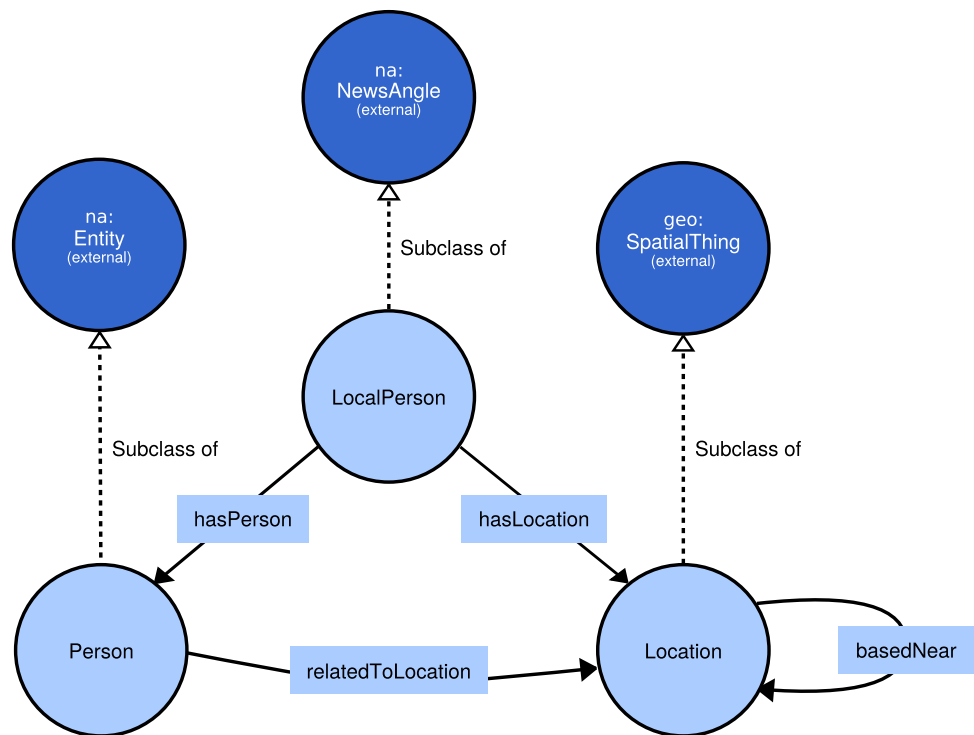
see if the angle fits along with *extended patterns* that suggest additional types of information to include in the presentation of the event. Matching events with news angles is a bidirectional process, in which the core facts of the event suggest candidate news angles and the candidate news angles in turn encourage additional facts to be sought, whether manually or by automated means.

Role in the architecture News angles are important both for detecting newsworthy events and for presenting them in ways that may interest the intended audience. We can represent each news angle as a core pattern to which an event must be matched and one or more extended patterns according to which the event graph can be enriched in potentially interesting ways. The part of an event graph that matches a news angle becomes a *fabula (sub-)graph*. The term *fabula* is adopted from literary theory [14] to denote *the facts that a story contains* in contrast to the *discourse*, which denotes how the those facts are told. Although our representations of news angles and fabulae might support automatic narration as well, our work on News Hunter is currently limited to proposing angled events as fabulae, leaving the writing to the journalist.

Concepts and relations Figure 6 shows how the core pattern of the **LocalPerson** news angle can be represented in OWL. It is a particularly simple angle, matched whenever a central

Fig. 6 Ontological representation of the local-person news angle

Figure showing ontological representation with na:NewsAngle, na:Entity, geo:SpatialThing (external), LocalPerson, hasPerson, hasLocation, Person, Location, relatedToLocation, basedNear, and Subclass of relations.

Person in an event graph is **relatedTo** a particular **Location** that is of importance to the journalist's intended audience, or to another **Location basedNear** that location. Figure 5 already showed how such an **Event** can be **matched** by a **NewsAngle** to form a fabula.

Figure 7 illustrates the core pattern of a more complex news angle, that of **Nepotism** [46], in which a **PowerfulPerson controls** a **Value** which a **GainingPerson** achieves **access** to because of her/his **privateRelation** (typically a family relation) to the **PowerfulPerson**. Because nepotism proper also requires causality, the angle in Fig. 7 represents a weaker *potential nepotism* that mandates further investigation by journalists.

