NeOn-project.org

NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – "Semantic-based knowledge and content systems"

D2.4.2 Multilingual and Localization Support for Ontologies

Deliverable Co-ordinator: Elena Montiel-Ponsoda (UPM), Wim Peters (USFD)

Deliverable Co-ordinating Institution: Universidad Politécnica de Madrid (UPM), University of Sheffield (USFD)

Other Authors: Guadalupe Aguado de Cea (UPM); Mauricio Espinoza (UPM); Asunción Gómez-Pérez (UMP); Margherita Sini (FAO)

This deliverable aims at providing support for the localization of NeOn ontologies, and as a result at obtaining multilingual ontologies. The main contributions of this document are: (1) an enhanced version of the linguistic model for storing multilingual information in NeOn ontologies, the so-called *Linguistic Information Repository* (LIR), and the initial evaluation results; and (2) a description of the new version of the *LabelTranslator* NeOn plug-in that supports the LIR model and the automatic localization of NeOn ontologies thanks to its three main components: the GUI, the localization algorithm, and the repository. Initial versions of both, the LIR model and the LabelTranslator plug-in, were presented in *D2.4.1 Multilingual Ontology Support* [18] of this project. The versions presented in this deliverable give support to the Multilingual Ontology Meta-Model proposed for NeOn in D2.4.1.

Document Identifier:	NEON/2008/D2.4.2/v1.0	Date due:	August 29th, 2008
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	August 29th, 2008
Project start date:	March 1, 2006	Version:	v1.0
Project duration:	4 years	State:	Final
		Distribution:	Public

2006–2008 © Copyright lies with the respective authors and their institutions.

NeOn Consortium

This document is a part of the NeOn research project funded by the IST Programme of the Commission of the European Communities by the grant number IST-2005-027595. The following partners are involved in the project:

Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Martin Dzbor, Enrico Motta E-mail address: {m.dzbor, e.motta} @open.ac.uk	**Universität Karlsruhe – TH (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 11 D-76128 Karlsruhe, Germany Contact person: Peter Haase E-mail address: pha@aifb.uni-karlsruhe.de
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es	**Software AG (SAG)** Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com
Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com	**Institut 'Jožef Stefan' (JSI)** Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si
Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l'Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr	**University of Sheffield (USFD)** Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk
Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de	**Consiglio Nazionale delle Ricerche (CNR)** Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it
Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de	**Food and Agriculture Organization of the United Nations (FAO)** Viale delle Terme di Caracalla 1 00100 Rome Italy Contact person: Marta Iglesias E-mail address: marta.iglesias@fao.org
Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.parientelobo@atosorigin.com	**Laboratorios KIN, S.A. (KIN)** C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

UPM

USFD

FAO

Change Log

Version	Date	Amended by	Changes
0.10	16-06-2008	W. Peters	Table of Contents
0.15	07-07-2008	E. Montiel-Ponsoda	Executive Summary and Introduction
0.20	14-07-2008	W. Peters, E. Montiel-Ponsoda	Section 2. Description of the new version of the LIR model
0.25	22-07-08	E. Montiel-Ponsoda	Revision of Introduction
0.30	28-07-2008	E. Montiel-Ponsoda	Revision of Section 2
0.35	31-07-2008	M. Sini, E. Montiel-Ponsoda	Section 4. Description of LIR initial tests against FAO resources
0.40	11-08-2008	A. Gómez-Pérez	Revision and proposals of changes
0.50	11-08-2008	M. Espinoza	Section 6. Description of the new version of the localization algorithm and LabelTranslator NeOn plug-in
0.55	12-08-2008	E. Montiel-Ponsoda	Integration of Section 6 in the document, and revision
0.60	14-08-2008	E. Montiel-Ponsoda	Revision and proposals of changes
0.65	20-08-2008	G. Aguado de Cea	Revision and proposal of changes
0.70	20-08-2008	M. Sini	Revision of Sections 2 and 4
0.75	21-08-2008	W. Peters	Revision and proposals of changes
0.80	21-08-2008	M. Espinoza	Revision of Section 6 and proposals of changes
0.85	27-08-2008	W. Peters	Section 5: LIR discussion and future lines of work
0.90	28-08-2008	E. Montiel-Ponsoda, E. Espinoza	Final revision
0.95	16-09-2008	E. Montiel-Ponsoda	Inclusion of Q.A. reviewer's comments in Sections 2, 3 and 4
0.96	17-09-2008	W.Peters	Final version
1.0	14-10-2008	A. Tumilowicz	Final QA

2006–2008 © Copyright lies with the respective authors and their institutions.

Executive Summary

Multilinguality in ontologies has become an impending need for institutions worldwide with valuable linguistic resources in different natural languages. Since most ontologies are developed in one language, obtaining multilingual ontologies implies to *localize* or "adapt them to a concrete language and culture community", as defined in [28,29]. For this end, we designed (1) a model called *Linguistic Information Repository* (LIR) that associated with the ontology meta-model would provide NeOn ontologies with the necessary linguistic information in several natural languages to allow its localization, and (2) the *LabelTranslator* NeOn plug-in, which based on a translation algorithm would allow the automatic translation of ontology labels from a source natural language to one or several target languages.

The LIR model [24] was introduced in D2.4.1 *Multilingual Ontology Support* [18], and its development was the result of an exhaustive analysis of the following aspects:

a) methods, techniques and tools used in the localization of well-known lexical resources (LRs), and representation models used by those LRs to store multilingual information (sections 5 to 11 of D2.4.1)

b) current standards for encoding lexical (LMF) and terminological information (TMF, SKOS Core) in order to guarantee interoperability with multilingual resources (section 12.2 of D2.4.1)

c) multilinguality requirements expressed in NeOn WPs (WP1, WP6, WP7 and WP8) (section 12.3 of D2.4.1)

d) possibilities for representing multilinguality at the various levels of a Knowledge Representation Base: Interface level, Metadata level, Knowledge Representation level, and Data level (sections 13 and 14 of D2.4.1)

e) advantages and disadvantages of 2 different modalities for representing multilingual information in ontologies: (1) inclusion of multilingual data in the ontology meta-model, (2) association of the ontology meta-model to an external linguistic model (sections 14.2 and 15)

After this thorough investigation on methods, techniques and tools for the localization of lexical and ontological resources and on how multilinguality could be represented in ontologies, D2.4.1 (Section 17) was also devoted to the description of the first version of the LabelTranslator plug-in. This version of the tool provided support to the translation of ontology labels embedded in the ontology. A user interface was provided, and the tool was integrated as a plug-in in the NeOn Toolkit. The main functionalities and components of LabelTranslator were exhaustively explained in sections 17.3.1 and 17.3.2 of D2.4.1. However, at that time the current state of the plug-in only used the linguistic features available in the NeOn meta-model characterized by simple references between concepts and labels, as offered by the standard `owl:comment` and `rdfs:label` properties. Therefore, further development was needed to give support to the great amount of linguistic information contained in the LIR, which was deemed necessary for an adequate localization of NeOn ontologies.

Thus, the main goal of this deliverable is to present enhanced versions of the LIR model and the LabelTranslator plug-in.

The scope and main contributions of this deliverable are:

1) A description of the second version of the LIR model that provides NeOn ontologies with multilingual information, as well as its implementation in OWL.

2) An overview of the initial experiments on the evaluation of the LIR against FAO lexical resources.

3) A detailed description of the second version of the LabelTranslator NeOn plug-in laying particular emphasis on the three main components of its new architecture: a) the GUI, b) the new translation algorithm, and c) the repository component.

2006–2008 © Copyright lies with the respective authors and their institutions.

Table of Contents

List of tables

List of figures

2006–2008 © Copyright lies with the respective authors and their institutions.

1. Introduction

Multilinguality in ontologies is nowadays demanded by institutions worldwide with a huge number of resources available in different languages. Within the NeOn project, both use cases (FAO and the Spanish Pharmaceutical Industry) have expressed the need for semantically structuring the information they have in different natural languages.

The FAO, as an international organization with five official languages -English, French, Spanish, Arabic and Chinese- has a great variety of heterogeneous and multilingual linguistic resources with different levels of granularity, among them:

- Glossaries: FAOTERM[1], the Fisheries Glossary[2], the Aquaculture Glossary[3], and Globefish[4], etc (in English, French, Spanish, Arabic and Chinese)
- Thesauri: AGROVOC[5] (currently in 17 languages), ASFA[6] (in English, French and Spanish), for example.
- Databases: AGRIS[7], FIGIS[8], etc.
- Non-FAO resources usually accessed by FAO personnel: Fishbase[9], Onefish[10], Ocean Atlas[11], among others

All FAO resources are multilingual resources in, at least, three of the institution official languages. Depending on the type of linguistic resource we are referring to, the sort and quantity of information will vary. For example, by the time we are writing this document, AGROVOC contains related terms in seventeen languages as well as scope notes about some use particularities of those terms, but no definitions. The FAOTERM glossary, however, contains the searched term and, usually, its definition in the five official languages of the FAO. (For a more detailed description of these two resources see D2.4.1, sections 5 and 8). We should also note that the amount of information contained for each language in FAO resources also varies from one language to the other. Some resources are more complete in the traditional languages managed by the FAO, English, Spanish and French, and show important gaps in the rest of languages. One of the main problems also appears when users in different locations edit (update and remove) asynchronously specific modules of a certain resource in a specific language, and changes are not propagated to the rest of languages or resources containing similar information. Last but not least, the issue of cultural discrepancies becomes evident in case of concepts that exist in one culture but not in others, as for example, some varieties of Asian rice unknown for most European cultures.

In the case of the Spanish Pharmaceutical Industry, multilinguality is also necessary because of the different languages that are spoken in some Spanish regions, and which have the same official

[1] http://www.fao.org/faoterm/

[2] http://www.fao.org/fi/glossary/default.asp

[3] http://www.fao.org/fi/glossary/aquaculture

[4] http://www.globefish.org/index.php?id=481&easysitestatid=-277727447Globefish

[5] http://www.fao.org/aims/ag_intro.htm

[6] http://www4.fao.org/asfa/asfa_es.htm

[7] http://www.fao.org/AGRIS/

[8] http://www.fao.org/fishery

[9] http://www.fishbase.org/home.htm

[10] http://www.onefish.org/global/index.jsp

[11] http://www.oceansatlas.org/Oceans Atlas

2006–2008 © Copyright lies with the respective authors and their institutions.

status as the Spanish language in those regions. These languages are Catalan, Basque and Galician. Not only may available data and sources of information be written in one of those languages, but also information results have to be customized for the different regional administrations in the language selected by the user. In this case, ontologies are also required to support multilinguality. Therefore, the specific requirements regarding multilinguality expressed in different NeOn WPs (WP1, WP6, WP7, WP8) were analyzed in detail in D2.4.1, and considered in the design of the NeOn Multilingual Ontology Meta-model.

1.1 Deliverable Main Goals and Contributions

The scenarios presented in section 1 attempt to be illustrative examples of the need for semantically organizing huge amounts of multilingual information that some organizations have. When providing ontologies with multilingual data, one of the activities identified in the NeOn ontology network development process is the Ontology Localization Activity, that consists in "adapting an ontology to a concrete language and culture community", as defined in [28,29]. The Ontology Localization Activity results in heterogeneous multilingual ontology meta-models depending on the ontology layers implied in the localizing activity. Therefore, the **first goal** of this deliverable is to describe the different ways we have identified for representing multilinguality in ontologies depending on the ontology layers implied in the Ontology Localization Activity.

Once we have analyzed the different multilingual modelling modalities, our **second aim** is to describe the option we have selected and the motivation underlying it, which was already outlined in the first version of this deliverable (D2.4.1- section 14.2). This option relies on the association of the ontology meta-model to a linguistic model called *Linguistic Information Repository* (LIR), whose second version will be fully described in this deliverable, highlighting the differences with the previous version.

The **third contribution** of this deliverable is related with some initial tests that were conducted at the FAO in order to confirm that the LIR meets FAO needs. For this aim, the LIR was evaluated against two lexical resources of the FAO: the AGROVOC Concept Server and the FAOTERM glossary. Then, we discuss some of the improvements that can be made to the LIR in future versions.

Finally, the **fourth contribution** of this document regards the technical support we aim at giving to the LIR, so that the multilingual information that is to be associated with ontologies by instantiating the LIR, can be introduced in the model in an automatic way. This will be achieved by means of the second version of the *LabelTranslator* plug-in and the new translation algorithm it uses. Both plug-in and algorithm have been enhanced in different ways to provide an automatic localization of the information contained in the LIR, namely, an automatic translation of the labels associated to the ontology elements from one source language to one or various target languages, and an automatic association of additional lexical and terminological information related to those labels by accessing external lexical resources.

1.2 Deliverable Structure

This deliverable is structured as follows:

- Section 2 deals with the most representative approaches for modeling multilinguality in ontologies, which essentially depend on the ontology layers that undergo localization [17]. Two of these approaches were briefly described in D2.4.1. In the current version of D2.4.2, a more comprehensive analysis is performed, in which a systematic description of the three options we have identified for representing multilinguality in ontologies is provided, as well as a brief state-of-the-art on the most relevant models that follow these approaches.

