The `newsEvents` Ontology⋆
An Ontology for Describing Business Events

Uta Lösch and Nadejda Nikitina

AIFB, Universität Karlsruhe, Germany
{uhe,nani}@aifb.uni-karlsruhe.de

Abstract. In the broader context of the development of an ontology-based new event detection system, we are developing the `newsEvents` ontology which allows modeling business events, the affected entities and relations between them. This paper presents requirements for and a first version of this ontology. A pattern-based approach to the design of the ontology was taken. Thereby a new useful pattern - the `EventRole` pattern - was identified and specified.

1   Introduction

The analysis of news and the estimation of their market impact is an important issue for traders in financial markets. As the amount of news that is made available is huge, at least a partial automation of the analysis process is desirable. The goal is to filter relevant news, to provide an aggreggated view on their content and to thus enable the users to only read those news that are really relevant for them. The relevance of a piece of news depends on their relevance for the traders' interests (i.e. is it about an entity that the trader is interested in) and its novelty. These two aspects are rather independent of each other and can thus be approached separately.

The idea of our work is to study which benefits ontologies may offer for the two described aspects. In order to study this problem, an ontology is needed which allows to describe relevant entities and events in the business news context. To the best of our knowledge no such ontology exists. We therefore decided to build one for the use in the kind of system we have described above. In this paper, the `newsEvents` ontology, which was developed for that purpose, will be presented.

In the task of relating news to a user's interest ontology-based annotations may provide a condensed representation of a text's content. A query on these annotations can be used to describe the user's information need. A similar system has been described by [4].

For the task of discovering new events, to which we refer as the new event detection problem, the annotation of texts with ontology entities allows for assessing their content and relating news to each other based on these annotations. The introduction of additional features that are based on annotations in the clustering task seems promising as it allows a more specific analysis of the content and as it enables a better distinction between similar events.

More precisely, annotations will be obtained from a state-of-the-art tool (OpenCalais[1] has been chosen for this task). To include the annotations in the clustering task a similarity measure that takes into account the similarity between annotations will be introduced. However, the provided annotations do not offer much structure, only a list of annotation types and some properties for each type are defined. A taxonomy on the annotation types is not defined. The `newsEvents` ontology will allow for the description of the current state of the domain and of events as reported in news. The ontology provides a formalisation of this kind of information that includes more background knowledge than OpenCalais' annotations.

The rest of the paper is structured as follows: Section 2 describes related work, section 3 describes our requirements for the ontology, section 4 describes the ontology we built, section 5 describes the newly defined EventRole pattern, before we conclude in section 6.

2 Related Work

While to the best of our knowledge no ontology exists which aims at describing companies, their relations among each other and the most important events, there is a number of models which address part of the scope of the intended ontology.

OpenCalais has defined a schema which is used for the annotation of news texts, in which events and entities are annotated. While the scope is quite similar to the scope of the `newsEvents` ontology, it only defines a list of annotation types and, for the complex annotations, the slots that have to be filled. The goal when defining the `newsEvents` ontology was to provide a model of the domain that provides more background knowledge on event and entity types and that is able to relate different annotations to each other.

The classification of news according to their content is very important for news providers. Therefore, various annotation languages have been defined for this purpose: while IIM[2] is today only used for annotating photos, its successors NITF[3] and NewsML[4] are still in use. However, their primary concern is to have a standardized format for providing and exchanging news.

---

[1] http://www.opencalais.com
[2] The Information Interchange Model, http://www.iptc.org/IIM/
[3] The News Industry Text Format, http://www.nitf.org
[4] The News Meta Language, http://www.newsml.org

The most important entity types (and also some events) are also defined in top-level ontologies such as Cyc[5] or Proton[6]. However, these ontologies are not detailed enough for our use case.

There is a number of ontologies in the finance domain. The LSDIS_Finance ontology[7] and the dip Ontology [1] both describe actors and products in the stock markets. However, these ontologies do not include any description of events.

3   Competency questions

As described above, the `newsEvents`[8] ontology shall be used for describing events which are relevant in a business context and for assessing similarity between different events. It will also be used for modeling the current information that is available about the economy and for determining whether a user might be interested in a new event that is reported. In the future, it may also be considered whether there are recurring patterns of series of events where the next events may be anticipated. It may also be useful for studying the impact of events on financial markets.

As the latter use cases may only be interesting in the future and are not the main motivation for developing the ontology, no special attention has been paid to them when developing the first version of the ontology. It will however be extended accordingly in the future.

When developing the ontology, the goal was to have as little modeling efforts as possible. Therefore, a pattern-based approach to its design was taken.