Related ontologies While there are many existing ontologies that capture central concepts and relations for describing events, and a few ontologies exist for annotating textual items too, we are not aware of previous work on representing news angles as ontologies—or indeed in any reasoning-ready or otherwise machine-processable form.

Reasoning Because they may involve identical or taxonomically related concepts and relations, the library of news angles will form a more or less connected news-angle ontology. The central concepts and relations in this slowly evolving ontology suggests which types of resources and relations that need to be represented in event graphs and lifted from news items.

We are exploring different ways of matching news angles with event sub-graphs. For example, Listings 5 and 6 show SPARQL queries that realise the core patterns of the news angles in Figs. 6 and 7. Each query

Listing 5 A SPARQL query that constructs matches to the local person angle represented in Figure 6.

```
CONSTRUCT {
    [   rdf:type na:LocalPerson;
        na:hasPerson ?person;
        na:hasLocation ?location
    ] .
    ?person na:relatedToLocation ?location .
} WHERE {
    ?person rdf:type na:Person .
    ?location rdf:type na:Location .
    ?person na:relatedToLocation / na:basedNear?
        ?location .
}
```

searches the knowledge graph and constructs a core fabula graph for each match of the angle to an event. In Listing 5, `na:relatedToLocation/na:basedNear?` is a *property path* stating that the person must be related to the location of interest or, optionally, to another location near it. It is an example of how OWL ontologies must sometimes be extended with rules or other additional restrictions to fully represent angles.

We envisage a News Hunter architecture in which many collaborating agents specialise in maintaining and leveraging specific concepts and relations in the connected news-angle ontology, continuously looking for changes that could enable or disable particular angles in response to unfolding events.

Fig. 7 Ontological representation of the nepotism news angle
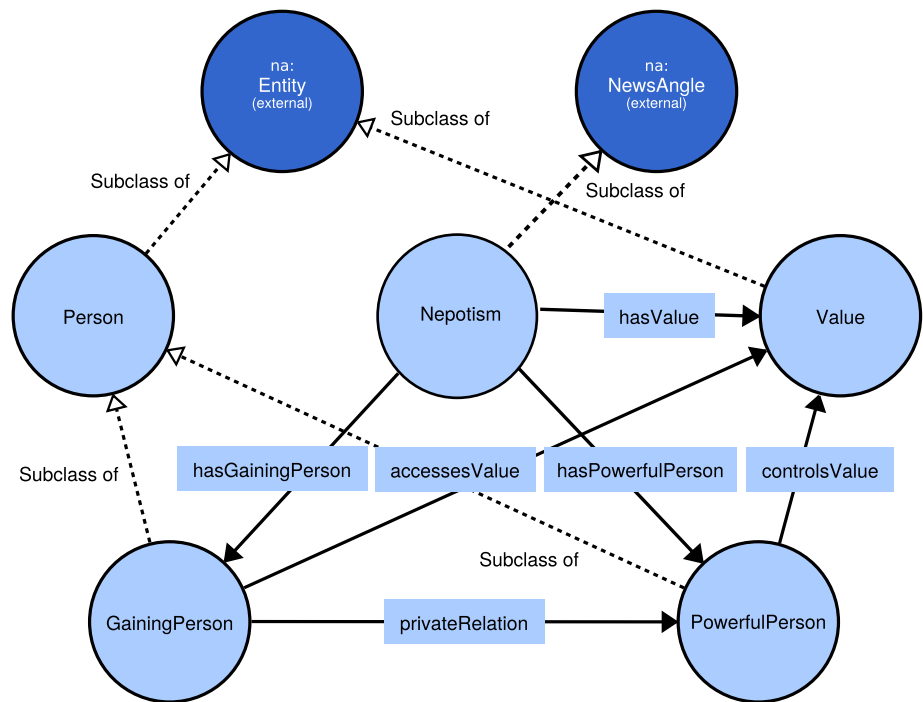
Listing 6 A SPARQL query that constructs matches to the nepotism angle represented in Figure 7.

```
CONSTRUCT {
    [   rdf:type na:Nepotism;
        na:hasPowerfulPerson ?powerfulPerson;
        na:hasGainingPerson ?gainingPerson;
        na:hasValue ?value
    ] .
    ?powerfulPerson na:controlsValue ?value .
    ?gainingPerson na:accessesValue ?value .
    ?powerfulPerson na:privateRelation ?gainingPerson .
} WHERE {
    ?powerfulPerson rdf:type na:PowerfulPerson .
    ?gainingPerson rdf:type na:GainingPerson .
    ?value rdf:type na:Value .
    ?powerfulPerson na:controlsValue ?value .
    ?gainingPerson na:accessesValue ?value .
    ?powerfulPerson na:privateRelation ?gainingPerson .
}
```

Listing 7 Small fabula graph that results from matching facts about prime-minister Khayre with the local-person news angle.