- Section 3 presents the second version of the LIR [17], and describes in detail the classes and properties that compose it. In this section, a subsection is devoted to the main changes undergone by the LIR with respect to the previous version reported in [24,18] and the motivation behind those changes. Finally, the OWL code of the LIR is attached by way of annex (see Annex 1).
- Section 4 describes some of the experiments conducted at the FAO in order to evaluate the suitability of the LIR model.
- Section 5 discusses the benefits of integrating the LIR in a network of ontologies that represent and structure linguistic and terminological data, and suggests further refinements to the LIR in future versions.
- Section 6 presents the second version of the LabelTranslator plug-in [6,7]. This section is subdivided into 5 different subsections. The first one is devoted to the limitations of the first version of the LabelTranslator plug-in. The second section describes the main innovations of the second version of the plug-in. The third section deals with the *GUI Component* of the LabelTranslator. The forth is focused on a detailed description of what has been called the *Ontology Localization Component*, i.e. the new translation algorithm that allows the automatic localization of ontology labels from a source language into a target language. Finally, the fifth section describes the *Repository Component* of the LabelTranslator NeOn plug-in and the synchronization process of linguistic and conceptual knowledge in NeOn ontologies.

2. Approaches for modelling multilinguality in ontologies

Ontologies, as any other systems for representing knowledge, make use of "thoughts of reference", also known as concepts, to refer to the real world. Comparing ontologies to linguistic systems, we may state that both have three main components: *signs* or symbols used to designate *concepts* or thoughts of mind, which refer to *phenomena* in the real world. In linguistic semantics this idea was represented as a triangle by Ogden and Richards in [20] (cf. Figure 1). A few years later, Morris [19] in his approach to semiotics made a similar distinction by dividing the sign into: *sign vehicle* (syntax), *designatum* (semantic) and *interpreter* (pragmatic), stating in this way that the understanding of the world was dependant of the viewpoint of the interpreter. A comparable division was reported by [14], when talking about heterogeneity of database systems. Heterogeneity could be **semantic**, having to do with content or content organization, or **non-semantic**, related to the structure or syntax used to express content.

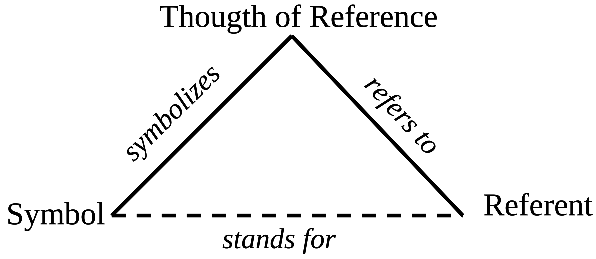


Figure 1. Ogden and Richards' triangle

Coming back to the ontology field, multiple authors have tackled this topic and we can basically distinguish 6 layers in any ontology, as summarized in [1], namely (cf. Figure 2):

I. Lexical layer: characters and symbols that make up the syntax (ASCII encoding, UNICODE, etc.)

2006–2008 © Copyright lies with the respective authors and their institutions.

II. Syntactic layer: structure of characters and symbols, i.e., the grammar. It embraces different representation languages (e.g. RDF(S), OWL, etc.)

III. Representation paradigm layer: paradigm followed in the representation of the ontology (frames, semantic networks, DL, etc.) that allows a certain way of expressing and structuring knowledge

IV. Terminological layer: terms or labels selected to name ontology elements

V. Conceptual layer: related to conceptualization decisions, such as granularity, expressiveness, perspective, etc.

VI. Pragmatic layer: final layout of the model according to user's needs

According to this, we may state that *terminological*, *conceptual*, and *pragmatic* layers are the ones involved in the Ontology Localization Activity (OLA), as illustrated in Figure 2. Terminology plays a decisive role in the localization activity since it is related with the names we give to the different ontology elements. As a result of this activity, ontology labels will be expressed in more than one natural language. Regarding the conceptual layer, certain ontologies may require adaptation of their conceptual structure in order to fit in the *thoughts of reference* of a specific linguistic and cultural community.

Finally, with regard to the pragmatic layer, the way in which multilingual information is presented to the user may also influence the acceptance and effectiveness of the model by the end user. The rest of layers -lexical, syntactic and representation paradigm layers- will not be affected by the localizing activity, and their design options will be normally independent of the inclusion of multilingual information.

Although the number of multilingual ontologies is still quite small in comparison with the ontologies available in the web (cf. OntoSelect[12] or Watson[13]), some models for representing multilingual data have been developed in order to respond to different needs for linguistic information. By way of summary, we may state that there are three main ways of modelling multilinguality in ontologies, depending on the extent to which each layer is involved in the localization activity (each modelling modality will be explained in more detain in the following sections):

- Including multilingual data in the *ontology meta-model*[14]: this implies localization at the terminological layer since the ontology conceptualization remains unmodified
- Combining the *ontology meta-model* with a *mapping model*: this allows localization at the conceptual layer since conceptualizations in different languages are mapped to each other
- Associating the *ontology meta-model* to a *multilingual linguistic model:* localization is performed at the terminological layer, although conceptual layer adaptations are also allowed

The pragmatic layer will not be considered at this stage.

The first and third modelling modalities identified here were briefly explained in D2.4.1. In this version, we wanted to offer a more exhaustive analysis, as well as a brief state-of-the-art on some of the most relevant models that follow the identified approaches. We have also provided a description of the advantages and disadvantages of each modelling option, which support our decision of following the third modelling modality, which implies the association of the ontology meta-model to a multilingual linguistic model (see section 2.3).

---

[12] http://olp.dfki.de/ontoselect/

[13] http://watson.kmi.open.ac.uk/WatsonWUI/

[14] A meta-model is the ontology for a model, or, in other words, a meta-model is a specification/model of a specification/modeling language

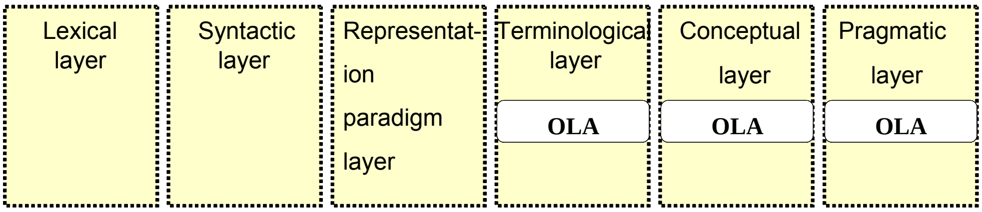


Figure 2. Ontology heterogeneity layers affected by the OLA (Ontology Localization Activity)

2.1 Including multilingual data in the *ontology meta-model*: localization at the terminological layer

Including multilingual data in the *ontology meta-model* is the most widespread modelling modality within the ontological community nowadays. It consists of making use of the `rdfs:label` and `rdfs:comment` properties to define labels and descriptions in natural language for classes (see RDF(S)[15]). This system allows localizing the ontology model at the terminological layer, as labels for ontology classes can be expressed in various natural languages (see Figure 3). In a similar way, the Simple Knowledge Organization System (SKOS[16]) data model for semantically structuring thesauri, taxonomies, classification schemes, etc., allows labelling concepts with multilingual strings, and even establishing relations between linguistics labels, as to determine which is the preferred label (`skos:prefLabel`) and the alternative one (`skos:altLabel`), which was not possible with RDF(S) properties. Apart from relations between labels, SKOS provides fuzzy semantic relations between concepts as `skos:narrower`, `skos:broader` and `skos:related`.

Disadvantages: Limitation of the amount of linguistic information that can be included in the ontology limited to strings without information about senses in their respective languages, nor provenance of the information, which makes concept localization to different natural languages quite difficult. Full synonym relation or 100% equivalence is assumed among labels in different languages associated with the same class, since it is not possible to establish more specific relations between the linguistic elements associated with ontology classes. Furthermore, this representation way does not allow performance of complex operations with linguistic elements, since no lexical relations exist among them.

Advantages: Labels can be integrated in the ontology in as many languages as wished by the user. This model provides a suitable representation form for highly specialized domain ontologies, e.g. in the technical field, since that sort of knowledge is shared among different linguistic and cultural communities and the equivalence relation among labels in different languages is usually adequate.

[15] www.w3.org/TR/rdf-schema/

[16] http://www.w3.org/2004/02/skos/specs

2006–2008 © Copyright lies with the respective authors and their institutions.

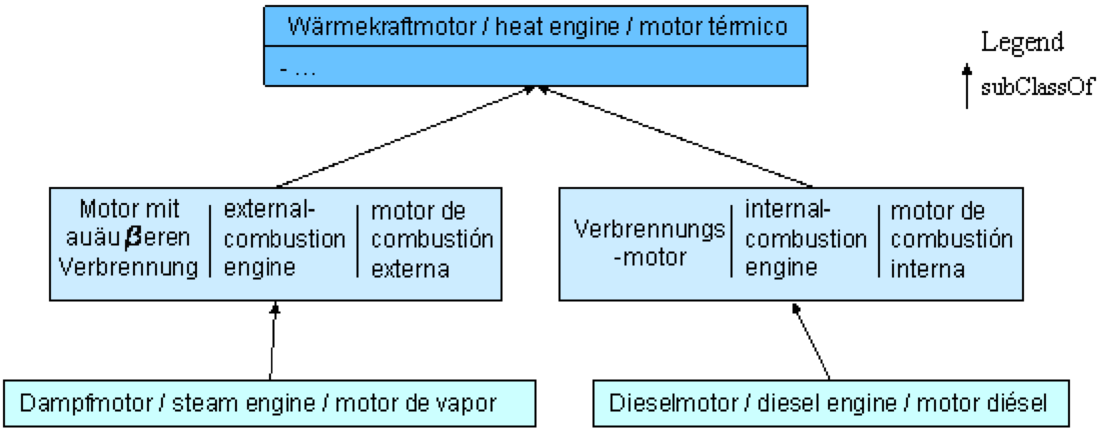


Figure 3. Multilingual Data Included in Ontology Meta-model

2.2 Combining the *ontology meta-model* with a *mapping model*: localization at the conceptual layer

According to this approach there are various modelling ways depending on the mapping arity and the graph form, as for example:

- Binary mappings in an orthogonal graph. In this case, each monolingual ontology organizes knowledge of a certain culture, and is mapped to the rest of ontologies in a pair-wise fashion. See Figure 4.
- Binary mappings in a radial graph. In this option monolingual ontologies are mapped to each other through an interlingua consisting of a set of common concepts that allow establishing equivalences (see Figure 5). The most representative model following this approach is the one adopted by the EuroWordNet (EWN) multilingual lexicon.

This modelling option implies localization at the conceptual layer. Knowledge is organized according to the structures of a specific cultural community of users.

Disadvantages: A huge effort is needed in order to conceptualize the same domain in different natural languages. Three types of expertise are required: domain expertise, linguistic expertise, and ontology engineering expertise. The establishment of alignments between conceptualizations in different languages is by no means trivial. Although concept equivalents among localized ontologies are reliable and reflect cultural differences, the quantity of linguistic information embedded in the ontology is often limited to labels and definitions associated with ontology classes.

Advantages: This option allows maintaining conceptualizations in each language, and, in this sense, it is suitable for ontologies modelling knowledge highly dependent of a certain culture, such as the judiciary.

