



Contents lists available at ScienceDirect

# Expert Systems with Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

## The NEWS ontology: Design and applications

Norberto Fernández, Damaris Fuentes, Luis Sánchez\*, Jesús A. Fisteus

Department of Telematics Engineering, University Carlos III of Madrid, Universidad 30, E-28911 Leganés, Madrid, Spain

### ARTICLE INFO

#### Keywords:

NEWS ontology  
News agencies  
Semantic Web  
NEWSPlugin  
Knowledge management in journalism

### ABSTRACT

News agencies are one of the main actors in the news business. These agencies produce daily big amounts of digital contents, which need to be managed and distributed to their customers. The NEWS (News Engine Web Services) project, already completed, had as goal to develop tools that, using state of the art technologies, may help news agencies to improve their production and distribution processes. As a central piece in the NEWS system, an ontology for the news domain was implemented. In this paper, we describe the NEWS ontology, its intended uses in the context of the production and distribution processes of a news agency, and the NEWSPlugin, a Protégé extension developed to partially automate its maintenance.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

Nowadays, news agencies are one of the main actors in the news business. Agencies like Reuters<sup>1</sup> or the Spanish EFE<sup>2</sup> produce thousands of electronic documents every day, which need to be managed and distributed to their customers.

The next subsections describe several factors that make the process of news item management and distribution complex.

#### 1.1. Volume of content

Agencies produce enormous amounts of contents every day. For instance, in the case of EFE, the daily production includes more than 3000 text/markup files and more than 1000 photos. These documents need to be stored in the agency's repositories, where they are available to be queried, updated, etc. Scalability is thus a crucial issue for agencies: their repositories need to be prepared to handle millions of documents. For example, the database EFE uses to store images (called *Fototeca*) stores currently almost one and a half million pictures.

#### 1.2. Content heterogeneity

Agencies produce contents in different languages and formats. In the case of EFE, the production includes images, videos, audio files, text and markup-based files in languages such as Spanish, English, Catalanian or Portuguese.

#### 1.3. Customer heterogeneity

The typical customers of news agencies are heterogeneous in nature: TV and radio channels, newspapers, companies' press cabinets, individuals, etc. These different customers have different information needs. For example, the press cabinet of a company is usually interested in news items talking about the company or its rivals, whereas a sports TV channel is interested mostly in news items describing sport events. Agencies should provide their clients with the information that is relevant to them, avoiding the well-known information overload problem.

#### 1.4. Up-to-date information

Customers are not only interested in obtaining top-quality news items of relevance to their interests, but also demand agencies to provide the information to them as soon as possible because, in many cases, a delay can imply economic losses.

The NEWS (News Engine Web Services) project,<sup>3</sup> currently completed, was a research and development project funded by the European Union whose main goal was to provide tools which can help the news agencies to address all the previously described difficulties. In order to do so, the project partners applied state of the art technologies, like Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) technologies and Web services, to the news agencies' production environment. Specifically, we worked in the context of the project with two European news agencies, the above mentioned EFE and the Italian ANSA.<sup>4</sup>

The NEWS system, described in Fernández et al. (2006), consists of a set of distributed components which interact using Web

\* Corresponding author.

E-mail addresses: [berto@it.uc3m.es](mailto:berto@it.uc3m.es) (N. Fernández), [dfuentes@inv.it.uc3m.es](mailto:dfuentes@inv.it.uc3m.es) (D. Fuentes), [luis@it.uc3m.es](mailto:luis@it.uc3m.es) (L. Sánchez), [jaf@it.uc3m.es](mailto:jaf@it.uc3m.es) (J.A. Fisteus).

<sup>1</sup> <http://www.reuters.com>.

<sup>2</sup> <http://www.efe.es>.

<sup>3</sup> <http://www.news-project.com>.

<sup>4</sup> <http://www.ansa.it>.

services as communication mechanism. The components provide facilities like automatic extraction of metadata from news items' contents, and storage and retrieval of news items and their metadata.

As a central piece in the NEWS architecture, an ontology was implemented. This ontology acts as a shared model among all the system components, providing, among other things, a vocabulary to formalize the metadata used for annotating every news item. It also allows the usage of basic reasoning techniques to implement intelligent content search and distribution services.

In this paper, we describe the current version of the NEWS ontology, its intended uses in the context of the production and distribution processes of a news agency, and the tool implemented to ease its management. Both the ontology and the tool are publicly available for download at [Web Technologies Laboratory \(2008\)](#).

The rest of this paper is structured as follows: Section 2 briefly revisits the NEWS project main achievements. Section 3 describes the NEWS ontology in depth. Section 4 shows the evolution of the NEWS ontology, describing the most important changes and enhancements in its design through the project life-cycle. Maintaining an ontology in an evolving domain as the one covered by the NEWS project is a difficult task. In order to partially automate this task, a tool in the form of an extension for the Protégé ontology editor has been developed. This tool is described in Section 5. Section 6 describes practical applications of the NEWS ontology in the NEWS system. Finally, Sections 7, with related work, and Section 8, which presents the main conclusions achieved regarding the NEWS ontology development, close this paper.

## 2. The NEWS project

NEWS was a research and development project funded by the European Commission under contract FP6 001906 in the framework of the Information Society Technologies (IST) programme. The project consortium was composed of two news agencies, EFE and ANSA, the DFKI research institute, Ontology Ltd. company, and the University Carlos III of Madrid.

### 2.1. NEWS agencies current workflow

As explained in the introductory section, the main goal of the project was the development of semantic based tools and systems to be deployed in news agencies in order to improve their productivity and revenues. In order to do so, the project partners analyzed the situation in the agencies, trying to find the weaknesses and points of improvement. This analysis phase covered both the production and distribution processes. The situation at the agencies when the NEWS project started, was as follows:

#### 2.1.1. Production

The journalists of a news agency use its proprietary tools to edit the contents of news items. These tools allow not only to create or modify the contents of news items, but also to manually annotate those contents by filling in some forms. This manual annotation process included the addition of information like the author, language, creation timestamp and location or keywords. Multimedia news items were also annotated with a textual description of their contents. Another metadatum which was also added to every news item was the category or categories of such news item in a certain categorization system. These categories reflect a topic which is covered by a certain news item, such as *sports*, *economy*, etc. In both EFE and ANSA, the categorization system was proprietary, and, thus, the different news agencies used different terms to describe their items.

Apart from allowing the edition of news items, the tools provided by the agencies offer common features that a journalist may need when creating a new document: money converters, dictionaries, access to the agencies' knowledge bases (which contain biographies, history and geopolitical data of countries, etc.) and also access to the agencies' data warehouse where the news items produced in the past by the agency are available for documentation purposes.

Once a news item content and annotations are provided by the journalist, a new news item is generated. News items are represented as documents in a certain format. The formats currently in use include binary ones like Information Interchange Model (IIM)<sup>5</sup> and XML-based ones, like News Industry Text Format (NITF)<sup>6</sup> and NewsML,<sup>7</sup> both standardized by the International Press Telecommunications Council (IPTC).<sup>8</sup> The generated news item is then stored into the data warehouse infrastructure of the news agency and distributed to customers.