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

For example, a *local-person* agent would specialise in deriving new **Person**–**relatedToLocation**–**Location** facts from the knowledge graph.

Example Listing 7 shows the core *fabula graph* that results from matching the facts in Listing 4 with the news angle

in Figure 6. This graph comprises only four facts, possibly derived by a *local-person agent* from facts stating that Khayre has worked for the Refugee Council located in Norway. Although the graph is simple, the facts it contains are important as they form the core fabula of the angled news report, to which potentially interesting related facts from the LOD cloud can be added. To guide identification of such related facts, the core pattern in Fig. 6 could be augmented into an extended pattern for the local-person angle, also represented as an OWL ontology.

6 Discussion

6.1 Novelty

We have proposed a family of 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, and we suggest for the first time how ontologies for annotating items, events, and news angles can be combined in a journalistic knowledge platform. We also think that the idea of augmenting a journalistic knowledge graph with support for news angles is in itself new.

As such, the *knowledge graph-driven* News Hunter platform can be a useful example of an emerging type of *model-driven information system* that we think is becoming increasingly important. From a systems modelling perspective, development of knowledge graph-driven systems can

involve reuse of existing ontologies and other vocabularies with large user communities outside the enterprise. Commitment to such ontologies makes a wide array of information sources, services, and software readily available for the system under development. At the same time, the fluent nature of the LOD cloud calls for model-driven designs that can leverage new information sources and services quickly and easily as they become available and also replace existing ones as they disappear. Commitment to common ontologies thereby also makes the system under development tied-in to an evolving ecology of sources, services, and software bound together by an ontology that is defined and maintained collaboratively by stakeholders external to the enterprise. Hence, knowledge graph-driven software systems development widens systems modelling to involve new types of long-term strategic concerns about which ontologies and LOD communities the enterprise should align with and to what extent and how it should align. As Sect. 6.7 will mention, developing a system whose components will all read from and update the same (set of) central knowledge graph(s) also calls for well-considered modularisation strategies, which can be ontology-driven too. While none of these concerns are new in themselves, knowledge graph-driven information systems bring them to the fore and combine then in new ways that deserve attention.

6.2 Use cases

We hope the News Hunter platform can help journalists with central tasks such as: detecting newsworthy events quickly and precisely; identifying appropriate angles on those events; and contextualising those angles up with relevant background and other related information. To make these and other uses of the platform clear, we have specified eleven use cases with extensions and variants to drive development of the platform. One particular important example is *What's my angle?* [46], which comprises the following steps:

1. A journalist types a working news report into the front end.
2. News Hunter lifts the working report and returns IRIs for named entities, concepts/topics/categories, relations, and sentiments in the report.
3. News Hunter retrieves angles that fit the working report.
4. News Hunter recommends the most suitable angles.
5. The front end makes recommendations to the journalist.

6.3 Relation to existing annotation-related ontologies

We have systematically compared our proposed annotation ontology with related ontologies from the literature, such as:

- The International Press Telecommunications Council (IPTC) has proposed *NewsML G2* [48][11] as part of its news architecture. Although not based on RDF or OWL, it offers an XML vocabulary and data format for exchanging news-related information in an industrial environment, allowing news items to be annotated with concepts and named entities from controlled vocabularies.
- The BBC's *Linked Open Data Platform* [21] includes the *BBC Things*[12], which is an online reference catalogue of people, places, organisations, and events that matter to the BBC and its audience. It is used to annotate the BBC's archival content.
- Although not specific to news, the *Tag Ontology* [32] focusses on the relations between an agent, an arbitrary resource, and one or more tags. It is extended by the *Meaning-of-a-Tag (MoaT)* ontology [35], which defines relations to the concepts that the tags and resources are about.
- The *SIOC (Semantically Interlinked Online Communities)* ontology [6] makes social media items and posts available as Linked Open Data [5] and lets them be tagged and categorised.