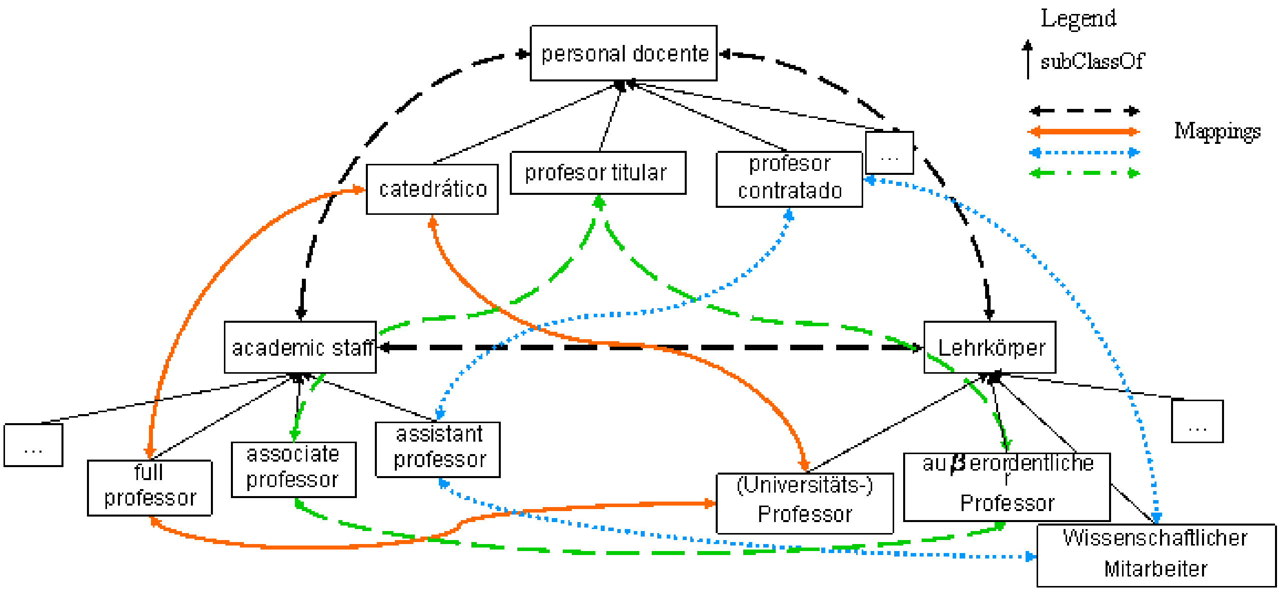


Figure 4. Ontology Meta-model combined with a Mappings model (Binary Mappings in an Orthogonal Graph)

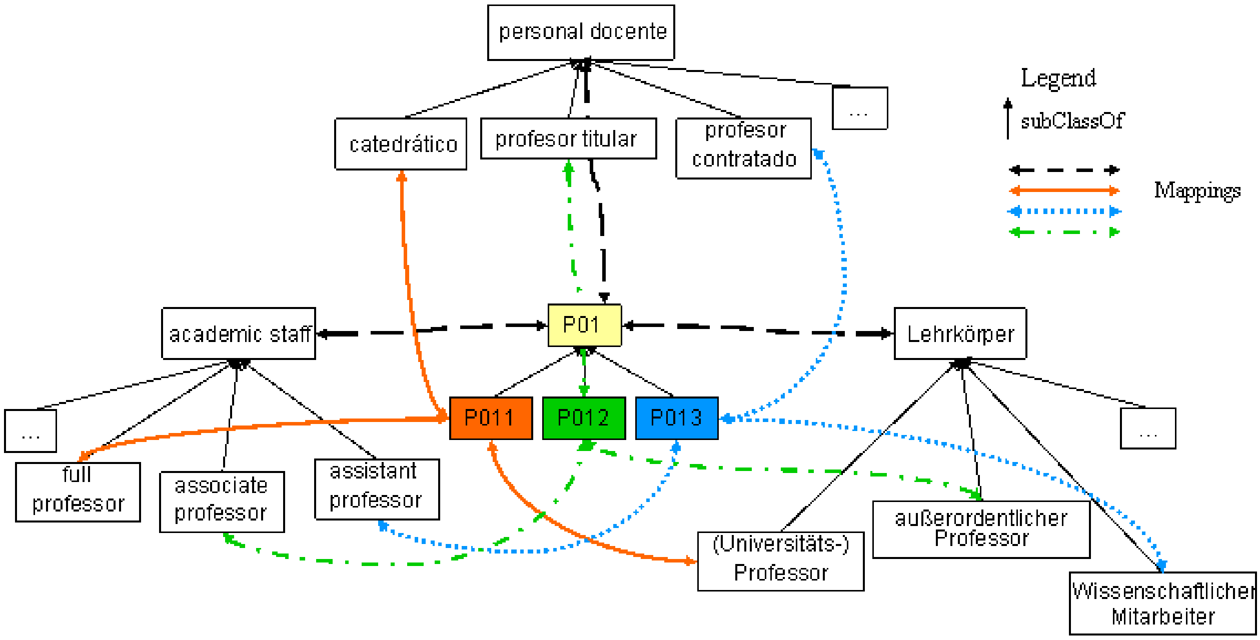


Figure 5. Ontology Meta-model combined with a Mappings model (Binary Mappings in a Radial Graph)

2.3 Associating the *ontology meta-model* with a *multilingual linguistic model*: localization at the terminological and conceptual layers

In this modelling option, the elements of the ontology have links to multilingual data stored outside the ontology. The model for representing and organizing the linguistic information can be a data base (as in GENOMA-KB[17] or OncoTerm[18]), an ontology (as in the case of LingInfo [3], LexOnto [5], or the AGROVOC Concept Server [15]), etc. In this approach, conceptual and terminological layers are kept separate, and the localizing activity is mainly carried out at the terminological layer. However, the ontology conceptualization layer can also undergo modifications, as the creation of language specific ontology modules, in order to meet localization needs.

Disadvantages: Since there is just one conceptualization, some language specificities could be lost, unless captured in language specific ontology modules, i.e., at the conceptual layer, or in the linguistic model, i.e., at the terminological layer.

Advantages: This representation form allows the inclusion of as much linguistic information as wished, as well as the possibility of establishing links among the linguistic elements within one language or across languages. In this sense, nuances or differences between languages can also be reported and even formalized in the terminological layer, in order to avoid the 100% equivalence correspondence among the different names of ontology elements. Relevant information as, e.g., the provenance of the linguistic elements, can also be included. This system also allows linguists or domain experts without ontology development expertise access to the terminological layer in a distributed environment.

[17] http://genoma.iula.upf.edu:8080/genoma

[18] http://www.ugr.es/~oncoterm/alpha-index.html

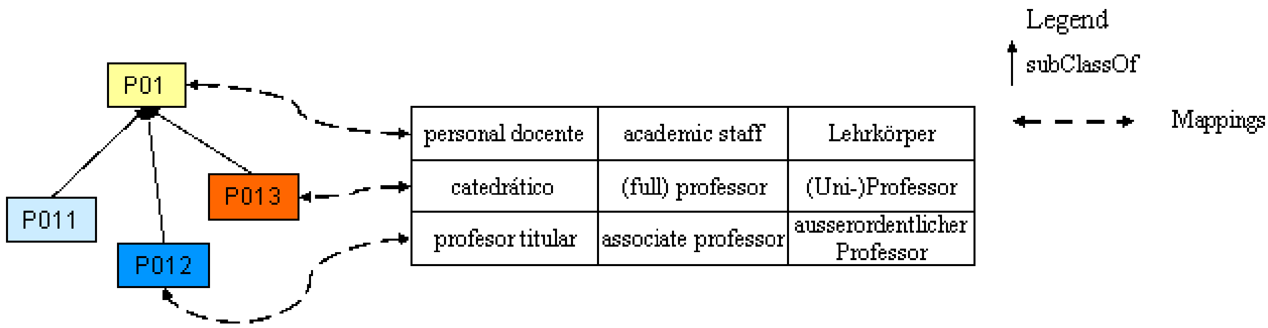


Figure 6. Ontology Meta-model associated with a Multilingual Linguistic Model

3. Second Version of the Linguistic Information Repository (LIR)

From the different modelling options presented above, the LIR model follows the option that consists in associating a multilingual linguistic model to the ontology meta-model (explained in section 2.3). In this way, the LIR allows the representation of the necessary linguistic information for localizing ontologies at its terminological level. The LIR has been conceived as a hub to interconnect various standard descriptions for linguistic and terminological knowledge, on the one hand, and ontological concepts, on the other. As the most important standard initiatives considered in the development of the LIR, we have analyzed two ISO standards, the Terminological Markup Framework (TMF) and the Lexical Markup Framework (LMF), described in the previous version of this deliverable (D2.4.1 – section 2 [18]). However, the hub of classes that compose the LIR is not intended to be the superset of all concepts in all standards, but to contain enough information in itself in order to allow concept label selection and translation. Fine-grained lexical knowledge, such as, morphological decomposition and syntactic complementation, are available from LMF and can be obtained by navigating those models. The LIR also serves the objective of integrating and aggregating multilingual information contained in heterogeneous and distributed lexical resources by guaranteeing a homogeneous access to the information.

Finally, it should be mentioned that the rationale underlying LIR is not to design a lexicon for different natural languages and then establish links to ontology concepts, but to associate multilingual linguistic knowledge to the conceptual knowledge represented by the ontology. What the LIR does is to associate *word senses* –as defined by Hirst [11]- in different languages to ontology concepts, although word senses and concepts can not be assumed to overlap. The reason for not completely overlapping is that word senses are tightly related to the particular vision of a language and its culture, whereas ontology concepts try to capture objects of the real world, and are defined and organized according to expert criteria agreed by consensus. The LIR goes more in the line of what Pustejovsky [25] defined as *Sense Enumeration Lexicon*, in which a unique sense is associated with a word string. From a translational viewpoint, this is a suitable approach for obtaining a multilingual system.

3.1 Main changes to the first version of the LIR

The following changes to the first version of the model presented in D2.4.1 have been implemented after a period of discussions between the involved partners. In order to better understand them, we have included a figure representing the previous LIR model [24] (cf. **Error! Reference source not found.**) and the current version of the LIR model [17] (cf. Figure 8).

2006–2008 © Copyright lies with the respective authors and their institutions.

The main changes to the LIR model have been motivated by three different types of reasons: pragmatical reasons, linguistic needs expressed by NeOn Use Cases, and ontology consistency problems. These changes are described in the following:

a) **Pragmatical reasons**. The LIR model takes into account useful possibilities for users that may want to visualize or edit certain linguistic information associated to ontology elements in just one language, or in several natural languages at the same time.

1. In this sense, the first big change has to do with the addition of the classes **Language** and **LanguageCode** from *FAO's languagecode ontology* (http://www.fao.org/aims/aos/languagecode.owl) (see Figure 8). These classes represent natural languages as objects, and the ISO 639-1 and 639-2 codes from *FAO's languagecode ontology*. **ISO639-1** and **ISO639-2** are subclasses of LanguageCode. Language and LanguageCode are related through the object property *hasLanguageCode.*

The integration of these classes from the *FAO's languagecode ontology* into the network allows structured access to multilingual information, and, if necessary, properties of the languages are also covered. The Language class is connected to various LIR classes by means of the object property *belongsToLanguage* and its inverse *hasLinguisticExpression*, indicating provenance and cultural links of orthographic and semantic LIR classes with one or more languages.

This property and its inverse replace the *xml:lang* attribute of the class Sense in the first version of the LIR (see **Error! Reference source not found.**). However, *xml:lang* is maintained in most classes of the second version of the LIR as an optional attribute to reflect the language code from ISO639-2 associated with the range of the *belongsToLanguage* object property in the *FAO's languagecode ontology*. This redundancy is under review until all querying issues are resolved.

One of the most important benefits of the inclusion of these classes is that users can restrict linguistic information retrieval to certain classes associated with a specific natural language. Let us imagine that a terminologist working for the FAO only wants to visualize those Lexicalizations and their related Notes that belong to the French language. This would be possible thanks to these new classes.

2. The second change has to do with the removal of the **PartOfSpeech** class and its data type property *category* (see **Error! Reference source not found.**), and the inclusion *partOfSpeech* as a data type property of the class **LexicalEntry** in the second version of the LIR model for reasons of intuitive clarity.

3. The addition of the data type property *sourceType* (domain: Source) with different subproperties (*nameSpaceIdentifier, bibliographicReference, sourceIdentifier, text*).

Some of these values have been adopted from ISO 26220, and a free text field for the description of the source has been added. This allows to better account for the source of provenance of the linguistic elements captured in the LIR.

b) **Use Case needs**. NeOn Use Cases have explicitly expressed the need for certain classes or properties which had not been taken into account in the first version of the LIR model (**Error! Reference source not found.**), but that turned out to be linguistically relevant, also for other eventual NeOn external users.

1. The addition of the object property *hasDialectalVariant* (domain and range: Lexicalization), and the corresponding *DialectalVariant* Boolean termType attribute, in order to capture dialectal variation within one language. Further more, the new data property *belongsToDialect* allows the encoding of the name of the dialect the Lexicalization belongs to.

This addition of this property was motivated by use case needs. Both NeOn use cases deemed this property as highly relevant in order to express regional or dialectal variants they came across in their domains.

2. The addition of the data type properties *scientificName* and *commonName* as Boolean attributes of Lexicalization, in addition to the relations *hasScientificName* and *hasCommonName* (domain and range: LexicalEntry), in order to cover cases where e.g. the only Lexicalization is a scientific name, and no common name variant exists. See the discussion in section 3.3, point 8.

3. The addition of the object property *hasAntonym* (domain and range: LexicalEntry) to capture that sort of lexical information so relevant in most lexicons (e.g. WordNet).

c) **Ontology consistency problems**. Some of the design decisions taken in the first version of the LIR were considered to cause some consistency problems, and we decided to change them to avoid such problems.

1. The reorganization of some Boolean data type properties (*fullForm*, *shortForm*, *abbreviation*, *acronym*, *logicalExpression*, *equation*, *formula*, *symbol*) as subproperties of the TMF-derived *termType* data type property.

2006–2008 © Copyright lies with the respective authors and their institutions.