#### 2.1.2. Distribution

The analyzed agencies use two different methods to distribute the contents that they produce: *pull* distribution services and *push* distribution services.

In the *pull* services, the agencies offer interfaces to their customers to access their repositories and query them, looking for news items. Commercial information retrieval tools like Autonomy<sup>9</sup> or Personal Library Software (PLS)<sup>10</sup> constitute the basis of these services.

In the *push* services, the agencies define streams of news items so that all the news items produced by the agency that match certain criteria are sent through these streams. Customers simply select the streams of interest and subscribe to them. In order to decide if a certain news item needs to be included in a stream, the metadata added during the production process (category, language, etc.) are used.

### 2.2. NEWS improvement proposals

After analyzing the depicted scenario, a number of proposals for improvement were suggested by the NEWS consortium. Basically, they covered three main aspects:

#### 2.2.1. Categorization

The process of tagging news items with categories was performed by hand, a process both prone to error and time consuming. Furthermore, this imposes a limitation in the complexity of the categorization system: as the journalists had to categorize by themselves the news items, using a categorization system with too many categories makes difficult to find the right one to be used. The agencies were aware of this situation, and due to this, the systems in use were small.<sup>11</sup>

Another limitation comes from the fact that the categorization systems were proprietary, which reduces interoperability between agencies and has the consequence that a customer of several news agencies must deal with the different systems.

In the NEWS system, this task is improved by using a human-supervised automatic categorization tool, which also allows the usage of more complex categorization systems like the one

<sup>5</sup> <http://www.iptc.org/IIM/>.

<sup>6</sup> <http://www.nitf.org>.

<sup>7</sup> <http://www.newsml.org>.

<sup>8</sup> The IPTC (<http://www.iptc.org>) develops and maintains technical standards for improved news exchange that are used by virtually every major news organization in the world.

<sup>9</sup> <http://www.autonomy.com>.

<sup>10</sup> A company disappeared in 1998 when it was acquired by AOL.

<sup>11</sup> For instance, in the case of EFE it included less than 20 categories.

provided by the IPTC: the Subject Code NewsCodes,<sup>12</sup> which at the time of writing consists of more than 1300 categories. By using this standard, agencies also improve their interoperability.

### 2.2.2. Content annotation

As explained before, when the project started, both EFE and ANSA annotated their news items with metadata. But these metadata were mostly used for internal management of news items at the agency, and included information like the author, timestamp and language of contents. The content itself was not annotated. Hence, for example, the basic entities (organizations, persons, places, etc.) or events mentioned inside the news items were not tagged.

In order to improve that situation, the project consortium developed components and algorithms (Fernández, Blázquez, Sánchez, & Bernardi, 2007) that, using text analysis and statistical techniques, could parse the text of a news item and automatically detect the entities mentioned there, linking them to instances in the NEWS ontology. Additional mechanisms were also developed to detect and annotate the main events described in a news item. As indicated in Fernández et al. (2006), these annotation techniques are used both with textual news items and multimedia news items, because the agencies' journalists have already added a textual description to all the multimedia contents, in order to allow the usage of text information retrieval algorithms with binary contents. Though the operation of the annotation components is mainly automatic, a human-supervision step was introduced in the NEWS system workflow, in order to ensure the quality of the results by providing training information useful to improve the system performance.

The content annotations obtained were used within the NEWS system to implement fine-grained news item selection, allowing the development of advanced distribution services like, for instance, personalized *push* services providing news items talking about a certain person or event, *more-like-this pull* suggestion services, etc.

### 2.2.3. Multilingualism

One of the main drawbacks of the information retrieval tools used by the agencies is that they work with text, which is language-dependent. As both EFE and ANSA produce contents in several languages, the situation could be improved by providing multilingual information retrieval tools. In the context of the project, and taking into account the needs of the agencies involved, three are the languages supported: English, Spanish and Italian.

In order to implement the multilingual search capabilities, the NEWS system exploits the semantic annotations in the content, because such annotations are formalized using ontology identifiers, which are language-independent.

We will not describe here each of the components of the NEWS system in detail; the interested reader can find a comprehensive description of the project achievements in Fernández et al. (2006). In the rest of the paper we will focus on describing all the work done in the project regarding the NEWS ontology design, implementation and maintenance.

## 3. The NEWS ontology

### 3.1. Design requirements

As a first step in the process of ontology development, we analyzed the intended usage scenario of the NEWS system, briefly described in the previous section, and how it might affect the design

of the ontology. Some of the restrictions introduced by the scenario are:

#### 3.1.1. Knowledge representation requirements

The NEWS ontology should provide the basic vocabulary for news item categorization and content annotation, two of the lines of improvement suggested by the NEWS consortium. We are also interested in using the ontology for implementing intelligent information distribution services. Since complex reasoning services are time-consuming, they are not feasible for the huge amounts of news items produced daily by news agencies and their strict time requirements. Therefore, we restricted reasoning services over the NEWS ontology to query expansion both in the categorization module and in the content module. The tests made by the news agencies shown that these limited but efficient reasoning services were effective to retrieve news items belonging to certain category or dealing with some instance of interest (persons, locations, organizations).

In addition to the vocabulary for content and categorization annotations, we formalized the current metadata added by the news agencies with management purposes. In this way, all the metadata that should be attached to a news item may be defined using the NEWS ontology.

Even though, as we have said, reasoning in the NEWS system was restricted to query expansion, we tried to capture as much knowledge as possible about the metadata modelled in the NEWS ontology. Therefore, we developed an OWL-DL ontology and added axioms (transitivity, inverse, symmetry, etc.) to the ontology where appropriate. This provides a better model of the news domain and can be used in a future extension of the NEWS system.

#### 3.1.2. Standard compatibility

There are several standards available in the state of the art which should be taken into account when modeling the domain with the NEWS ontology. Among them, we emphasize the following:

- News representation standards like NewsML or NITF, that, as we have seen in Section 2, are widely used by the agencies in the project.
- NewsCodes, a set of controlled vocabularies (at the time of writing, there are almost 30) defined by the IPTC which provide values for certain elements and attributes within NewsML or NITF documents.
- Standards like the Dublin Core (DC)<sup>13</sup> and Publishing Requirements for Industry Standard Metadata (PRISM),<sup>14</sup> which define metadata vocabularies, useful in describing news items.
- Semantic Web standards, including metadata and ontology representation languages like Resource Description Framework (RDF)<sup>15</sup> and the Web Ontology Language (OWL),<sup>16</sup> both defined by the W3C.<sup>17</sup>

Compatibility with standards is highly desirable in order to make the system interoperable with the agencies' legacy systems, but also with the information systems of the potential customers.

#### 3.1.3. Multilingualism

As any other component, the ontology should be designed taking into account that the NEWS system should support three languages: English, Spanish and Italian. In order to meet this goal,

<sup>13</sup> <http://dublincore.org>.

<sup>14</sup> <http://www.prismstandard.org>.

<sup>15</sup> <http://www.w3.org/RDF/>.