[19,20] present overviews of early annotation ontologies before 2010. More recent proposals include SCOT [19] and MUTO [26]. However, none of them accounts for the **RelationAnnotation** in Fig. 4, which is essential for annotating news items with actual item graphs, because it represents the relation (an OWL-object property) between two **Entities** represented by other **Annotations** of the same **Item**. Other terms we have not found in the related ontologies are:

- **Confidence**, **relevance**, and **strength**, which are essential for representing the uncertain and graded semantic annotations produced by NL lifters.
- The explicit representation of the **annotator Agent** along with each **Annotation**, and the possibility of explicit assignment of **confidence** to **annotator** and **source Agents** as well.

Hence, existing annotation ontologies are not sufficient for News Hunter, although they offer many interesting paths for further linking and extension of our proposal.

6.4 Relation to existing event-related ontologies

We have also systematically compared our proposed event ontology with related ontologies from the literature, which include:

---

[11] https://iptc.org/standards/newsml-g2/.

[12] https://www.bbc.co.uk/things/.

- *IPTC's EventsML G2* [48][13] is an XML vocabulary and data format for "conveying event information in a news industry environment", with focus on receiving, storing, exchanging, and publishing information about persistent (archival) and topical (ongoing) events and their coverage.
- The BBC Ontologies[14] include the *News Storyline Ontology*, which is a generic model for describing and organising the story lines that news organisations tell about events, but which offers no detailed description of the events themselves. The *BBC Core Ontology* defines event subtypes (for music, sports, politics, etc.) that are instantiated by *BBC Things*[15]. There are also specialised ontologies for business news, politics, and sports.
- The *Event and (Implied) Situation Ontology (ESO)* [41] is an OWL2 ontology used in the NewsReader project [38]. It targets economical and financial news and organises different types of events in a taxonomy. It is extended by the *Circumstantial Event Ontology (CEO)* [40] intended to capture chains of newsworthy calamity events.
- Although not specific to news, the *Event Ontology* [37] is an OWL ontology developed to support the Music Ontology[16].
- *ACE (Automatic Content Extraction)*[17] is promoted by the Linguistic Data Consortium to drive research on natural-language processing through standardised annotation tasks and training and evaluation materials. It provides a detailed framework for describing different types of events along with their relations to other events and to the agents and other entities they involve.
- Other common ontologies that deal with events and situations include:
  Linked Open Description of Events (LODE) [42], which is an intentionally minimal model of events aimed to facilitate interoperability;
  DOLCE+DnS UltraLite (DUL) [10], which is a simplification and extension of the DOLCE [11] and the Descriptions and Situations ontologies;
  the *F Model of Events* [39], which builds a modularised event model over DUL to cover participation in, composition of, causality and correlation of, documentation/representation of, and interpretation of events;
  the *Simple Event Model (SEM)* [50]), which provides core classes for describing events in terms of their actors, places, and times;
  the *Rich Event Ontology (REO)* [7], which is an OWL ontology that unifies existing semantic role-labelling

(SRL) schemas (like ESO and CEO) and augments them with causal and temporal relations between events; and the *Comprehensive Event Ontology (CEVO)* [43], which is an event ontology and lexicon designed identify semantic relations between entities that appear in a texts or knowledge graphs.

Linked Open Data resources such as Schema.org, DBpedia, and Wikidata also define terms for describing events and related phenomena. However, none of them accounts for how **Events** are **describedBy Items** and **match Events**, which are our most central concerns in Fig. 5. Other terms we have not found in the related ontologies are:

- **Confidence**, **relevance**, and **strength**, which are derived from the corresponding datatype properties of the **Items** that **describe** the **Event**.
- The **RelationDescriptor**, which represents a semantic relation between a pair of entities that describe the same event, using the **hasRelation** property to indicate the semantic relationship intended. Whereas existing ontologies such as ESO, CEO, and CEVO also represent relations (through their focus on verbs), they do not provide concepts for representing event graphs.

Hence, existing event ontologies are not sufficient for News Hunter, although they offer many interesting paths for future extensions. In particular, they can be used to define the **Entity**-subclasses and **owl:ObjectProperty**-subproperties in event graphs.

6.5 Relation to existing news-angle related ontologies

While there are many existing ontologies that capture central concepts and relations for describing annotations and events, we are not aware of any work on representing news angles as ontologies—or indeed in any reasoning-ready or otherwise machine-processable format. To the best of our knowledge, this is an original contribution of the News Hunter platform, of our ongoing News Angler project, and of this paper.