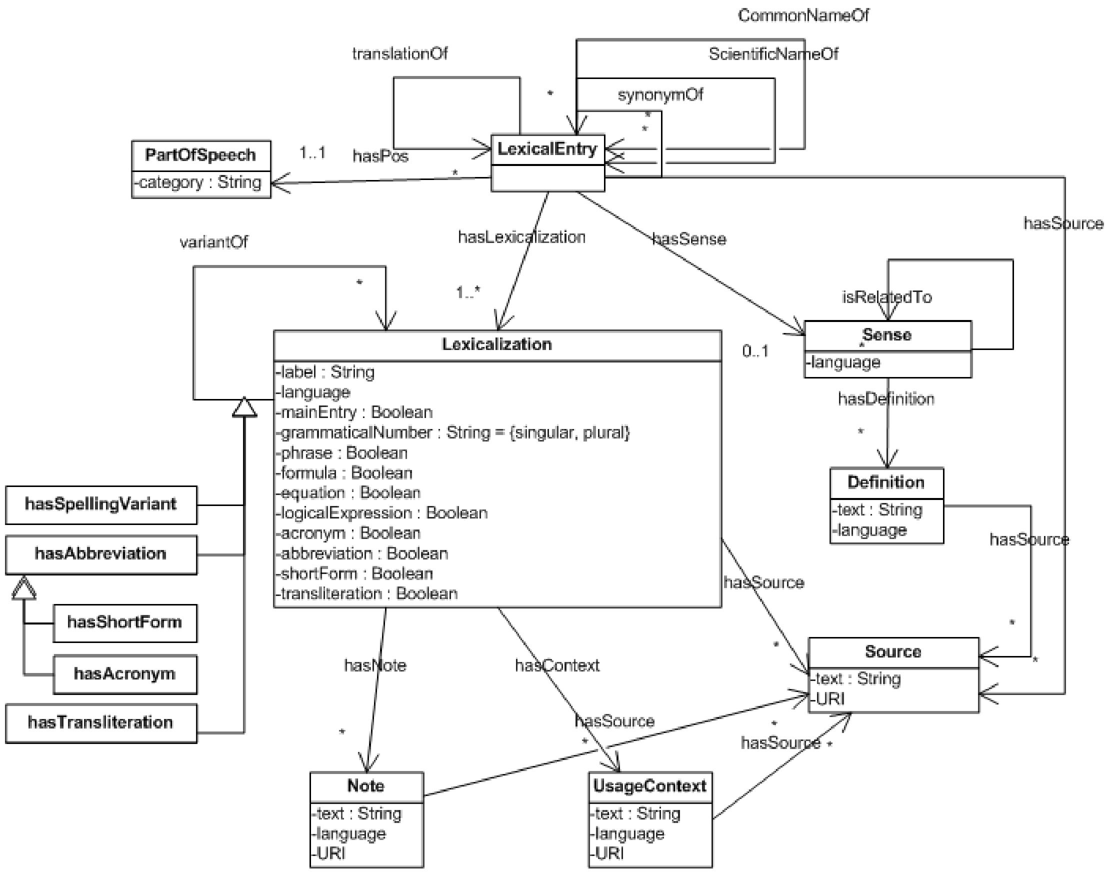


Figure 7. First version of the LIR model [24]

In the next two sections (3.3 and 3.4), our aim is to give a detailed description of the nine classes that compose the LIR model and their corresponding data type properties. For this end, we provide definitions for classes and data type properties, as well as their correspondence with equivalent classes in the linguistic and terminological ISO standard initiatives taken into account for the design of the LIR (LMF, TMF and TBX), as indicated in each case.

This list adheres to the following stylistic and typographic conventions:

- Classes are rendered in bold face when heading a section
- Properties are in italics
- Object properties are also called relations
- Data properties are also called attributes
- “Lexicalization” is synonymous with “word form”
- “Sense” is synonymous with “meaning”

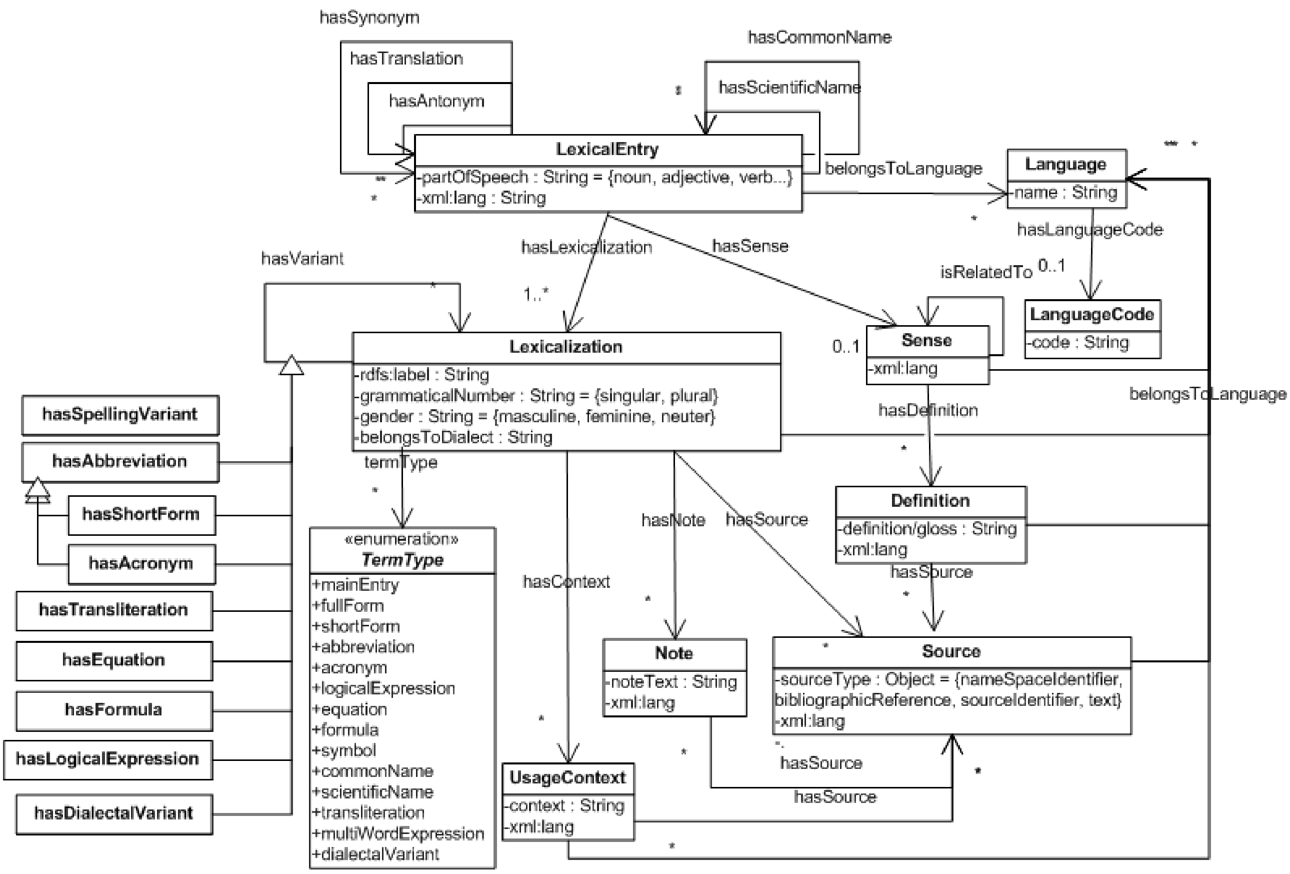


Figure 8. Second version of the LIR model [17]

3.2 Description of the LIR Classes

In the following we describe the nine classes that compose the LIR.

1. LexicalEntry: a lexeme, which is a unit of form and meaning.

A lexeme is an ordered collection of related word forms, having the same lexical meaning[19] [26].

Please note that the meaning shared by the word forms is lexical, not grammatical. In other words, meaning differences between e.g. singular/plural are not covered by lexical meaning.

The LexicalEntry class manages the link between the classes Sense and Lexicalization. It is an abstract class, of which each instance is a combination of a set of lexicalizations and zero or one sense.

The LexicalEntry class has the following attributes:

- partOfSpeech: the grammatical class of the LexicalEntry.

  Traditionally, members of the set of word forms incorporated into a particular lexeme are selected on the basis of part of speech, inflectional behaviour and meaning. Within the LIR, this means that lexemes are pre-filtered by the major syntactic class by means of the *partOfSpeech* attribute. This corresponds with the encoding of part of speech in LMF. By

[19] See also Wikipedia: http://en.wikipedia.org/wiki/Lexeme

2006–2008 © Copyright lies with the respective authors and their institutions.

doing so, the repetition of *partOfSpeech* for all **Lexicalization** instances is avoided, since Lexicalizations are deemed to belong to the same major part of speech. Synonymy relations across major part of speech boundaries will need to be implemented at the **LexicalEntry** level.

- *Xml:lang*: optional attribute reflecting the language code from ISO639-2[20] associated with the range of the *belongsToLanguage* object property (see previous section) in the *FAO's languagecode ontology*.

2. Sense: a language-specific unit of intentional lexical semantic description.

Optional attribute:

- *xml:lang*: reflects the language code from ISO639-2[21] associated with the range of the *belongsToLanguage* object property (see previous section) in the *FAO's languagecode ontology*. This allows us to model idiosyncratic differences between language specific meanings.
  Within LIR, there are two possible ways to model language specificity:

  a) Based on the principled viewpoint that lexical entries by default express language specific notions, Sense is necessarily considered language specific as well. This is assumed in e.g. the EuroWordNet model [32].

  Besides, two lexical entries in different languages are associated with different senses. If the lexical entries mean the same, we link them with the *hasSynonym* relation (see section 3.3, point 2).

  Other types of equivalence relations between lexical entries, such as *equivalence near-synonymy* and *equivalence hypernymy*, can be modelled in LIR by postulating sub-relations of the *isRelatedTo* relation between senses.

  b) Terminological entries in e.g. TMF[22] and TBX[23] define one sense for a multilingual set of terms. The assumption behind this is that terms - as opposed to lexical items in general - have a very precisely defined meaning within a domain, which is shared across language boundaries.

  In order to model this terminological case, we can either apply a), or link each LexicalEntry to one and the same Sense, i.e. the meaning of the terminological entry. This would represent a variant of an interlingua approach assuming full equivalence between language-specific lexical entries.

The LIR is capable of modelling both options. A choice needs to be made for each use case.

The translational or conceptual equivalence between lexical entries is expressed by the relations *hasTranslation* and *hasSynonym* (see section 3.3, points 2 and 3).

3. Lexicalization: a word form

[20] http://www.loc.gov/standards/iso639-2/php/English_list.php

[21] http://www.loc.gov/standards/iso639-2/php/English_list.php

[22] http://www.loria.fr/projets/TMF/

[23] http://www.lisa.org/standards/tbx/

This class corresponds with the LMF[24] Form Representation class. The choice of this data category means that the lexicalizations of concepts are deemed word forms rather than lemmas or citation forms, and therefore also include inflected forms, such as plurals.

The notion of Lemma as the canonical form (citation form) representing the set of related word forms such as inflections, is equivalent to Lexicalization with attribute *mainEntry* (see below) set to true.

The class Lexicalization has the following attributes:

- *rdfs:label*: string representing the word form.
- *xml:lang*: (optional) language code from ISO639-2[25] associated with the range of the *belongsToLanguage:* object property (see previous section) in the *FAO's languagecode ontology*.
- *grammaticalNumber:* captures the morpho-syntactic features of the lexicalization, and can take the following values: "singular", "plural" and "other".
- *gender:* captures grammatical and inflectional features of the lexicalization, and can take the following values: "masculine", "feminine", and "neuter".
- *belongsToDialect*: the dialect name to which the Lexicalization belongs.

Further, it contains a set of descriptions for term types taken from TMF[26] and TBX-Lite[27], split up into:

a) Term type attributes represented as a set of Boolean attributes or values of the *termType* attribute itself that describe a number of term types (sub-properties of the data property *termType*):

- *mainEntry*: the concept designation that has been chosen to head a terminological record (ISO12620: section 02.01.01). By default, this field contains the lemma or citation form.
- *formula*: figures, symbols or the like used to express a concept briefly, such as a mathematical or chemical formula (ISO12620: section 02.01.14).
- *equation*: an expression used to represent a concept based on the statement that two mathematical expressions are, for instance, equal as identified by the equal sign (=), or assigned to one another by a similar sign (ISO12620: section 02.01.15).
- *symbol:* a designation of a concept by letters, numerals, pictograms or any combination thereof (ISO12620: section 02.01.13).
- *logicalExpression*: an expression used to represent a concept based on mathematical or logical relations, such as statements of inequality, set relationships, Boolean operations, and the like (ISO12620: section 02.01.16).
- *scientificName:* a term that is part of an international scientific nomenclature as adopted by an appropriate scientific body (ISO12620: section 02.01.04).
- *commonName:* a synonym for an international scientific term that is used in general discourse in a given language (ISO12620: section 02.01.05).
- *fullForm:* the complete representation of a term for which there is an abbreviated form (ISO12620: section 02.01.07).