<sup>16</sup> <http://www.w3.org/2004/OWL/>.

<sup>17</sup> <http://www.w3.org/>.

<sup>12</sup> <http://www.iptc.org/NewsCodes>.

**Table 1**  
Multilingual properties.

Name	URL	Description
altLabel	<a href="http://www.w3.org/2004/02/skos/core#altLabel">http://www.w3.org/2004/02/skos/core#altLabel</a>	An alternative lexical label for a resource
prefLabel	<a href="http://www.w3.org/2004/02/skos/core#prefLabel">http://www.w3.org/2004/02/skos/core#prefLabel</a>	The preferred lexical label for a resource, in a given language
definition	<a href="http://www.w3.org/2004/02/skos/core#definition">http://www.w3.org/2004/02/skos/core#definition</a>	A statement or formal explanation of the meaning of a concept

and following the requirement of standard compatibility stated above, we used the W3C SKOS<sup>18</sup> (Simple Knowledge Organization System) vocabulary to define multilingual labels and comments for the ontology elements (classes, properties, instances). In concrete, we used the following SKOS properties: *prefLabel*, *altLabel* and *definition*, as shown in Table 1.

In summary, the above discussion leads to the following conclusions that conditioned the ontology design:

- (i) We decided to structure the ontology into three main modules:
  - Categorization module, containing the taxonomy for news item classification.
  - Management module, containing definitions of elements useful in the semantic annotation of news items with management metadata.
  - Content annotation metadata module, containing definitions of elements useful in the semantic annotation of news item contents.
- (ii) Labels and comments in the three languages of interest in the project were added to the elements (classes, properties, instances) in the ontology using the SKOS vocabulary.
- (iii) We implemented the ontology using the OWL language. More precisely, in order to limit the inference complexity and achieve good computational properties, we use the OWL-DL flavor. RDF is also used to represent metadata about news items.
- (iv) When implementing the different ontology modules, we took into account the related standards. For instance, the categorization module is based on IPTC standards, and the management module includes elements taken from standards like the NewsCodes or PRISM.

The next sections describe in more detail the three ontology modules and the standards and sources which inspire their design.

### 3.2. Categorization module

One of the objectives of NEWS is to provide means for automatic news categorization. The categorization module of the NEWS ontology provides the basic vocabulary, the classes, used in such process. This module is based on the categorization system of IPTC, the Subject Code NewsCodes.

The Subject Code NewsCodes defines a set of values used to describe the subject (s) covered by the contents of a news item. These values are known as *Subject References* by the IPTC. A single news object can provide information about different issues so it can have several Subject References. Each *Subject Reference* is identified by an eight-digit decimal string. The first two digits represent the *Subject*, which can take values in a range of 17 possibilities such as, for example, *Politics* (11), *Labour* (09), *Religion and Belief* (12), or *Science and Technology* (13). The next three digits represent the *Subject Matter*, which is optional (000 means none). The last three digits can contain 000 (no value) or a number representing a *Subject Detail*. Basically, the *Subject*, *Subject Matter* and *Subject Detail* act as a

kind of three level taxonomy (from more general to more specific). For instance, inside the *Sports* category (15000000), the *American football* subcategory (15003000) can be found, and within this one, the subcategory *National Football League (NFL)* (15003001) is included.

The NEWS Categorization module, takes the Subject Code NewsCodes values and defines a tree of classes whose root is the *NewsItem* class defined in the content module (see Section 3.4). The result is a taxonomy of over 1300 classes.

The concrete news items generated by news agencies will be instances of one or more subclasses of *NewsItem* so, for example, a news item related to economy and politics, will be annotated with the following RDF triples:

```
# IPTC Subject economy, business and finance
news_item_uri rdf:type uri_for_cat_04000000.
# IPTC Subject politics
news_item_uri rdf:type uri_for_cat_11000000.
```

Fig. 1 shows the taxonomy of the categorization module. The structure of this taxonomy is exploited in the NEWS system to carry out query expansion. For instance, when a customer requests news items about sport events (category 15000000), all the news items classified as belonging to category *Soccer* (15054000) are also suggested as relevant to the customer.

### 3.3. Management module

The NEWS ontology must provide a vocabulary for news item semantic annotations. As we have seen, this includes both metadata describing the news contents and metadata intended for use in news item life-cycle management by news agencies (priority, creation date, author, etc.). The vocabulary for this life-cycle management metadata is mainly provided by the management ontology module.

Basically, this module contains a set of definitions of properties which can be used to describe instances of the class *NewsItem*, defined in the content module (see Section 3.4). We can divide these properties according to their function into:

**Management metadata properties** contain information relevant to the management process of a news item, like for example its urgency (*has\_urgency*), its status (*has\_status*, it is *usable*, *cancelled*, etc.) or the UTC time it has been created (*has\_creation\_time\_UTC*). An interesting special kind of properties in this group are, for example, those which relate two news items, stating that one is an update of the other (*is\_update\_of*) or that one is derived from the other (*is\_derived\_from*).

**Administrative metadata properties** provide information about the provenance of a certain news item. Examples of this kind of properties are: *has\_provider*, *has\_creator* or *has\_contributor*.

**Rights-related metadata properties** contain information about the rights pertaining to a certain news item. For example: *has\_rights\_usage\_type*, *has\_rights\_geography*, *has\_rights\_start\_date* or *has\_rights\_holder*.

**Descriptive metadata properties** with information describing the contents of a news item as a whole like, for example, the

<sup>18</sup> <http://www.w3.org/2004/02/skos>.

The screenshot shows two main panels: 'SUBCLASS EXPLORER' on the left and 'CLASS EDITOR' on the right.

**SUBCLASS EXPLORER:** Shows a hierarchical tree of classes. The root is 'owl:Thing'. Underneath is 'content:NewsItem'. Below that are several subclasses, including '\_01000000' through '\_17000000'. The class '\_15000000' is currently selected and highlighted.

**CLASS EDITOR:** Shows the details for the selected class '15000000'. It contains a table with the following properties:

Property	Description
rdfs:comment	Competitive exercise involving physical effort. Organizations and bo
rdfs:comment	Esercizio competitivo che richiede uno sforzo fisico. Organizzazioni
rdfs:comment	Ejercicio competitivo, que involucra esfuerzo físico. Organizaciones
rdfs:label	sport
rdfs:label	Sport
rdfs:label	deporte

Below the table, there are icons for class operations and a small diagram showing the class 'content:NewsItem'.

Fig. 1. An excerpt of the NEWS categorization module taxonomy.

location where the news story has occurred (*has\_location*), the language used in elaborating the piece of news (*has\_language*), or the expected target audience of the item (*has\_interest*). A property which is specially important is the one which relates the news item with the instances that are explicitly mentioned in the contents of that news item (*has\_instance\_occurrence*). These instances are provided by the content annotation module (see Section 3.4).

**Physical, content-related metadata properties** also provide information about the contents, but from the physical point of view: its binary size (*has\_binary\_size*), the codec used in coding it (for multimedia news items, *has\_audio\_codec*, *has\_video\_codec*), its MIME type (*has\_mime\_type*) or the number of words that it contains (for textual news items, *has\_word\_count*), among others.