6.6 Further ontology development

We expect the proposed ontologies to evolve as we develop the News Hunter platform and proof-of-concept prototype further. It is possible that Fig. 6 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.

Additional ontologies may also be needed, for example, to: organise different types of input items; represent avail-

---

[13] https://iptc.org/standards/eventsml-g2/

[14] https://www.bbc.co.uk/ontologies.

[15] https://www.bbc.co.uk/things/.

[16] http://musicontology.com/specification/

[17] https://www.ldc.upenn.edu/collaborations/past-projects/ace.

able 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 as separate ontologies in this paper, we see them as alternative thematic windows into a single logically contiguous, but perhaps physically distributed, knowledge graph.

6.7 Architectural considerations

Our paper has focussed on ontology—and thus on knowledge graph—structure, leaving architectural and most reasoning issues to parallel [1,9,46] and future work. The thematic ontology organisation can potentially also shape News Hunter's processing structure, so that different subsystems, perhaps implemented as collaborating agents, can take responsibility for different thematic sub-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. The different entities used to describe items and events could also be managed by different sub-systems according to types such as person, organisation, and location, each of which requires different forms of enrichment and analysis. Finally, metadata about privacy, provenance, and terms of use could also be maintained by different sub-systems. In such an architecture, the knowledge graph would be split by sub-ontology into contiguous and sometimes overlapping *thematic sub-graphs* that comprise both ontology definitions (TBox), corresponding RDF-facts represented as triples (ABox), and associated software agents with responsibility for maintaining and leveraging facts of a particular type.

6.8 Fake news

Fake news that originates from fraudulent social messaging accounts, deceptive web sites, and bogus news feeds has received increasing attention in recent years. We argue that a journalistic knowledge platform like News Hunter can make false information easier to detect because it is able to constantly triangulate multiple real-time sources of information about the same event. It is also able to corroborate parts of the information with background information taken from reference sources. This opens for triangulation- and fact-based approaches to identifying fake news, which complement the current focus on identifying fake news items by their sources and textual features. The representation of harvested information in a knowledge graph may also open for a graph-based approach to identifying fake news, using graph features instead of or in addition to the textual features used in current machine-learning approaches to fake news identification.

Care must be taken to ensure that malicious actors do not exploit journalistic knowledge platforms like News Hunter to generate fake news by infusing false news items into the knowledge graph. We have therefore designed the ontology to keep traces of all events back to the items they aggregate and further back to the persons or other agents who contributed them, making it possible to retract information derived from news sources or items that later turn out to be wholly or partially false.

7 Conclusion

Our work on the News Hunter platform and prototype has opened many interesting paths for further work: developing the platform 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; understanding what makes a news event or report interesting in a particular context; and selecting the most suitable and appropriate empirical research goals and evaluation approaches for our project.

In a world of ever-increasing information, journalists are not the only ones facing the challenge of detecting interesting events and situations in big data sets and presenting those events and situations in 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.

References

1. Al-Moslmi, T., Ocaña, M.G., Opdahl, A.L., Veres, C.: Named entity extraction for knowledge graphs: a literature overview. IEEE Access **8**, 32862–32881 (2020)

2. Al-Moslmi, T., Ocaña, M.G., Opdahl, A.L., Tessem, B.: Detecting Newsworthy Events in a Journalistic Platform. Paper presented at the European Data and Computational Journalism Conference, Malaga, Spain (June 2019)

3. Allemang, D., Hendler, J.: Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. Elsevier, Amsterdam (2011)

4. Berven, A., Christensen, O.A., Moldeklev, S., Opdahl, A.L., Villanger, K.J.: News Hunter: building and mining knowledge graphs for newsroom systems. NOKOBIT **26**, 1–11 (2018)

5. Bizer, C., Heath, T., Berners-Lee, T. Linked data: the story so far. In: Semantic Services, Interoperability and Web Applications: emerging concepts. pp. 205–227. IGI Global (2011)

6. Breslin, J.G, Harth, A., Bojars, U., Decker, S.: Towards semantically-interlinked online communities. In: European Semantic Web Conference, pp. 500–514. Springer (2005)

7. Brown, S., Bonial, C., Obrst, L., Palmer, M.: The rich event ontology. In: Proceedings of Events and Stories in the News Workshop, pp. 87–97 (2017)