[24] http://www.lexicalmarkupframework.org/

[25] http://www.loc.gov/standards/iso639-2/php/English_list.php

[26] http://www.ttt.org/oscar/xlt/webtutorial/datcats02.htm

[27] http://www.lisa.org/standards/tbx/

2006–2008 © Copyright lies with the respective authors and their institutions.

- *acronym:* an abbreviated form of a term made up of letters from the full form of a multiword term strung together into a sequence pronounced only syllabically (ISO12620: section 02.01.08.04).
- *shortForm*: an abbreviated form that includes fewer words than the full form, e.g. "Intergovernmental Group of Twenty-four on International Monetary Affairs" vs. "Group of Twenty-four" (ISO12620: section 02.01.08.02).
- *abbreviation*: a term resulting from the omission of any part of the full term while designating the same concept, e.g. adjective vs. adj. (ISO12620: section 02.01.08).
- *transliteration*: a form of a term resulting from an operation whereby the characters of an alphabetic writing system are represented by characters from another alphabetic writing system (ISO12620: section 02.01.10).
- *multiWordExpression*: this attribute is equivalent to ISO12620 *Phrase*, defined as a phraseological unit containing any group of two or more words that are frequently expressed together and that comprise more than one concept. The individual words in a phrase usually function in more than one grammatical category (part of speech) within the syntax of a sentence, e.g. "work offline" (ISO12620: section 02.01.18).
- *dialectalVariant*: this attribute indicates whether a word form originates from a dialect.

b) A number of relations between Lexicalization classes expressed by the object property *hasVariant* and its following sub-properties.

In TMF and TBX, these term types are represented as attributes rather than relations. However, representing them as relations rather than as Boolean attributes ensures the proper link between unique source and target lexicalizations where term type attributes allow multiple derivations of relations.

The reason for using both a set of Boolean attributes and a set of relations is that relations cannot always be deduced from a set of attributes. For instance, if two lexicalizations are associated with a LexicalEntry, one of them as a full form, and one as an abbreviation, then it is impossible to determine with certainty if, on the basis of Boolean attributes, the full form lexicalization is related to the abbreviation.

Also, if a LexicalEntry contains two full form lexicalizations and one acronym, it is impossible to determine which full form is in the domain of the *hasAcronym* object property on the basis of attributes alone. For instance, the WordNet synset *{J, Joule, watt second}* (unit of electrical energy) contains three Lexicalizations, of which two are full forms, and one is an acronym. Using attributes alone will not enable the user to establish the right *hasAcronym* relation between any pair wise combination of these Lexicalizations.

Conversely, in cases where a LexicalEntry occurs in isolation, it is impossible to determine the term type of the Lexicalization on the basis of relations, because there are not any available. For instance, when there is only one LexicalEntry containing a scientific name, the relation *hasScientificName*, which holds between LexicalEntries (see below), cannot be used to characterize the Lexicalizations contained by the LexicalEntry as scientific name. In order to be able to do this, this *hasScientificName* relation needs at least a pair of lexical entries one of which contains a lexicalization with a *scientificName* attribute value "true", whereas the other needs a Lexicalization with the *commonName* attribute value "true". The attribute *ScientificName* is necessary to characterize the lexicalization from this isolated LexicalEntry as a scientific name.

- *hasSpellingVariant* (inverse: *isSpellingVariantOf*)
- *hasAbbreviation* (inverse: *isAbbreviationOf)*
- *hasAcronym* (inverse: *isAcronymOf*)

- *hasShortForm* (inverse: *isShortFormOf)*
- *hasTransliteration* (inverse:*isTransliterationOf*)

hasAcronym and *hasShortForm* are subtypes of *hasAbbreviation*. Although both have been officially disallowed in TMF, and the use of the more general attribute *Abbreviation* is prescribed, FAO requires these data categories.

4. Language: this is the language concept imported from *FAO's languagecode ontology* (http://www.fao.org/aims/aos/languagecode.owl). This ontology contains multilingual language names and ISO639 codes. It is linked to various LIR classes through the object property *belongsToLanguage* and its inverse *hasLinguisticExpression* (see below).

5. LanguageCode: the ISO 639-1 and 639-2 codes are standard labels for languages[28], which have been incorporated into *FAO's languagecode ontology*.

ISO 639-1 is the alpha-2 code (codes composed of 2 letters of the basic Latin alphabet). Multiple codes for the same language are to be considered synonyms.

ISO 639-2 is the alpha-3 code (codes composed of 3 letters of the basic Latin alphabet).

Both **ISO639-1** and **ISO639-2** are subclasses of LanguageCode.

Language and LanguageCode are related through the object property *hasLanguageCode* (see below).

6. Definition: a statement that describes a concept and permits its differentiation from other concepts within a system of concepts. (ISO12620: section 05.01)

The Definition class has the following attributes:

1. *definition/gloss*: string.

2. *xml:lang*: optional attribute to indicate the language in which the definition is written. It reflects the language code from ISO639-2[29] associated with the range of the *belongsToLanguage* object property (see previous section) in the *FAO's languagecode ontology*.

7. Source: the provenance of the linguistic/terminological information.

The class Source contains the following data properties:

1. *sourceType*, which itself has the following sub-properties:
   - *nameSpaceIdentifier*: URL/URI (see ISO12620: section 10.21).
   - *bibliographicReference*: a complete citation of the bibliographic information pertaining to a document or other resource (see ISO12620: section 10.19).
   - *sourceIdentifier*: the code assigned to a document in a terminological collection and used as both the identifier for a bibliographic entry and as a pointer in individual term entries to reference the bibliographic entry identified with this code (see ISO12620: section 10.20).

[28] http://en.wikipedia.org/wiki/ISO_639

[29] http://www.loc.gov/standards/iso639-2/php/English_list.php

2006–2008 © Copyright lies with the respective authors and their institutions.

- *text*: e.g. a textual description of the resource, or maybe a unique key into the resource specific information structure (for instance, in the case of a dictionary, the composite key lemma, part of speech and sense number).

2. *xml:lang*: optional attribute to reflect the language code from ISO639-2[30] associated with the range of the *belongsToLanguage* object property (see previous section) in the *FAO 's languagecode ontology.*

8. UsageContext: a text or part of a text in which a term occurs (ISO12620: section 05.03).

TBX describes this class as follows:

"Context sentences serve the following purposes:

- They prove that the term actually exists in real language.
- They can shed light on the meaning of the term.
- They can provide additional "encyclopaedic" information about the term that is not in the definition (the who, why, when, where, how).
- They can illustrate how the term is used in discourse (collocations, register, etc.). For instance, a context sentence could alert the translator that the term is colloquial.
- They can provide grammatical information (such as gender), stylistic clues (such as hyphenation or capitalization) as well as alternate forms (abbreviations and so forth).
- The requirement to include a context sentence for the target language term helps to prevent the terminologist from simply translating the source language term, by requiring him or her to find an equivalent designation of the concept actually in use in the target language. This helps to ensure authenticity of the target language term and helps to reduce influence of the source language on the target language."

Usage contexts can consist of plain text, and therefore be associated with Lexicalization in order to model the occurrence of word forms in a unit of context. On the other hand, the context can consist of semantically annotated text, as in the case of Semcor [9]. In the later case the UsageContext class is associated with LexicalEntry rather than Lexicalization. Therefore, the object property *isContextOf* (see below) has both LexicalEntry and Lexicalization as range.

The class UsageContext has the following attributes:

- *context*: the textual context in string format.
- *xml:lang*: optional attribute reflecting the language code from ISO639-2[31] associated with the range of the belongsToLanguage object property (see previous section) in the *FAO's languagecode ontology*.

9. Note: supplemental information pertaining to any other element in the data collection, regardless whether it is a term, term-related, descriptive, or administrative. (ISO12620: section 08)

This class can be linked to any class from the LIR model. For the moment, this sort of supplemental information envisages to be captured in a non-formal way through free text. It is possible that in a later stage these differences can be formalized to a greater extent. The Note class will function as an extension point for this potential further formalization.

---

[30] http://www.loc.gov/standards/iso639-2/php/English_list.php

[31] http://www.loc.gov/standards/iso639-2/php/English_list.php

Object properties cannot be within the domain and range of other object properties in OWL DL (which is the OWL variant we are committed to). They are therefore not connected to the Note class by means of the *hasNote* property (see section 3.3 below), and this supplemental information functionality is taken over by the *rdfs:comment* attribute.

The class Note has the following attributes:

- *noteText*: the content of the Note in string format
- *xml:lang*: optional attribute reflecting the language code from ISO639-2 associated with the range of the *belongsToLanguage* object property (see previous section) in the *FAO's languagecode ontology*.

3.3 Description of the LIR Relations between classes

1. *hasLexicalEntry*: the link between the ontology and the LIR.

This relation has, as yet, no semantic characterization apart from "is lexicalized by". It can be further parameterized in order to describe the nature of the mapping between lexical and conceptual knowledge. For instance, an element from a lightweight ontology can be linked to an LIR LexicalEntry with conceptual equivalence. Parameterization of the *hasLexicalEntry* relation is for now enabled as free text by means of the *rdfs: comment* attribute.

Domain:

OntologyElement (external to the LIR as part of http://owlodm.ontoware.org/OWL1.0)

Range: LexicalEntry

Inverse: ***isLexicalEntryOf***

2. *hasSynonym*: lexical semantic equivalence relation between LexicalEntries.

WordNet distinguishes between the lexical relations synonymy and antonymy (for the latter see no. 18) on the one hand, which depend on the lexemes involved in the relation, and conceptual relations between synsets on the other, which do not depend on the lexemes that constitute the synsets. The decision whether two lexical entries in different languages are synonyms, depends on the characterization of the synonymy relation. Since labels are elements from natural language, the logical notion of synonymy (the preservation of truth conditions in all contexts) is hardly ever applicable. Because of this fact, [16] suggest using a weaker notion of synonymy, namely 'semantic similarity', which is defined as "two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value" [16].

In the LIR model we are concerned with capturing lexical knowledge, which is connected, but not equivalent to, ontological knowledge in our model through the *hasLexicalEntry* relation (see above). Therefore we follow this lexical, rather than logical, notion of synonymy.

Domain: LexicalEntry

Range: LexicalEntry

Inverse: ***isSynonymOf***

3. *hasTranslation*: translation equivalence relation between LexicalEntries from different languages.

Domain: LexicalEntry

Range: LexicalEntry

2006–2008 © Copyright lies with the respective authors and their institutions.

Inverse: ***isTranslationOf***

4. *hasVariant*: this property and its sub-properties (points 5-9 below) reflect the *termType* data property associated with Lexicalization. The reason for this redundancy is given in the Lexicalization section above (section 3.2, point 3).

Domain: Lexicalization

Range: Lexicalization

Inverse: ***isVariantOf***

5. *hasSpellingVariant*: a relation between Lexicalizations describing variance in orthographic representation.

Domain: Lexicalization

Range: Lexicalization

Inverse: ***isSpellingVariantOf***

6. *hasTransliteration*: it is related to the *Transliteration* data property described above.

Domain: Lexicalization

Range: Lexicalization

Inverse: ***isTransliterationOf***

7. *hasAbbreviation*: it is related to the *Abbreviation* data property described above. This in turn subsumes the following relations: ***hasShortForm*** and ***hasAcronym***, which are related to the attributes ShortForm and Acronym described above.

Domain: Lexicalization

Range: Lexicalization

Inverse: ***isAbbreviationOf***; ***isShortFormOf***; ***isAcronymOf***

8. *hasScientificName* and ***hasCommonName*** have been defined as inverse relations between LexicalEntries. This gives us a more economical representation of this information, because it reduces the reduplication of this information at the Lexicalization level. If we maintain the *hasScientificName* relation as a relation between Lexicalizations, we need to encode this relation between each common name Lexicalization within each LexicalEntry and each scientific name Lexicalization, not only within a language, but also across languages, since the scientific name is the same for each language specific common name.

In many cases, the directionality of these relations enables the derivation of term types as Boolean attributes for Lexicalization classes. For instance:

X *hasScientificName* Y → X: *ScientificName*: 0; Y: *ScientificName*: 1

However, the following cannot be unambiguously derived:

X has Abbreviation Y → X: fullForm:1; Y: Abbreviation:1

X *hasSpellingVariant* → X: *mainEntry*; Y: Variant

In the first example, X can, for instance, also be a shortForm, albeit with less probability. In the second example, X can equally be a variant.

For a discussion of the shortcomings of these relations, and the use of data properties to remedy these shortcomings, see section 2.2.3B above.

Domain: LexicalEntry

Range: LexicalEntry

Inverse: ***isScientificNameOf***, ***isCommonNameOf***

9. *hasDialectalVariant*: it indicates whether a word form originates from a dialect. The name of the dialect is encoded by the *belongsToDialect* attribute.

Domain: Lexicalization

Range: Lexicalization

Inverse: ***isDialectalVariantOf***

10. *hasNote*: relation between any OntologyElement and Note.