**Others** like keywords (*has\_keyword*), the headline of the news item (*has\_headline*) or its abstract (*has\_abstract*) are also included in the description as specific properties.

Some of the properties included in the module (*has\_creator* is an example) directly adapt metadata properties specified in PRISM, or DC, for instance, by limiting the adequate domain and range for that properties. Other properties in the module are inspired by NITF or NewsML schemas. That is the case, for instance, of

*has\_headline*, which is defined taking into account that both NITF (element *headline*) and NewsML (element *Headline*) define elements with the purpose of specifying the news item headline.

Most of the IPTC NewsCodes are also taken into account in the management module. NewsCodes sets are usually represented as classes with a set of associated individuals. For instance, in order to use the values standardized in the *Status* NewsCodes set, we added a class, named *Status*, to the management module of the NEWS ontology and included as instances of that class each of the values defined in the IPTC vocabulary; that is, at the time of writing, *Canceled*, *Embargoed*, *Usable* and *Withheld*. A property (*has\_status*) is also defined so that a concrete instance of news item can be linked to the proper value of status.

The result is a module consisting of more than 50 classes, more than 90 properties and over 500 instances. A more detailed description of this module and its relations with metadata standards can be found in the NEWS deliverables available at [Web Technologies Laboratory \(2008\)](#).

### 3.4. Content annotation module

The content annotation module provides a basic vocabulary for news content annotation. It is also used to tie together the different ontology modules: it defines the class *NewsItem*, that is the root

class of the categorization module, and the domain of almost all the properties defined in the management module.

One of the difficulties of building an ontology to be used to annotate the contents of a news item, is that almost everything in the world can appear in a piece of news. Providing a formal model of all the things in the world is a huge task but, fortunately, agencies are not interested in formally annotating all the contents of a news item, but only some important entities: persons, organizations, locations, etc.

In order to address this requirement, in NEWS we relied on a top-level ontology which provides general concepts that are independent of the application domain. In order to reuse resources already available in the state of the art, our content module was built taking SUMO (Niles & Pease, 2001) and MILO (Niles & Terry, 2004) as basis. These ontologies were developed by the IEEE Standard Upper Ontology Working Group<sup>19</sup> with the aim of developing a standard upper ontology to support applications such as data interoperability, information retrieval or natural language processing.

One of the main problems when building this ontology module was to select what concepts from SUMO/MILO should be included or discarded. SUMO/MILO are very wide ontologies with hundreds of concepts, but we believe that some of these concepts (like *Arthropod* or *ComplexNumber*) were of little utility in the news domain. So, a pruning strategy should be applied to filter non-relevant concepts. The approach we followed to address this problem was to use a middle-out strategy, like suggested in ontology building methodologies (Uschold & Grnninger, 1996). The main idea was to look at the basic entities included in journalism standards (inline annotation elements from NITF, values of IPTC Topic Type News-Codes, etc.) and map these entities to classes in SUMO/MILO. These classes should be included in our ontology in order to be compatible with standards. Once we had these seed classes we started the process of pruning. For each seed we looked for its ancestors until the top concept of the ontology (*Entity*) and included them, obtaining a seed tree as result. For each class included in this tree, we then looked for non-included descendants. The decision about whether to include them or not is based on criteria like:

- Relation to at least one of the categories in the first level of the categorization taxonomy. For instance, the concept *Book* is related to IPTC subject 01000000, which can be used as a type of news items talking about arts, culture and entertainment. We can think, for example, of a news item which informs about the presentation of a new book of a well-known writer.
- Usefulness in other components of the NEWS ontology. For example, the concept *HumanLanguage* is useful for the management module, because instances of this concept are used as the range of the property *has\_language* of class *NewsItem*, which relates the instance of news item with its language.

The resultant taxonomy was then reviewed in order to add/remove classes if needed.

After deciding the concept taxonomy of the content module, we proceeded by analyzing SUMO/MILO relations which included at least one of the concepts of the taxonomy. Again, using criteria of usefulness for news annotation, we decided which relations should be included. As a result, new concepts had to be included. When both concepts and relations from SUMO/MILO were defined, we proceeded by adding instances. Those instances were taken from different information sources like ISO country codes standard (ISO 3166:1997–1), CIA WorldFact Book (CIA, 2004), NASDAQ companies codes, SUMO/MILO instances, Wikipedia<sup>20</sup> articles, etc.

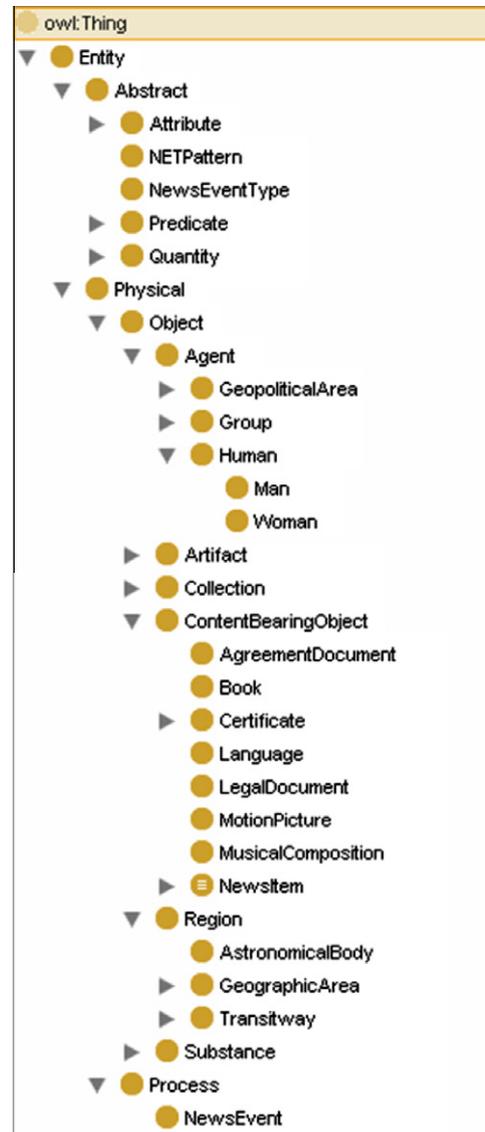


Fig. 2. Taxonomy of NEWS content annotation module.

The result is a generic top-level ontology with more than 200 classes and more than 100 properties. It also contains more than 6000 instances of different classes: countries, languages, currencies, cities, companies, persons, etc. Fig. 2 shows the tree with some of the elements included in this version.

A concrete part of the content annotation module that deserves special attention is the event model. According to the requirements defined by the news agencies that participated in the NEWS project, news events mentioned in a news item are one of the most interesting types of content metadata. In NEWS we introduced the concept of news event in our ontology and developed a specific model that allows to define a number of properties that characterize an event. This model is described with detail in the following subsection.

#### 3.4.1. Event model

An event in the news domain is defined as “something that happens and is subject to news coverage” (IPTC EventsML-G2 specification<sup>21</sup>). Of course, materializing this general definition in concrete mechanisms to identify events is difficult. In order to recognize in

<sup>19</sup> <http://suo.ieee.org>.