8. Corcoglioniti, F., Rospocher, M., Aprosio, A.P.: Frame-based ontology population with PIKES. IEEE Trans. Knowl. Data Eng. **28**(12), 3261–3275 (2016)

9. Ocaña, M.G., Nyre, L., Opdahl, A.L., Tessem, B., Trattner, C., Veres, C.: Towards a big data platform for news angles. In: Proceedings of 4th Norwegian Big Data Symposium (NOBIDS 2018), vol. 2316, pp. 17–29. CEUR Workshop Proceedings (2018)

10. Gangemi, A.: Dolce+DnS Ultralite (2009). RDF+OWL ontology at http://ontologydesignpatterns.org/ont/dul/DUL_v23.owl. Accessed 26 Sept 2019

11. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: *International Conference on Knowledge Engineering and Knowledge Management*, pp. 166–181. Springer (2002)

12. Gangemi, A., Presutti, V., Recupero, D.R., Nuzzolese, A.G., Draicchio, F., Mongiovì, M.: Semantic web machine reading with FRED. Semant. Web **8**(6), 873–893 (2017)

13. Germann, U., Liepins, R., Gosko, D., Barzdins, G.: SUMMA: Integrating multiple NLP technologies into an open-source platform for multilingual media monitoring. In: Proceedings of Workshop for NLP Open Source Software (NLP-OSS), pp. 47–51 (2018)

14. Gervas, P.: Computational approaches to storytelling and creativity. AI Mag. **30**(3), 49–62 (2009)

15. Gregor, S., Hevner, A.R: Positioning and presenting design science research for maximum impact. MIS Q. **37**(2), 337–355 (2013)

16. Harcup, T., O'Neill, D.: What is news? News values revisited (again). Journal. Stud. **18**(12), 1470–1488 (2017)

17. Hevner, A.R.: A three cycle view of design science research. Scand. J. Inf. Syst. **19**(2), 4 (2007)

18. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)

19. Kim, H.-L., Decker, S., Breslin, J.G.: Representing and sharing folksonomies with semantics. J. Inf. Sci. **36**(1), 57–72 (2010)

20. Kim, H.L., Passant, A., Breslin, J.G., Scerri, S., Decker, S.: Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces. In: 2008 IEEE International Conference on Semantic Computing, pp. 315–322. IEEE (2008)

21. 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: European Semantic Web Conference, pp. 723–737. Springer (2009)

22. Kolitsas, N., Ganea, O.-E., Hofmann, T.: End-to-end neural entity linking (2018). arXiv:1808.07699

23. Latar, N.L.: The robot journalist in the age of social physics: the end of human journalism? In: The New World of Transitioned Media, pp. 65–80. Springer (2015)

24. Leban, G., Fortuna, B., Brank, J., Grobelnik, M.: Event registry: learning about world events from news. In: Proceedings of 23rd International Conference on World Wide Web, pp. 107–110. ACM (2014)

25. 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 25th ACM International Conference on Information and Knowledge Management, pp. 207–216. ACM (2016)

26. Lohmann, S., Díaz, P., Aedo, I.: MUTO: the modular unified tagging ontology. In: Proceedings of 7th International Conference on Semantic Systems, pp. 95–104. ACM (2011)

27. Lohmann, S., Link, V., Marbach, E., Negru, S.: WebVOWL: web-based visualization of ontologies. In: International Conference on Knowledge Engineering and Knowledge Management, pp. 154–158. Springer (2014)

28. Machill, M., Beiler, M.: The importance of the internet for journalistic research: a multi-method study of the research performed by journalists working for daily newspapers, radio, television and online. Journal. Stud. **10**(2), 178–203 (2009)

29. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia spotlight: shedding light on the web of documents. In: Proceedings of 7th International Conference on Semantic Systems, pp. 1–8 (2011)

30. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)

31. Miroshnichenko, A.: AI to bypass creativity. Will robots replace journalists? (The answer is "yes"). Information **9**(7), 1–20 (2018)

32. Newman, R., Ayers, D., Russell, S.: Tag Ontology (2005). http://www.holygoat.co.uk/projects/tags. Accessed 07 Apr 2006 version through https://web.archive.org/web/20060407215007/http://www.holygoat.co.uk/projects/tags/

33. Opdahl, A.L., Berven, A., Alipour, K., Christensen, O.A., Villanger, K.J.: Knowledge graphs for newsroom systems. NOKOBIT **24**, 1–4 (2016)