Domain: LexicalEntry, Lexicalization, Sense, Source, Definition, UsageContext

Range: Note

Inverse: ***isNoteOf***

11. *hasSource*: it associates various classes with Source.

Domain: LexicalEntry, Lexicalization, Sense, Note, Definition, UsageContext

Range: Note

Inverse: ***isSourceOf***

12. *hasDefinition*: it associates Sense with Definition.

Domain: Sense

Rage: Definition

Inverse: ***isDefinitionOf***

13. *hasSense*: it associates LexicalEntry with Sense.

Domain: LexicalEntry

Range: Sense

Inverse: ***isSenseOf***

14. *belongsToLanguage*: it associates language origin with a number of classes.

Domain: LexicalEntry, Lexicalization, Sense, Definition, Note, Source, UsageContext

Range: Language

Inverse: ***hasLinguisticExpression***

2006–2008 © Copyright lies with the respective authors and their institutions.

15. *hasContext*: it links contextual information with word forms and lexemes.

Domain: LexicalEntry, Lexicalization

Range: UsageContext

Inverse: ***isContextOf***

16. *isRelatedTo*: this property denotes a general notion of lexical semantic relatedness between Senses.

Domain: Sense

Range: Sense

17. *hasLanguageCode*: this relation has been imported from *FAO's languagecode ontology* (http://www.fao.org/aims/aos/languagecode.owl). It links the FAO Language class to the FAO LanguageCode class with its subclasses **ISO639-1** and **ISO639-2**.

Domain: Language

Range: LanguageCode

inverse: ***isCodeOf***

18. *hasAntonym:* lexical semantic relation between the LexicalEntries expressing semantic opposition.

WordNet distinguishes between the lexical relations synonymy (see no. 2) and antonymy on the one hand, which depend on the lexemes involved in the relation, and conceptual relations between synsets on the other, which do not depend on the lexemes that constitute the synsets.

Domain: LexicalEntry

Range: LexicalEntry

Inverse: ***isAntonymOf***

3.4 OWL Version of the LIR Model

The LIR model has been implemented in OWL. The OWL code is included as Annex 1 to this deliverable. Its URL is http://gate.ac.uk/gate-extras/neon/ontologies/lir1.7.owl

4. Initial tests of the LIR model against FAO resources

In this section we describe some of the evaluation tests conducted at the FAO in order to asses the suitability of the LIR model for FAO linguistic and multilingual needs. We expect to conduct similar experiments with resources from the Pharmaceutical Industry, which represents the second Use Case of the NeOn project.

As already explained in the introduction (section 1), the FAO is an international organization with multiple multilingual resources with different levels of granularity. Two of the most used multilingual lexical resources within the FAO are the AGROVOC thesaurus and the FAOTERM glossary. Both resources were analysed in detail in the previous version of this deliverable (see D2.4.1 [18] -

sections 5.1 and 8.1). They represent two types of traditional lexical resources used by many international organizations, such as the European Union (EU), to solve their multilingual needs. However, those kinds of resources present some problems and deficiencies that Semantic Web technologies could solve.

In the next sections, our aim is to give a brief overview of the main deficiencies FAO resources have by using the example of the AGROVOC thesaurus, and that could be solved in the LIR model.

4.1 Deficiencies of traditional Lexical Resources: the AGROVOC thesaurus

The current AGROVOC thesaurus, developed by FAO in collaboration with the EU in the early '80s, is term oriented. It is a thesaurus, therefore no ambiguities are possible at term level, which will need to be disambiguated by referring to the keyword itself, with consequent mix of single and more generic terms (generally written in brackets after the ambiguous term); e.g. *Alabama (Lepidoptera)* and *Alabama (USA)*, or *Biofilms (bioreactors)* and *Biofilms (microbial contaminants)* and *Biofilms (packaging)*.

In addition to that, in a traditional thesaurus it is not possible to set more than one translation per term. According to this, for example, the English term *Field size* can be translated in French as *Taille des parcelles* or *Dimension des parcelles*. In the current thesaurus one of the translations is assigned as the translation of the descriptor, and the other as an associated non-descriptor. But, in order to realize powerful information retrieval (IR) systems, both French terms should or could actually be used interchangeably.

The traditional thesaurus relationships (Broader Term (BT), Narrower Term (NT), Related Term (RT), USE and UsedFor) do not cover all possible associations between terms, in the sense that it is not possible to retrieve and distinguish an *acronym* from a *full form* description, a *synonym* from a *translation*, or a *scientific name* from a *common name*. With the modern Semantic Web techniques however, it would be of great interest to be able to express such differences.

Another feature that FAO would like to have implemented (but a traditional thesaurus could not) is the possibility to specify lexical variants for dialects or local languages for a geographical region, such as the ones we could find between Spanish used in Spain and Spanish used in Latin America, or Thai expressions which may vary depending on the city they are used in.

In order to be able to allow the representation of more semantic values (meaning of relationships, more linguistic information, etc.), FAO has investigated the development of a Concept Server [15]: a pool of well identified concepts with clear connections to its lexicalizations in multiple languages, and, in addition to concept-to-concept relationships, with relationships between the lexicalizations themselves. The concept server has developed a specific OWL model in which the distinction of concepts, terms and their corresponding written forms are well represented.

The basic idea behind the Concept Server is to have three levels of objects: concepts, terms, and strings. A repository of specific relationships between concepts, terms, and strings has also been developed. With the NeOn LIR model, the same information as the Concept Server has been implemented and further elaborated. Therefore, the LIR model can be considered as a further refinement of the Concept Server. For example, relationships at the string level in the Concept Server have been implemented simply by assigning a specific data type property. An example of this is to be seen in Figure 9, in which we show a spelling variant of the word "organization", namely, "organisation", inserted in the server as a string associated with a term by means of a data type property.

2006–2008 © Copyright lies with the respective authors and their institutions.

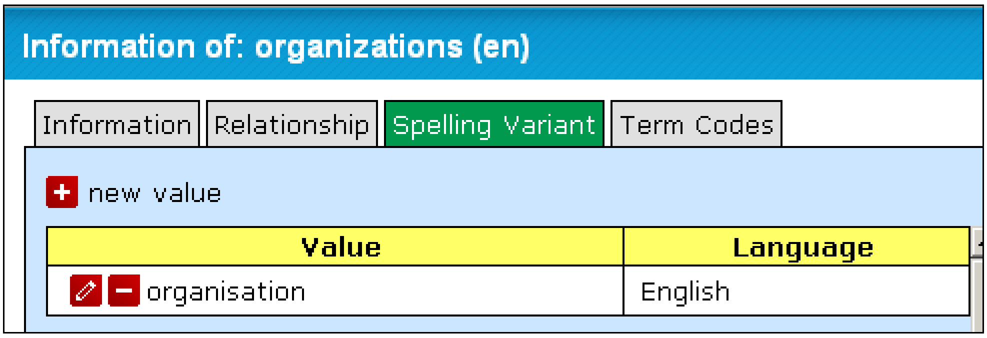


Figure 9. A spelling variant in the Concept Server is inserted as a string associated with a term via a data type property.

In the Concept Server the domain of relationships such as *hasSpellingVariant* is set to c_lexicalization, whereas the range is a simple string (see Figure 10).

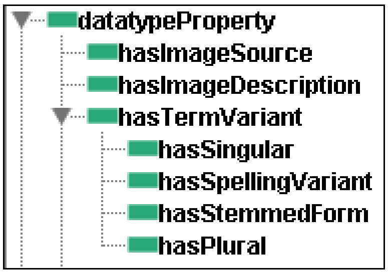


Figure 10. Some data type properties in the Concept Server

FAO is now creating specific modules from the AGROVOC to the Concept Server. These modules can be tested with the AGROVOC Concept Server Workbench (http://www.fao.org/aims/agrovoccs.jsp), and they can be accessed online at ftp://ftp.fao.org/gi/gil/gilws/aims/kos/agrovoc_formats/owl/concept-server-modules/v0.2.

In order to further test the LIR model, FAO staff is loading with the Neon Toolkit all these modules and they will be also analyzed with the LIR structure. Initial tests indicate that the LIR model rapidly load in the NeOn Toolkit without any problem (see Figure 11).

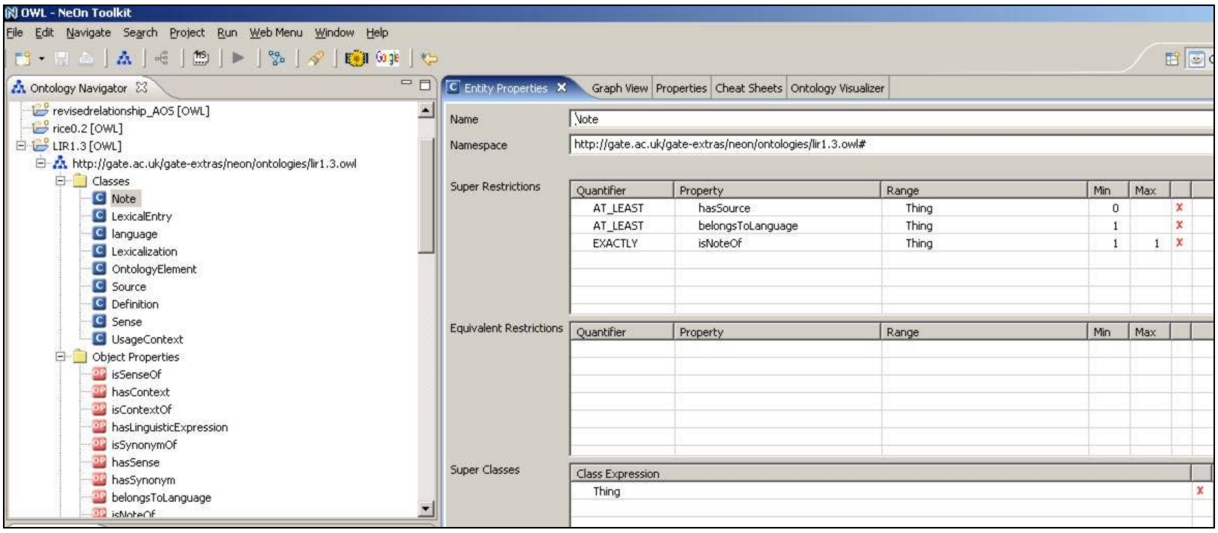


Figure 11. LIR model loaded in the NeOn Toolkit

Further benefits for FAO ontological models about agriculture could be achieved from the LIR thanks to the definition of quantifier for relationships. These identify how many relationships could or should exist between the terminological elements of the modules.

4.2 Benefits of the LIR to FAO Needs

The LIR model allows a very granular specification of relationships between elements of an ontology. In particular, it identifies well-defined relationships at the terminological layer that is used to represent ontological concepts.

In FAO, there are several resources that could benefit from this specification: the AGROVOC Thesaurus (or better the AGROVOC Concept Server, which is the corresponding ontology created from the original thesaurus), and the FAOTERM system, a multilingual glossary that identifies synonyms, acronyms, full forms and translations of Agricultural terminology. In addition, all recently developed domain-specialized ontologies could also be modelled following the LIR paradigm.

The examples below show how some problems mentioned in section 4.1 above have been solved by modelling with the LIR model the multilingual information associated with an ontology. Those problems are:

- Establishment of well-defined relations within one language
- Establishment of well-defined relations across languages
- Conceptualization mismatches among different cultures and languages
- Representation of non-native language expressions

In the following, we try to illustrate how the LIR solves these problems by means of real FAO examples.

a) Example 1: Establishment of well-defined relations within one language

The first example concerns the establishment of relations among lexical elements belonging to the same language. Specifically, this case exemplifies the use of various

2006–2008 © Copyright lies with the respective authors and their institutions.

acronyms and one *full form* attached to one and the same concept. The LIR model does not only allow specifying the language for all lexicalizations, but helps identify the exact acronyms and full forms across synonyms or translations.

In the example given in Figure 12, three lexical entries (01:LexicalEntry, 02:LexicalEntry and 03:LexicalEntry) are associated with the same concept (C21:Class), which means that they are all terms that identify one and the same concept. Two lexical entries (01:LexicalEntry and 02:LexicalEntry) belong to the same Language (English), whereas the third lexical entry (03:LexicalEntry) belongs to French. The two English lexical entries are considered synonyms, and translations of the French lexical entry. Each lexical entry contains two lexicalizations. For example: 01:LexicalEntry includes 011:Lexicalization and 0111:Lexicalization, whose labels are *FAO* and *Food and Agriculture Organization*, respectively. *FAO* is the acronym for *Food and Agriculture Organization*, and, moreover, it is considered the main entry. *FAO of the UN* and *Food and Agriculture Organization of the United Nations* are deemed synonyms of FAO and *Food and Agriculture Organization,* and translations of *OAA* and *Organisation des Nations Unies pour l'Alimentation et l'Agriculture* in the French language.

Thanks to the LIR it is possible to retrieve synonyms within the same language associated with the same concept, and distinguish different term types such as acronyms and full forms.