<sup>20</sup> <http://www.wikipedia.org>.

<sup>21</sup> <http://www.iptc.org/EventsML/>.

the contents of news items the mention of events (e.g. *Oracle buys SUN*) of a certain event type (e.g. *company acquisition*), three pieces of information are used in NEWS:

- The **categories** of the news item. A set of *compatible categories* are defined for each different event type. For instance, the *economy, business and finance* category (04000000) is included within the compatible categories of the *company acquisition* event type. The automatic event-detection system of NEWS checks that a news item is classified in at least one of the compatible categories of a certain event type before annotating the news item with a concrete event of such type.
- The **instances** the news item has been annotated with. The mention of a particular instance in the contents of a news item is sometimes a necessary requirement to decide that the news item describes an event of a certain type. For instance, the event-detection system requires that the instance *Iraq* is mentioned in the news item to recognize an event of type *Iraq war terrorist action*.
- **Patterns** detected in the contents of the news item. In NEWS, an event type definition includes a set of language-dependent patterns. These patterns have an SPO (Subject-Predicate-Object) structure, where subject and object are defined as one of the following three entity types: person (PER), organization (ORG) or location (LOC), and the predicate can include a verb and/or some free text. For instance, the *company acquisition* event type has, among others, the following associated patterns:

```
en: ORG buy ORG
en: ORG bought ORG
es: ORG compra ORG
es: PER ha comprado ORG
it: ORG acquistati ORG
it: PER ha acquistato ORG
```

An event of a certain event type will be recognized in a news item when the following conditions hold:

- (i) The news item belongs to at least one of the compatible categories of the event type and mentions all the required instances (if any).
- (ii) There is a matching between any of the patterns defined for the event type and the text of the news item.

A specific pattern recognition software based on Apache Lucene<sup>22</sup> was developed in NEWS to analyze the textual contents of news items and to detect the occurrence of event type patterns. Using that software, the sentence *Oracle Corp. has bought yesterday SUN Microsystems*, would be matched with the pattern *en: ORG bought ORG*. A detailed description of that software can be found in Fernández et al. (2006).

In the content annotation module of the NEWS ontology, events are modelled as instances of the class *NewsEvent*, subclass of class *Process*. The event types are instances of the class *NewsEventType*. Hierarchies of event types could be defined by defining subclasses of *NewsEventType* if necessary. The property *hasEventType* is used in order to relate a concrete event with its type. Properties have also been created to define the patterns of an event type and its required instances and compatible categories.

As aforementioned, each event type can have several related patterns and compatible categories. Not all of these patterns and categories have to be present in each concrete event occurrence. For instance, though the *company acquisition* event type has several associated patterns, a news item of category

04000000 where only one of these patterns is detected would be annotated with and event of that event type. Furthermore, the patterns defined in the event type are generic, in the sense that they use PER, ORG and LOC as subject or predicate (e.g. *en: ORG bought ORG*), whereas concrete occurrences in events involve concrete entities (e.g. *Oracle bought SUN*). Due to this, properties are provided in the ontology to define the instances, categories and patterns involved in a concrete event detected by the system, as well as the timestamp of the news item where the event was detected.

For each event type defined in the ontology, the class of all events of that type and the class of all the news items in which an event of that type is mentioned can be defined in OWL, like in the following example:

```
<owl:Class rdf:ID="CompanyAcquisitionEvents">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:hasValue>
        <NewsEventType
          rdf:ID="CompanyAcquisitionEventType"/>
      </owl:hasValue>
    <owl:onProperty>
      <owl:ObjectProperty rdf:ID="hasEventType"/>
    </owl:onProperty>
  </owl:Restriction>
</owl:equivalentClass>
<rdfs:subClassOf>
  <owl:Class rdf:ID="NewsEvent"/>
</rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="CompanyAcquisitionNewsItems">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf
        rdf:parseType="Collection">
          <owl:Restriction>
            <owl:someValuesFrom
              rdf:resource="CompanyAcquisitionE-
                vents"/>
            <owl:onProperty>
              <owl:ObjectProperty
                rdf:ID="has_instance_ocurrence"/>
            </owl:onProperty>
          </owl:Restriction>
          <owl:Class rdf:about="NewsItem"/>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
  <rdfs:subClassOf rdf:resource="04000000"/>
</owl:Class>
```

First the class *CompanyAcquisitionEvents* is defined as the class of all the events that have type *CompanyAcquisitionEventType*. Then the class *CompanyAcquisitionNewsItems* is defined as the class of all the news items with an event of type *CompanyAcquisitionEventType* (thus belonging to class *CompanyAcquisitionEvents*). We also state that class *CompanyAcquisitionNewsItems* is a subclass of class 04000000 (economy, business and finance). This would allow to infer the category of a news item if it would have been manually annotated by a journalist with an event of type *CompanyAcquisitionEventType* using the NEWS user interface.

Note that in order to recognize automatically the event using the model presented here it must have been previously annotated

<sup>22</sup> <http://lucene.apache.org/>.

with a compatible category and, therefore, the *subClassOf* axiom would not provide additional information.

#### 4. Evolving the ontology design

During the project life-cycle, the NEWS ontology evolved to better cover the needs of the news agencies. We started by developing a small prototype in RDF (S) (Fernández & Sánchez, 2004), which already included a reduced version of the three modules. For instance, the content annotation module contained only some classes to represent the basic entities defined in the NITF schema for inline annotation purposes, and the management module did not take into account properties defined in NewsML to describe multimedia news items.

A second version of the ontology was later developed (Fernández, Sánchez, Blázquez, & Villamor, 2006) extending the content and management modules. This version was implemented in TRIPLE (Sintek & Decker, 2002) in order to include expressive reasoning rules to be used in intelligent search facilities. Unfortunately, the powerful inference capabilities offered by TRIPLE were time-consuming, specially for an environment with time restrictions and where huge volumes of content and metadata need to be managed. Due to this, we decided to restrict the reasoning services to be used in the NEWS project to query expansion. Even though, ontologies were defined using OWL DL, for modelling purposes and to enable possible future extensions of the reasoning services of the NEWS system.

Apart from a change in the language used to implement the ontology, the modules also suffered modifications. The ontology in Fernández et al. (2006) included additional modules, that defined two aspects considered of interest for the project: time and measure units. A third module, the structural module, was also developed to provide compatibility between TRIPLE and RDFS semantics. In the current version of the ontology these three modules disappear. Time aspects are currently modeled in the OWL ontology using XML schema datatypes like *xsd:date* or *xsd:dateTime*. We have made this change because we have found that we can deal with time information using imperative algorithms much more efficiently than with a reasoner. A very basic model of measure units is provided in the content module, and is enough for the project purposes. Again, the processing of units of measure is done by means of imperative algorithms. The concept of unit of measure remains in the ontology only for modelling purposes. Finally, the structural module is not needed any more, because we are not using TRIPLE right now. The current version includes also aspects not covered by the previous two, like the event model described in Section 3.4.1 and multilingual support based on SKOS.