The first step in the development of the ontology was the definition of a set of competency questions. Competency questions are questions that the ontology should be able to answer. Ideally, the ontology should be able to answer all and only the competency questions - no superfluous additional information should be included in the model. For a more detailed discussion of competency questions see [3]. The following questions were defined:

- Related to the history of an event
  - *Is there any information on a specific event already available?* This question serves to determine whether a specific event has already been reported on. If this is not the case, the event is definitely new - information about it has to be integrated into the knowledge base. Additionally, its novelty will be high.
  - *In which order and in which timeframe was information on a specific event published?* The purpose of this question is to determine the development of the information that is made available on an event. This will help to study the development of the history of a single event. It may

---

[5] The Cyc knowledge base, http://www.cyc.com

[6] The proton ontology, http://proton.semanticweb.org

[7] http://lsdis.cs.uga.edu/projects/meteor-s/wsdl-s/ontologies/LSDIS_FInance.owl

[8] The ontology is available at
http://www.aifb.uni-karlsruhe.de/WBS/uhe/ontologies/newsEvents.owl

however also help to determine patterns of how event histories look like and thus to anticipate new information.

– Related to the assessment of similarity
  - *How similar are two entities?* The similarity of two entities may be defined through their properties, like entity type, name, position, location, industry, etc. The assessment of entity similarity is needed for the assessment of event similarity, but it will also help to identify entities on which similar events may have similar impact, entities which may have a similar history, or entities that may be interesting for a user based on the interests he has stated explicitely.
  - *How similar are two events?* The similarity of two events is needed for deciding whether two events are in fact the same. Furthermore, if two events are very similar, they may be interesting for the same users and the history of one event may allow for anticipating future developments concerning the other event.

– Related to relations between entities
– *Which products does a company produce? Which industry does a company belong to? Where is a company located?* Although these questions may seem very heterogenuous at first sight, they all serve for finding entities and events that are related to a user's interests. For example, if he is interested in cars, he may be interested in all news that relate to car manufacturing companies.

4   The `newsEvents` ontology

Based on the requirements presented in the previous section, we have developed a first version of the `newsEvents` ontology. It describes various entity types that are relevant in business news as well as important events, in which described entities may be involved.

For the development of the ontology we tried to follow a pattern-based approach. Especially, we found the use of content design patterns helpful. These are small ontologies - typically consisting of two to ten classes and relations among them, which describe typical modeling problems arising in different domains. These patterns were proposed by Gangemi and Presutti [2] [6]. The goal is facilitate the design of the ontology by providing building blocks which can be composed, specialized or instantiated and thus be adapted to a specific domain.

As OpenCalais is used for obtaining annotations, the annotation types Open-Calais defines were taken as a starting point for the definition of the concepts that should be represented in the ontology.

After the most important concept types were chosen, the next step consisted in identifying the patterns that could be reused for defining the `newsEvents` ontology. Using the patterns described below, the description of most of the event and entity types turned out to be rather straight forward. As the whole ontology is quite big and complex, we only show parts of it to illustrate the use of design patterns:

- *TimeIndexedParticipation.* This pattern describes the involvement of objects in an event at a certain point in time. This pattern can be reused for the description of events like `Acquisition`, `Merger`, `IPO`, etc. The various event types could be described as specialisations of the concept `Event`, its participants could be defined through the use of specialisations of the `hasParticipant` property and of the `Participant` class, which are defined in the *Participation* pattern. The association of an event with the time at which it is happening is possible through the `TimeIndexedParticipation` concept, which relates events and their participants to a time component.
- *Situation.* This very general pattern can be used for the description of complex relations, like `CompanyTicker`. The latter describes a company which is traded at a specific stock exchange and has a specific ticker symbol there. This could be modeled by defining the concept `CompanyTicker` as subconcept of `Situation`.
- *Place.* This pattern defines how places and locations should be described in an ontology. The `newsEvents` ontology describes various concepts, especially event types, that happen at a certain location. The place pattern allows to model these locations and relations among them.
- *ObjectRole.* This pattern allows to describe the different roles an entity of a specific type may play. This pattern is useful for defining the different roles an entity may play in an event or in a relation among entities. For example, an `Acquisition` describes the event of one company acquiring another one. Obviously, there are two entities of the same type, i.e. two companies involved. One company is the acquiring company, the other one the acquired company - the companies thus play different roles in the event. To address the problem of roles in events we have defined the `EventRole` pattern, which will be described in section 5.

A taxonomy of event and entity types has been defined, which is needed to allow for the calculation of similarities between different event types.

To distinguish between the different type of events, we introduced additional classes for describing events that have a similar meaning. For this purpose classes like `CompanyCollaboration`, `LegalIssue`, or `StockEvent` were introduced. None of these classes defines concrete events. It is merely used as a grouping element. For example, the class `CompanyCollaboration` has the subclasses `Alliance`, `BusinessRelation`, `JointVenture`, and `Merger`. Each of these events describes a way in which two companies may choose to collaborate.

Entities were also grouped in a hierarchy. Here, grouping was chosen according to which roles an entity may play in events. Most importantly, we created a class `LegalEntity` which describes both legal and natural persons, i.e. every entity that may be an actor in an event.