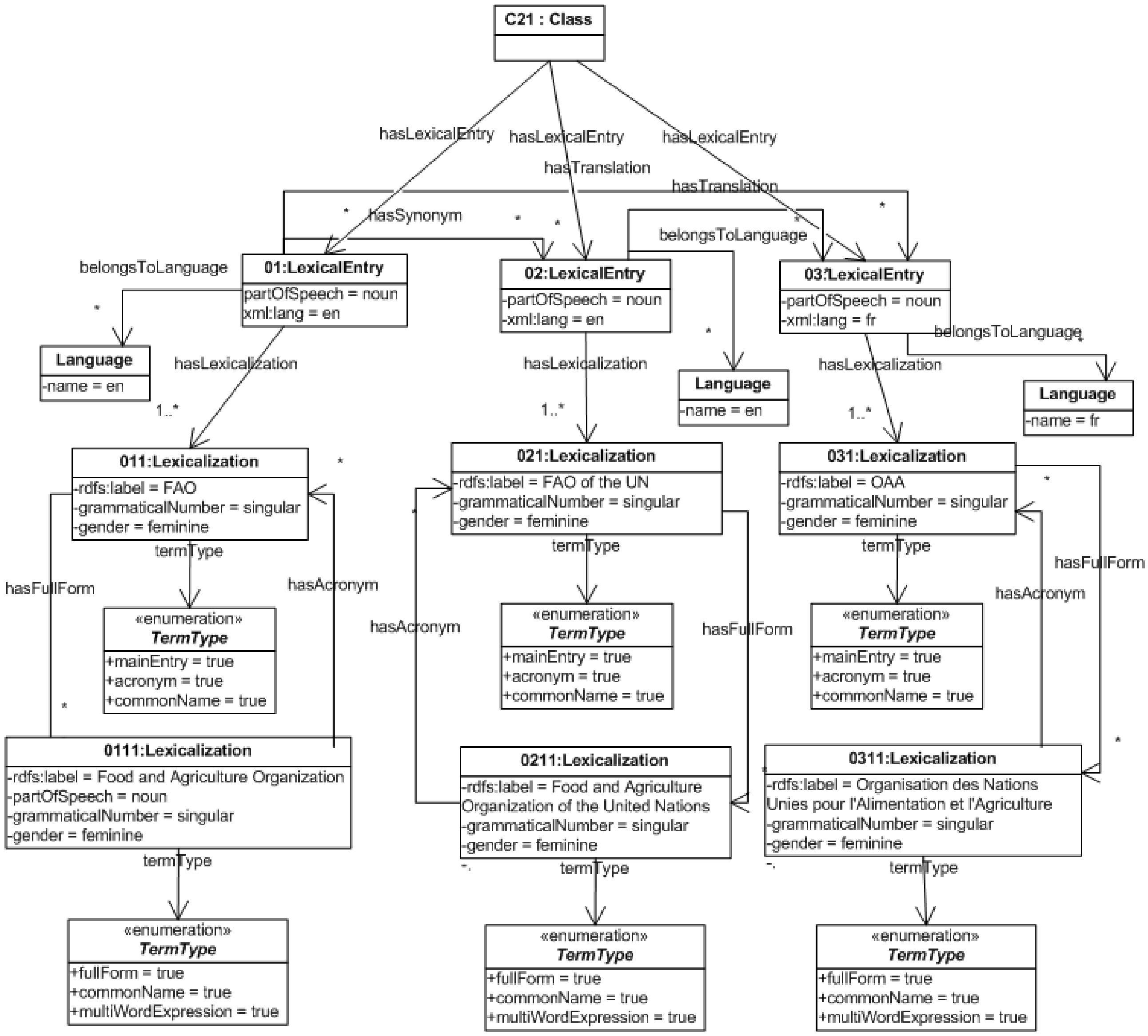


Figure 12. Representation of acronyms and full forms within a language

b) Example 2: Establishment of well-defined relations across languages

The second example highlights the possibility given by the LIR model to represent scientific names and use them across languages (scientific names are in Latin and are internationally accepted over scientific communities). Variants in the same language (e.g. *Buffaloes (syncerus)*) can therefore be connected to the same scientific term, such as the English *African buffaloes* and the Japanese □□□□□□*.*

We have illustrated in Figure 13 how the concept *buffaloes* (C133:Class) has four lexical entries associated (01:LexicalEntry, 02:LexicalEntry, 03:LexicalEntry, 04:LexicalEntry). Two belong to the English language and contain synonymous lexicalizations (011:Lexicalization and 021:Lexicalization) represented by the labels *African buffaloes* and *Buffaloes (syncerus)* respectively. Then, we have a lexicalization in Latin that represents the scientific name, and it is accordingly related with the rest of lexical entries by means of the object property *hasScientificName*. Finally, 04:LexicalEntry belongs to the Japanese language and has the label □□□□□□, which is also the common denomination in Japanese of the *Syncerus caffer* scientific name, and, at the same time, the translation of the two lexicalizations in English.

2006–2008 © Copyright lies with the respective authors and their institutions.

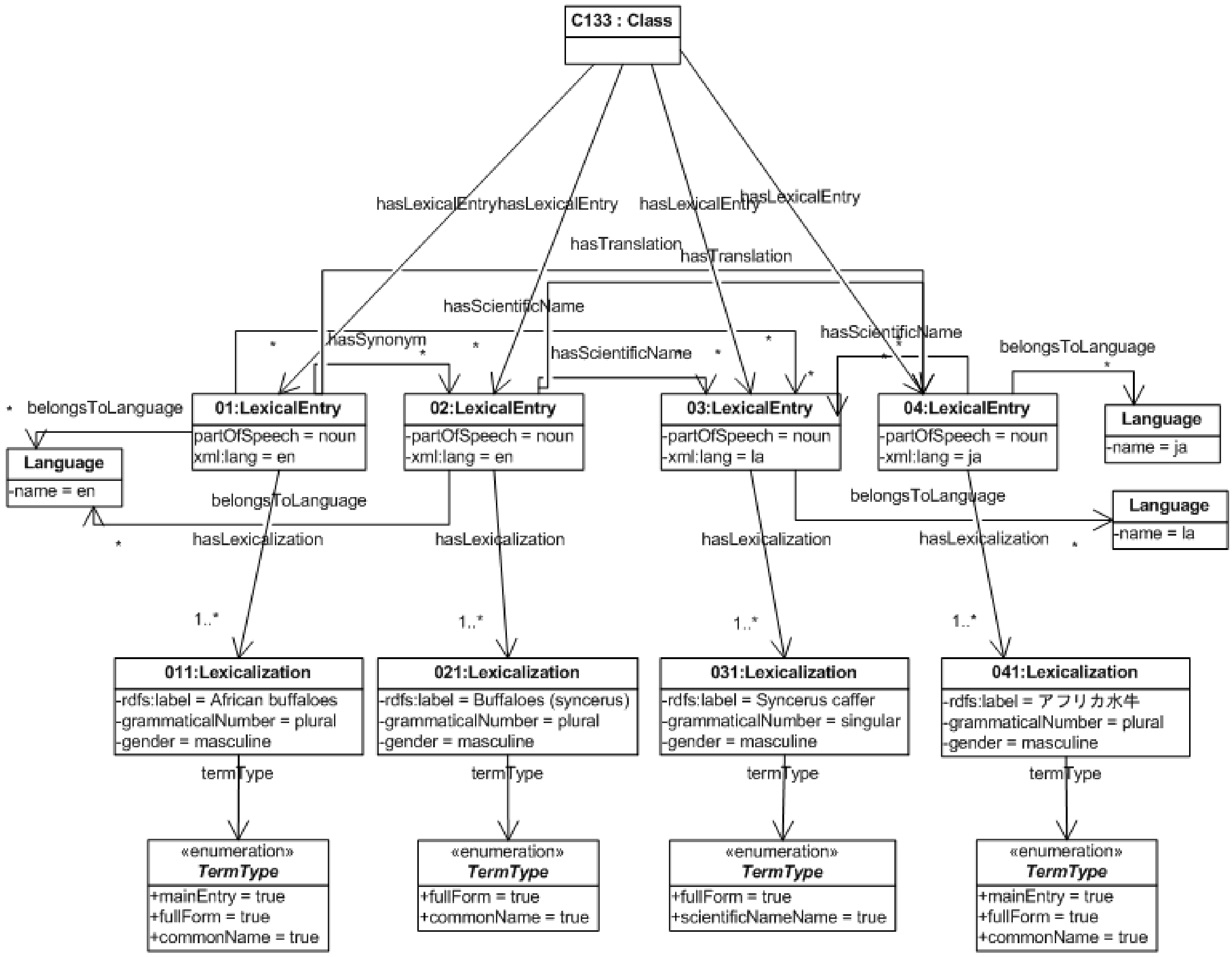


Figure 13. Representation of Scientific Names and Common Names across Languages

c) Example 3: Conceptualization mismatches among different cultures and languages

One of the main problems in representing lexicalizations is due to the lack of specific terminology in one language for a concept, i.e., when there are no lexicalizations in a language for a specific concept. Closely related to this, it is the case in which one concept that corresponds to one lexicalization in one language does not have an exact equivalent in the other language, but corresponds to more than one concept and, in its turn, it has also various lexicalizations in that target language. These are well-known translation difficulties among professional translators.

In order to be able to express that kind of translation specificities among languages and cultures, the LIR has foreseen the classes Sense, Definition and Note.

Let us imagine the case in which our ontology contains the class *river*. In English, river is defined as a natural stream of water of usually considerable volume. The French language has no exact equivalent, as far as we know. There is the term *course d'eau*, which is slightly more general, and could be considered a translation of *stream of water*, and the terms *fleuve* and *rivière*, which may be somewhat more specific. Broadly speaking, *fleuve* is a river that flows into the see, whereas *rivière* is a river that can flow into the sea or into another stream. In the LIR model, the classes Sense, Definition and Note allow us to explicitly define such nuances that exist among languages. We have tried to represent the following scenario in Figure 14. In this case, the ontology concept *river* (C2321:Class) has thee lexical entries associated (033:LexicalEntry, 031:LexicalEntry, and 030:LexicalEntry).

The lexicalization related to the English language is *river*, whereas there are two lexicalizations in French, *fleuve* and *rivière*. The French lexicalizations are considered synonyms in most cases, and consequently, both lexical entries are related by means of the *hasSynonym* object property. However, we contemplate the possibility of further specify the relations established among lexical entries in the future, if deemed necessary. Although they are considered synonyms and their senses are related, they are differently defined, as has been explained in the Definition and Note classes.

In this way, the LIR permits the association of a complete and complementary set of linguistic data to ontology elements that allow the localization of ontology elements to a certain language and culture.

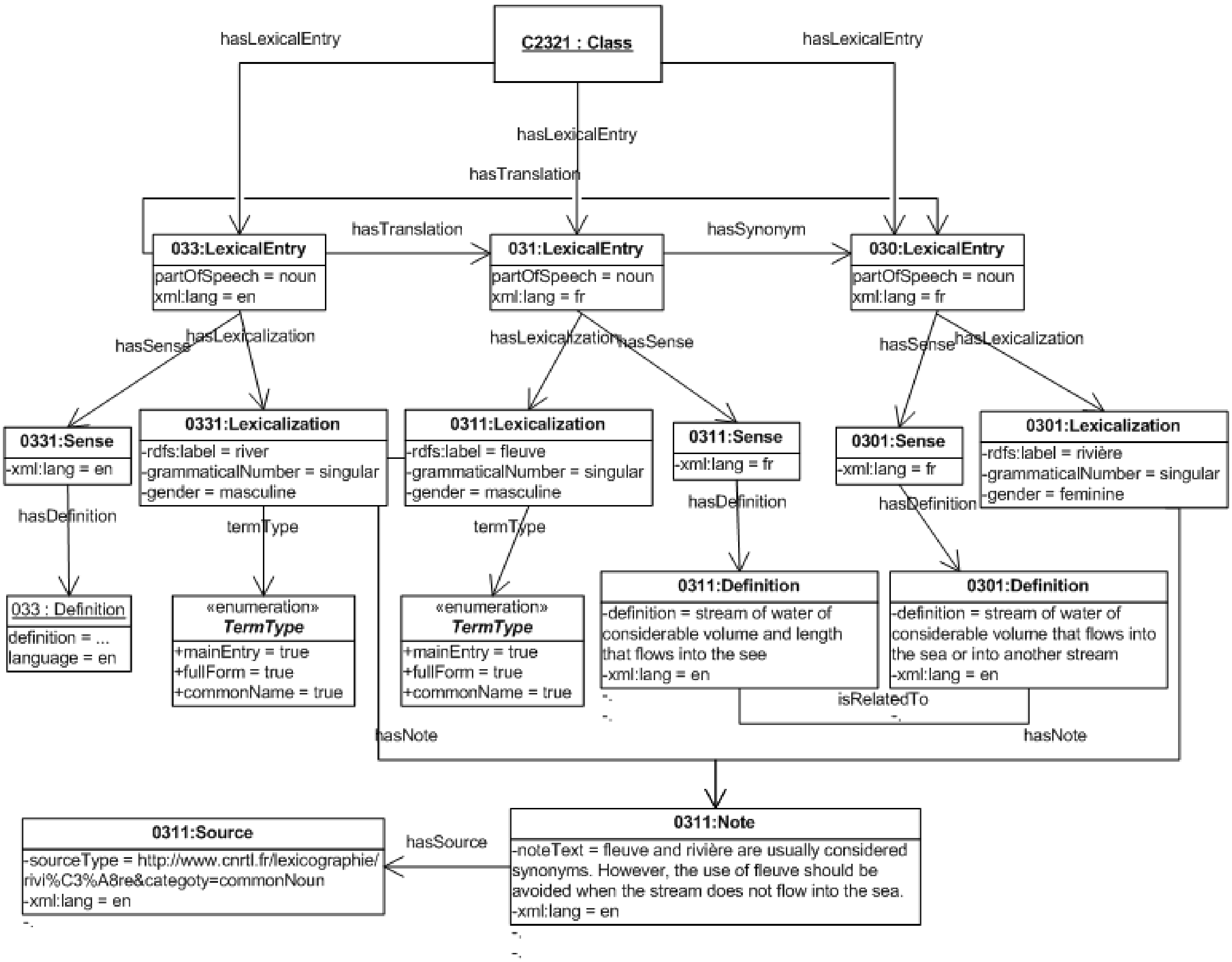


Figure 14. Representation of conceptualization mismatches among different cultures and languages

d) Example 4: Representation of non-native language expressions

The last example we wanted to include in this document is related with the possibility offered by the LIR of expressing that certain lexicalizations that belong to a specific language are used in another (foreign) language. This is the case of the word form *paella*, which is a Spanish word used in other languages such as English and Italian.