34. Opdahl, A.L., Tessem, B.: Towards ontological support for journalistic angles. In: Enterprise, Business-Process and Information Systems Modeling, pp. 279–294. Springer (2019)

35. Passant, A., Laublet, P.: Meaning of a tag: a collaborative approach to bridge the gap between tagging and linked data. LDOW **369** (2008)

36. Peroni, S., Gangemi, A., Vitali, F.: Dealing with markup semantics. In: Proceedings of 7th International Conference on Semantic Systems, pp. 111–118. ACM (2011)

37. Raimond, Y., Abdallah, S.: The Event Ontology (2007). http://motools.sourceforge.net/event/event.html. Accessed 23 Sept 2019

38. 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. **37**, 132–151 (2016)

39. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F—a model of events based on the foundational ontology DOLCE+DnS Ultralight. In: Proceedings of 5th International Conference on Knowledge Capture, pp. 137–144. ACM (2009)

40. Segers, R., Caselli, T., Vossen, P.: The Circumstantial Event Ontology (CEO). In: Proceedings of Events and Stories in the News Workshop, pp. 37–41 (2017)

41. Segers, R., Vossen, P., Rospocher, M., Serafini, L., Laparra, E., Rigau, G.: ESO: a frame based ontology for events and implied situations. In: Proceedings of MAPLEX (2015)

42. Shaw, R., Troncy, R., Hardman, L.: LODE: linking open descriptions of events. In: Asian Semantic Web Conference, pp. 153–167. Springer (2009)

43. Shekarpour, S., Alshargi, F., Thirunaravan, K., Shalin, V.L., Sheth, A.: CEVO: comprehensive event ontology enhancing cognitive annotation on relations. In: 2019 IEEE 13th International Conference on Semantic Computing (ICSC), pp. 385–391. IEEE (2019)

44. Shoemaker, P.J., Reese, S.D.: Mediating the message: Theories of influences on mass media content, 2nd edn. Longman, New York (1995)

45. Singhal, A.: Introducing the Knowledge Graph: Things, Not Strings. Official Google Blog, vol. 5 (2012)

46. Tessem, B., Opdahl, A.L.: Supporting journalistic news angles with models and analogies. In: 13th International Conference on Research Challenges in Information Science (RCIS), pp. 1–7. IEEE (2019)

47. Thurman, N.: Computational journalism. In: Karin, W.-J., Thomas, H. (eds.) The Handbook of Journalism Studies, chapter 12, 2nd edn. Routledge, New York (2019)

48. Troncy, R.: Bringing the IPTC news architecture into the semantic web. In: International Semantic Web Conference, pp. 483–498. Springer (2008)

49. Vaishnavi, V., Kuechler, B., Petter, S.: Design Research in Information Systems (2004). http://desrist.org/design-research-in-information-systems. Accessed 28 Oct 2019

50. Hage, W.R.V., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the Simple Event Model (SEM). Web Semant. Sci. Serv. Agents World Wide Web **9**(2), 128–136 (2011)

51. Vossen, P., Agerri, R., Aldabe, I., Cybulska, A., van Erp, M., Fokkens, A., Laparra, E., Minard, A.-L., Aprosio, A.P., Rigau, G.: NewsReader: using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. Knowl. Based Syst. **110**, 60–85 (2016)

Bjørnar Tessem is Professor in Information Science at the Department of Information Science and Media Studies, University of Bergen, Norway. His research and teaching cover a large span within information science, including theoretical and applied research in artificial intelligence and machine learning, design of new media with an emphasis on location-based media and journalism, and empirical studies of information systems development and agile software development. He has published in high-ranked journals and conferences in the whole breadth of his research.

Andreas L. Opdahl is Professor of Information Systems Development at the Department of Information Science and Media Studies, University of Bergen, Norway. He received his Ph.D. from the Norwegian University of Science and Technology. His research interests include ontologies and knowledge graphs, enterprise and IS modelling, as well as safety and security requirements. Opdahl is the author, co-author or co-editor of more than a hundred peer-reviewed research papers that have been cited more than 3800 times. He is a member of IFIP WG5.8 on Enterprise Interoperability and WG8.1 on Design and Evaluation of Information Systems. He serves regularly as a reviewer for premier international journals and on the program committees and as an organizer of renowned international conferences and workshops.