5   The `EventRole` pattern

When building the ontology, design patterns could be used to address most of the modeling problems that were encountered. However, no obvious solution seemed

to be available for the case where two entities of the same type are involved in the same event. This is however a recurring situation. Examples in our use case are acquired and acquiring company, provider and customer, plaintiff and suedEntity, etc. The occurrence of this kind of situation is not restricted to the business domain: an obvious example in an everyday context is a visit, where there is a visitor and a person that is visited.

The problem can be solved by composing the patterns *Participation* and *ObjectRole*. The idea is that the participants in an event are not described by their entity type, but by the role they are taking in the event. The resulting pattern is depicted in figure 1.

A new class `EventRole` has been introduced which serves as connector between the two original patterns. The `EventRole` class is used to describe the role that an entity plays in an event. Therefore, it is a specialization of the `Role` and the `Object` class. For each role an entity can play in an event, this class should be specialized. Additionally, the property `hasParticipant` should be specialized for each of the entity's roles and the thus defined properties should be declared disjoint. Thus, each object can only have one role in a given event.

To show how the pattern can be adapted to a specific use case, consider the definition of an acquisition event. An acquisition is the event of one company, the acquiring company, buying another one, the acquired company. Two companies are involved in this event, but it makes a huge difference for company A if it is acquired by company B or whether it acquires company B. Therefore, the event roles `AcquiringCompany` and `AcquiredCompany` have been defined. They both are roles of `Company`. The class `Acquisition` is then defined as a subclass of `Event`. Additionally, we defined the restriction that each acquisition has at least one acquiring company and at least one acquired company, but at most 2 participants. The properties `hasAcquiringCompany` and `hasAcquiredCompany` were specified as subproperties of the `hasParticipant` property. The two properties are defined as being disjoint.

6 Conclusion

The paper presented an ontology which can be used to describe events and entities in a business news context. Patterns have proven to be very useful for the design of the ontology as they directly solve many of the modeling issues that were encountered in the engineering process.

One of the recurring problems encountered while modeling the ontology was the description of the roles entities take in an event. The `EventRole` pattern has been proposed in order to solve this issue.

In the future, the proposed ontology will be refined and extended such that relations between events (especially causal relationships) may be included in the ontology.

Additionally, procedural knowledge (following our proposal in [5]) will be associated with the ontology, such that automatic updates of an entity's state after being affected by an event become possible. The ontology, which can then
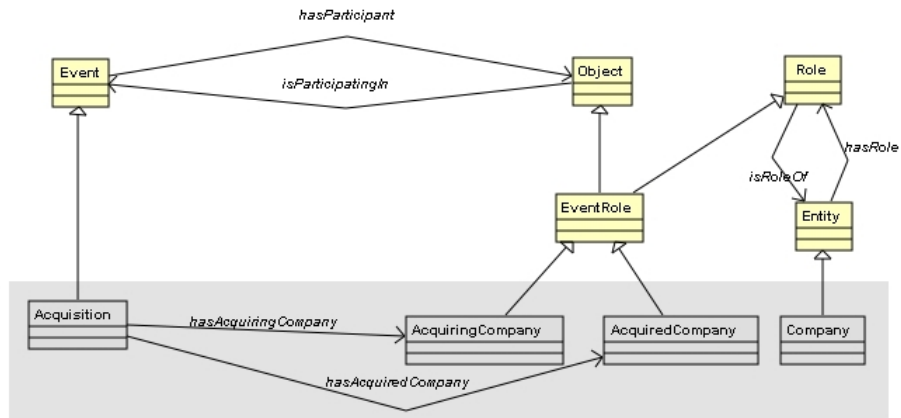
Fig. 1. The role event pattern and its adaptation for modeling acquisitions

automatically adapt to changes in the domain, will then be ready to be used in the new event detection process.

References

1. S. L. Alonso, J. L. Bas, S. Bellido, J. Contreras, R. Benjamins, and J. M. Gomez. Deliverable 10.7 - financial ontology. Technical report.
2. A. Gangemi. Ontology design patterns for semantic web content. *The Semantic Web ISWC 2005*, pages 262–276, 2005.
3. M. Gruninger and M. Fox. The role of competency questions in enterprise engineering, 1994.
4. C. Halaschek-Wiener and J. Hendler. Toward expressive syndication on the web. In *Proc. of the 16 th International World Wide Web Conference(WWW 2007)*, 2007.
5. U. Lösch, S. Rudolph, D. Vrandecic, and R. Studer. Tempus fugit - towards an ontology update language. In *6th European Semantic Web Conference (ESWC 09)*. Springer-Verlag, 2009.
6. V. Presutti and A. Gangemi. Content ontology design patterns as practical building blocks for web ontologies. In *ER '08: Proceedings of the 27th International Conference on Conceptual Modeling*, pages 128–141, Berlin, Heidelberg, 2008. Springer-Verlag.