#### 5. Ontology maintenance: NEWSPlugin

As we have seen in previous sections, important parts of the NEWS ontology are inspired by IPTC standards like NewsCodes. The contents in these NewsCodes are updated from time to time by the IPTC. For instance, new categories may be added to the Subject Code NewsCodes, which are the nucleus of the categorization module of the NEWS ontology.

As we have also seen, the content annotation module of the NEWS ontology provides instances which are used to annotate news items. As time goes by, the content module should evolve to reflect the changes in the world: new persons become famous, new companies appear, some countries disappear, etc.

A process of ontology maintenance is, thus, needed in order to keep the ontology useful for its purposes. As the NEWS ontology is implemented using standard languages (OWL), in principle it is pos-

sible to use any of the tools available in the state of the art, like Protégé,<sup>23</sup> for ontology maintenance. An expert (usually a knowledge engineer or documentalist working in the news agency) can open the ontology and carry out manually the necessary modifications.

This manual maintenance process is both time-consuming and error-prone, specially if we take into account that, due to the multilingualism requirement, the elements in the ontology should be tagged in three different languages. In order to help the knowledge engineer and simplify the NEWS ontology maintenance process, more specific tools are required. In our case, we have developed an extension of the Protégé ontology editor named NEWSPlugin.

In Fig. 3 the graphical user interface of the NEWSPlugin can be seen. The main functionalities of this tool are described in the following subsections.

##### 5.1. Maintenance of the categorization module

The main maintenance operation involving the categorization module is the addition of new categories as soon as the IPTC adds them to the Subject Code NewsCodes. For this process, the plugin provides a button which, when clicked, downloads the Subject Code vocabulary, available as XML files from the IPTC Web site, and processes these files obtaining the categories included at that moment. These categories are compared with the ones already available at the ontology to include the new ones. In order to support the multilingualism requirement, and taking into account those versions of the Subject Code in English, Italian and Spanish are available at the IPTC Web site, it is simply needed to configure the plugin to download the three versions of the vocabulary and use them to extract the labels and comments in the different languages.

Additionally, it is possible to remove the categories which are potentially deprecated and removed from IPTC standard, thus completely syncing the IPTC Subject Code NewsCodes with the categories available at the ontology. For this purpose, the plugin asks the user whether he/she wants to delete them. Nevertheless, this operation should be performed with care, because it would involve changing the annotations in the news item knowledge base, moving the news items annotated with the removed categories to the direct superclasses of these categories. For instance, if the category of identifier 15001002 is removed, then all news items belonging to this category should be moved to category 15001000. The NEWSPlugin does not handle the NEWS knowledge base of annotated news items, due to its enormous size, so it only provides the possibility of updating the ontology removing the deprecated classes, but leaving unaltered the knowledge base. Of course the plugin notifies to the user the categories which have been removed, in order to help him/her to carry out additional maintenance tasks in the knowledge base if required.

##### 5.2. Maintenance of the content annotation module

Regarding the content annotation module, the main operation provided by the NEWSPlugin is populating the classes of such module. The plugin shows the taxonomy of the NEWS ontology at the left of the window and the knowledge engineer can search or browse the ontology tree to find the class to be populated. Once the class is selected, the user needs to provide the label (and optionally the definition) of the new instance to be added in one of the three languages of interest in the context of the project. By clicking on a *Translation* button, the plugin invokes the Google Translator service<sup>24</sup> to obtain the representations of the labels

<sup>23</sup> <http://protege.stanford.edu>.

<sup>24</sup> [http://www.google.com/translate\\_t](http://www.google.com/translate_t).

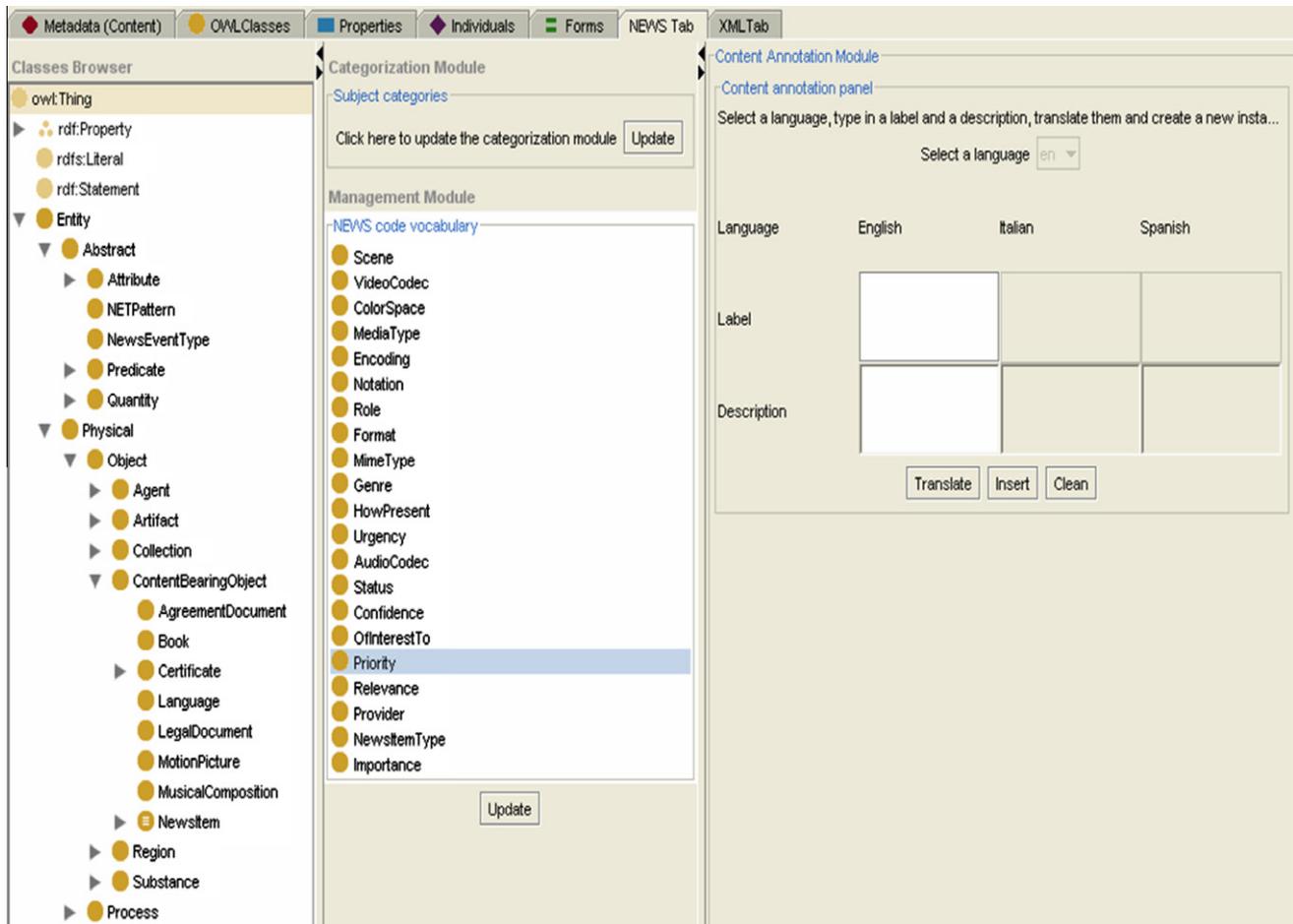


Fig. 3. Screen capture showing the graphical interface of the NEWSPlugin.

(and possibly the definitions) in the other two languages. As an automatic translation service is used, the knowledge engineer should carefully check the automatically generated labels and definitions. Once the revision process is finished, the user can simply click on a different button (*Insert*) that populates the previously selected class with the news instance, also defining the triples used to link that instance to its labels and definitions.

### 5.3. Maintenance of the management module

As explained in Section 3.3 the management module of the NEWS ontology is partially based on IPTC's NewsCodes. Most of these IPTC NewsCodes vocabularies are included in the module as classes with a set of associated individuals. Using the NEWSPlugin, the knowledge engineer can update the management module when the NewsCodes vocabularies change. In order to do so, a list with the classes in the module taken from the NewsCodes is provided at the NEWSPlugin interface. The user can select the class to be updated and, by clicking on the *Update* button, the instances in the associated class in the management module are updated. This process involves downloading from the IPTC Web site the XML file where the correspondent NewsCodes are defined, and parsing that file to obtain the available values at that moment. These values are compared with the ones in the ontology in order to decide whether or not to include them.

Removing instances, which are declared deprecated in the IPTC vocabulary and later removed from it, is in theory possible using the plugin. However, this needs to be done with care, because it involves the modification of the news item knowledge base, an oper-

ation which is not supported by the plugin. Again, as in the case of categories, when the plugin is used to delete instances, it notifies the fact to the user and he/she can abort this deletion.

One of the problems with this part of the plugin comes from the fact that most of the NewsCodes (apart from the Subject Code vocabulary mentioned above) defined by IPTC do not have alternative versions in different languages, only in English. This makes impossible the approach used with the categorization module, which relies in the data provided by the IPTC. In order to address this issue and keep the ontology multilingual, automatic translation is again used.

## 6. The NEWS ontology in practice

In the previous sections, we have described the current status of the NEWS ontology. But an ontology description would not be complete if we do not mention its actual usages within the intelligent system it was designed for. Taking into account the main application areas of ontologies in intelligent systems, described in *OntoWeb project consortium (2003)*, we classify the usages of the NEWS ontology inside the NEWS system as follows:

### 6.1. Information retrieval

The ontology acts as a vocabulary for news item categorization and annotation. Semantic metadata makes possible to search news items by category, instances they refer to, events they talk about, etc. Search is multilingual because queries are based on the identi-

fiers of the categories/instances/events in the ontology, which are language-independent. The formal model of the ontology allows simple reasoning used for query expansion (for instance, including subcategories in category-based search). It also makes possible to look for news items annotated with certain instances when the only information that we have about such instances is that a concrete property holds for them (for instance, look for news items annotated with instances that represent persons who *work\_at* Renault F1 Team).

### 6.2. E-commerce

The ontology has applications both in B2B and in B2C scenarios. In the case of B2B scenarios, the ontology serves as a formal model of the domain shared by the agencies involved in the project. This fosters interoperability between them by removing the burden due to the usage of different proprietary metadata formats/vocabularies. Selling/buying news items to/from the other agency should become a simpler task by agreeing to the NEWS ontology. In the B2C scenario, the ontology can be used with personalization purposes, allowing final users to define their interests based on the vocabulary provided by the NEWS ontology. A client profile includes references to elements like categories and concrete events/instances of interest. The news items whose annotations match the profile are sent to the client through customized *push* services like personalized RSS<sup>25</sup> channels.

### 6.3. Information management

The management module of the NEWS ontology provides a vocabulary which allows the definition of metadata regarding the news item life-cycle management. Aspects like authorship information, rights, etc. are covered by this module. An application of such information is, for example, finding adequate journalists to cover a certain event (e.g. a soccer match, a political meeting, etc.). This is simply implemented by searching the NEWS system datawarehouse to find the authors of former news items reporting on similar events. Another example: management metadata could be used to define the urgency of a certain news item. This is helpful in deciding the order in which news items are sent to clients or are processed by certain procedures in the agency's production workflow (automatic annotation, for instance).

## 7. Related work

The news domain has a number of features that make it interesting for running experiences using artificial intelligence techniques in real business: data heterogeneity, huge amounts of information to manage, multilingualism, economical and social interest, etc. Taking this into account, it is not strange to find projects in the state of the art related to the topic of applying knowledge management and (more recently) Semantic Web technologies to the journalism domain: PlanetOnto (Domingue & Motta, 2000), NAMIC (Basili et al., 2001), Neptuno (Castells et al., 2004) or PENG (Pasi, Bordogna, & Villa, 2006) are examples of initiatives similar to NEWS in aims and scope. But some of the requirements that the usage scenario imposes in NEWS (and thus in the NEWS ontology) are not present in the aforementioned initiatives: multilingualism is not considered in PlanetOnto, Neptuno or PENG, compatibility with professional journalism standards is missed in PlanetOnto, support for news item management metadata is out of the scope of NAMIC, etc.

Regarding concrete ontologies to model the news domain, we can divide our analysis in two main categories:

- (i) Ontologies that are not explicitly designed to be used in a professional journalism scenario, and thus do not provide all the required modelling aspects (for instance, do not take into account management metadata or compatibility with journalism standards). Due to this, these ontologies are not enough by themselves for the purposes of the NEWS project but, nevertheless, they should be taken into account, because they provide formal models that can be (at least partially) adapted and reused in the NEWS ontology. SUMO (Niles & Pease, 2001), MILO (Niles & Terry, 2004), KIMO (Popov, Kiryalov, Ognyanoff, Manov, & Kirilov, 2004), PROTON (Terziev, Kiryakov, & Manov, 2005), Upper Cyc Ontology (Lenat & Guha, 1990), OpenCyc (OpenCyc.org, 2002), the EuroWordNet Top Level ontology (Vossen, 1998) and the TAP knowledge base (Guha & McCool, 2003) are the most prominent examples of this group. The content annotation module of the NEWS ontology is partially inspired by SUMO and MILO.
- (ii) Initiatives that develop and use ontologies in the news domain. In this sense, the projects mentioned above: PlanetOnto, Neptuno, PENG, NAMIC, use ontologies, but, as we have said, the scopes of these projects lack some of the aspects covered by NEWS and, thus, their ontologies also lack some of the characteristics provided by the NEWS ontology. An additional related work, apart from these research projects, is (Lee, Chen, & Jian, 2003), where DAML+OIL ontologies are used to support the process of event extraction for Chinese news summarization. These ontologies do not have as requirements to be multilingual, nor compatible with professional journalism standards. News item management metadata are also not taken into account in the model. In Wang, Zhe, Kang, Wang, and Chen (2008), an ontology for knowledge about news regarding financial instruments is developed. It does not cover management metadata aspects and is designed to be used as model for news items in a very concrete domain, whereas the NEWS ontology does not impose domain restrictions. Finally, in García, Perdrix, Gil, and Oliva (2008), a tool to convert from XML schemas to OWL is defined. This tool is used to convert the schemas of several standards like Subject Code News-Codes, NewsML or NITF into OWL, implementing an ontological framework which is being tested in a newspaper. This work does take into account standards, and includes management metadata issues, but lacks multilingual aspects and offers a much more limited content annotation model than the one provided by the NEWS ontology.

## 8. Conclusions

In this paper we have explained thoroughly the ontology development process that we have carried out in the NEWS project. A maintenance tool for this ontology has been also described.

The ontology covers the different types of metadata that can be attached to a news item: management, categorization and content metadata.

Basic principles that guided the ontology development process were:

- Ontology reuse: we have defined our content module based on the well-known top-level ontology SUMO and its complement MILO.
- Standards compliance: IPTC metadata standards for the news domain have been used as the basis for the management and categorization modules, together with the metadata standards

<sup>25</sup> <http://www.rss-specifications.com/rss-specifications.htm>.

Dublin Core and PRISM. The SKOS initiative has also been reused for multilingualism support.

- News agencies requirements: news agencies are used to deal with metadata, although usually in an informal way. Therefore, they have a clear view of their needs. The requirements stated by the news agencies that participated in the NEWS project, EFE and ANSA, were carefully taken into account in the ontology development.
- Customization and maintainability: the content module includes top-level concepts (*Entity, Object, Process, ...*), that will make easy to extend such module as desired. This can be used to adapt the content module to the requirements of a given news agency (or a mass media). The same can be said about the categorization module, because new news item categories can be defined as subclasses of the standard IPTC categorization system, though it is not recommended for interoperability reasons.

The maintenance tool, the NEWSPlugin, extends the well-known ontology editor Protégé and allows to update all the ontology modules. It can also be used to keep in sync the management and categorization modules with the related IPTC standards when the later evolve.

As a result, the whole development presented here, including the ontology and its maintenance tool, NewsPlugin, fills the gap existing in the current news market, evolving to a more automatic and comprehensive model for news agencies. Both the ontology and the maintenance tool can be downloaded from [Web Technologies Laboratory \(2008\)](#).

## Acknowledgements

This work has been partially funded by the European Commission under contract FP6–001906 in the framework of the Information Society Technologies (IST) programme and by the Spanish Government under contract TSI2007–65393–c02–01. We wish also to acknowledge to the partners of the NEWS project for their cooperation.

## References

- Basili, R., Pazienza, M. T., Zanzotto, F., Catizone, R., Setzer, A., & Webb, N., et al. (2001). Multilingual authoring: The NAMIC approach. In *Proceedings of the workshop on human language technology and knowledge management, association for computational linguistics*.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.
- Castells, P., Perdrix, F., Pulido, E., Rico, M., Benjamins, R., Contreras, J., et al. (2004). Neptuno: Semantic web technologies for a digital newspaper archive. In C. Bussler, J. Davies, D. Fensel, & R. Studer (Eds.), *The semantic web: Research and applications: First european semantic web symposium, ESWS 2004*. LNCS (Vol. 3053). Springer.
- CIA, World Factbook. (2004). <http://www.cia.gov/cia/publications/factbook/> (January 2004).
- Domingue, J., & Motta, E. (2000). PlanetOnto: From news publishing to integrated knowledge management support. *IEEE Intelligent Systems*, 15(3), 26–32.
- Fernández, N., & Sánchez, L. (2004). Building an ontology for NEWS applications. In *Poster proceedings of the third international semantic web conference, ISWC 2004, Hiroshima, Japan*.
- Fernández, N., Blázquez, J. M., Arias, J., Sánchez, L., Sintek, M., Bernardi, A., et al. (2006). NEWS: Bringing semantic web technologies into news agencies. In *Proceedings of the fifth international semantic web conference, ISWC 2006*. LNCS (Vol. 4273). Athens, Georgia, USA: Springer.
- Fernández, N., Blázquez, J., Sánchez, L., & Bernardi, A. (2007). IdentityRank: Named entity disambiguation in the context of the NEWS project. In *Proceedings of the fourth european semantic web conference, ESWC 2007*. LNCS (Vol. 4519). Innsbruck, Austria: Springer.
- Fernández, N., Sánchez, L., Blázquez, J. M., & Villamor, J. (2006). A handbook of principles, concepts and applications in information Systems. *Integrated series in information systems* (Vol. 14, pp. 887–923). Springer. [Chapter :The NEWS Ontology for Professional Journalism Applications].
- García, R., Perdrix, F., Gil, R., & Oliva, M. (2008). The semantic web as a newspaper media convergence facilitator. *Journal of Web Semantics*, 6(2), 151–161.
- Guha, R., & McCool, R. (2003). TAP: A semantic web platform. *Computer Networks*, 42, 557–577.
- Lee, C.-S., Chen, Y.-J., & Jian, Z.-W. (2003). Ontology-based fuzzy event extraction agent for chinese e-news summarization. *Expert Systems with Applications*, 25(3), 431–447.
- Lenat, D., & Guha, R. (1990). *Building large knowledge-based systems: representation and inference in the cyc project*.
- Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the second international conference on formal ontology in information systems*.
- Niles, I., & Terry, A. (2004). The MLO: A general-purpose, mid-level ontology. In *2004 International conference on information and knowledge engineering (IKE'04)*.
- OntoWeb project consortium. (2003). Deliverable D2.3: Successful scenarios for ontology based applications, <<http://www.ontoweb.org/About/Deliverables/>>.
- OpenCyc.org . (2002). OpenCyc. Formalized Common Language, <<http://www.opencyc.org>>.
- Pasi, G., Bordogna, G., & Villa, R. (2006). The PENG system: Practice and experience. In *Proceedings of the 17th international conference on database and expert systems applications (DEXA 06)*.
- Popov, B., Kiryalov, A., Ognyanoff, D., Manov, D., & Kirilov, A. (2004). KIM – A semantic platform for information extraction and retrieval. *Journal of Natural Language Engineering*, 10(3–4), 375–392.
- Sintek, M., & Decker, S. (2002). A query, inference, and transformation language for the semantic web. In *International semantic web conference (ISWC)*.
- Terziev, I., Kiryalov, A., & Manov, D. (2005). D1.8.1 Base upper-level ontology (BULO) Guidance, <[http://proton.semanticweb.org/D1\\_8\\_1.pdf](http://proton.semanticweb.org/D1_8_1.pdf)>.
- Uschold, M., & Grnninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93–155.
- Vossen, P. (Ed.). (1998). *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic Publishers.
- Wang, S., Zhe, Z., Kang, Y., Wang, H., & Chen, X. (2008). An ontology for causal relationships between news and financial instruments. *Expert Systems with Applications*, 35(3), 569–580.
- Web Technologies Laboratory (WebTLab). (2008). WebTLab: Web Technologies Laboratory, research results, <<http://webtlab.it.uc3m.es/results/index.html>>.