The possibility given by the LIR model in this case is that using the *belongsToLanguage* link, we can express that a term is used in a specific country or a specific culture, and using the *xml:lang* attribute we can identify the real language of this term (see Figure 15).

2006–2008 © Copyright lies with the respective authors and their institutions.

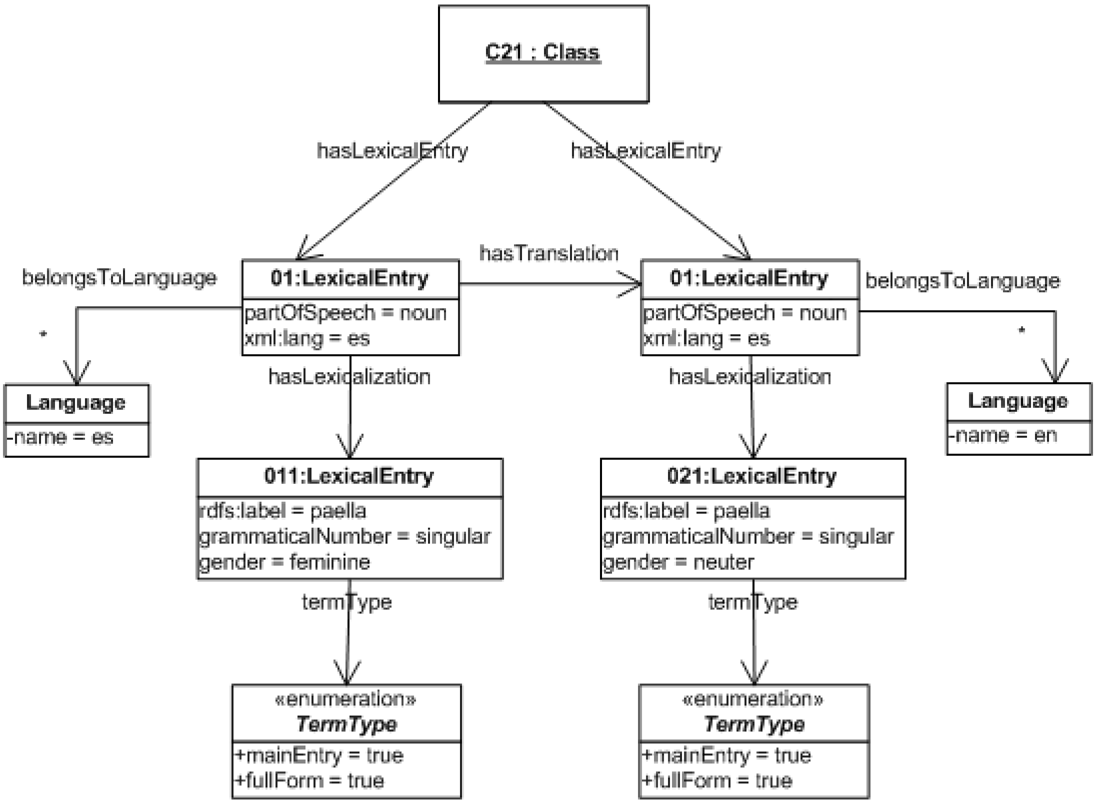


Figure 15. Representation of non-native language expressions

5. Discussion of the LIR and Future Work

The LIR model described in this deliverable represents a core set of linguistic information units that cover a range of linguistic phenomena. This range is deemed sufficient for immediate NeOn purposes, in that it covers the NeOn Use Case specifications (section 4).

The LIR model has incorporated a selection of data categories from existing standard representations for linguistic and terminological resource description such as the following:

- ISO 16642:2003, *Computer applications in terminology – TMF (Terminological Markup Framework)*
- ISO 24613 *Language Resource Management – LMF: Lexical Markup Framework*

This means that it is linked to these representations and any other representation schemas that are derived from or linked to these standards, at least with respect to the data categories they have in common. For a short description of some of these standards, please see D2.4.1 section 2.

The motivation underlying these representational choices is that the issue of lexical representation is far from resolved, and is important to achieve synergy and interoperability with world-wide interoperability initiatives using these representations.

Standards cover a percentage of the linguistic information that is available on line. Individual linguistic and terminological resources largely differ in the explicit linguistic information they expose, which may vary in format, content granularity and motivation (linguistic theories, intended users, purpose or system-oriented scope etc…) [21].

Various standardization initiatives are trying to enhance the interoperability between formats and levels of coverage. Most notably, the ISO Technical Committee 37, “Terminology and other

Language and Content Resources"[32], is developing a Data Category Registry (DCR) [12,13]. This registry will provide a reusable set of (standardized) data denoting linguistic concepts that cover a range of linguistic domains. The concepts in the DCR can be referenced to from all sorts of tools and resources. Therefore, the DCR acts as an intermediate between those tools and resources.

In brief, the DCR will contain every possible data category of a certain standard. Parts of the metadata model of a language resource can include these data categories, and thus share common semantics with other resources. It will be further developed and exploited in a number of research infrastructure projects such as CLARIN[33]. These are meant to form the basis for future eScience scenarios in which it is foreseen that researchers can seamlessly access and combine resources and services offered by various service and repository centres.

5.1 LIR within an ontology network

The LIR model presents a non-exhaustive list of data categories for linguistic description. Only limited by its present coverage and usage within NeOn, the LIR can be seen as the hub within a set of networked ontologies representing re-engineered versions of the standard representation formats.

These ontologies are either in existence or under development on the basis of other data formats such as DTDs. Figure 16 shows a networked ontology structure, in which three resources are linked through the LIR.

Because the LIR represents only a subset, it misses many data categories, and therefore the ability to cover the whole range of linguistic/terminological description. For instance, in Figure 17 below illustrates the LIR alignment with the data structure used by the AGROVOC Concept Server (CS) (section 4).

It is clear that apart from correspondences, there are also mismatches between the two schemas: there is a difference in descriptive granularity at the level of "scientific term". CS captures subcategories of scientific term, whereas LIR denotes only the general class.

Moreover, whereas the characterization of a lexicalization as scientific term is expressed in LIR by means of a data property, in CS this is expressed by means of a class with subclasses for more specific types of scientific term.

[32] http://en.wikipedia.org/wiki/ISO/TC37

[33] http://www.clarin.eu/structure

2006–2008 © Copyright lies with the respective authors and their institutions.

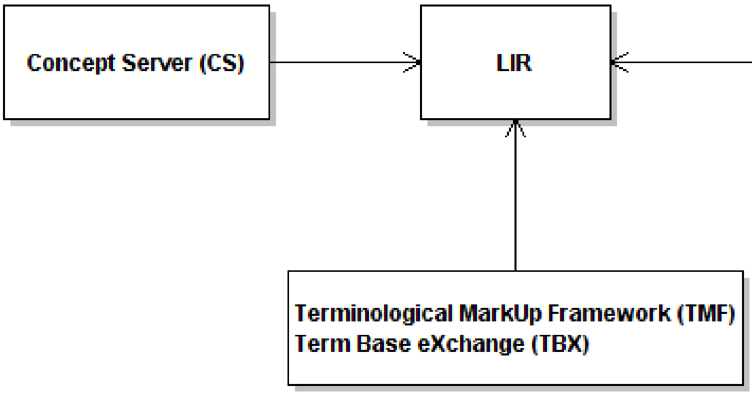


Figure 16. Networked ontologies for linguistic/terminological representation

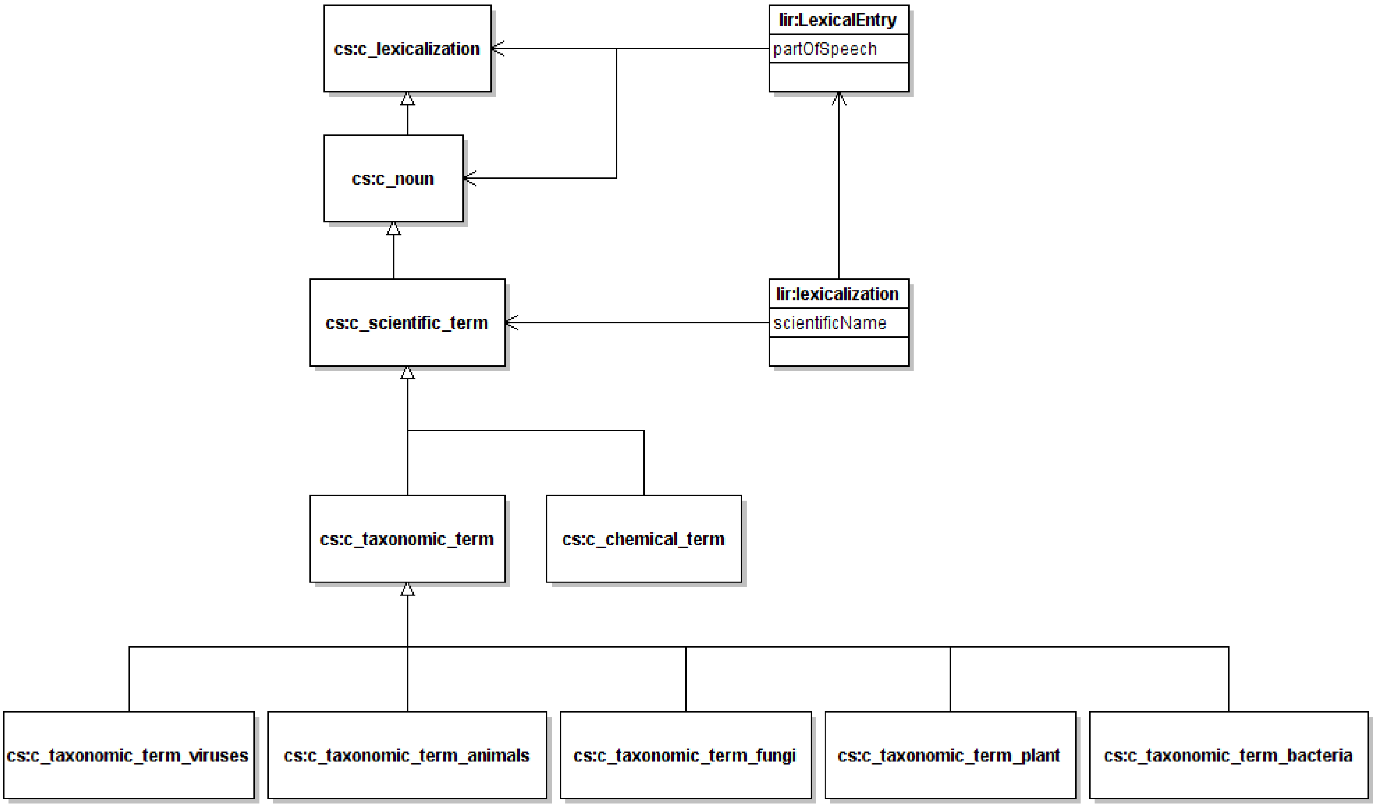


Figure 17. Alignment of the AGROVOC Concept Server and the LIR

Information not contained yet in the LIR, and provided by the other standard representations, will need to be linked through external links between these resources, circumventing the LIR (Figure 18).

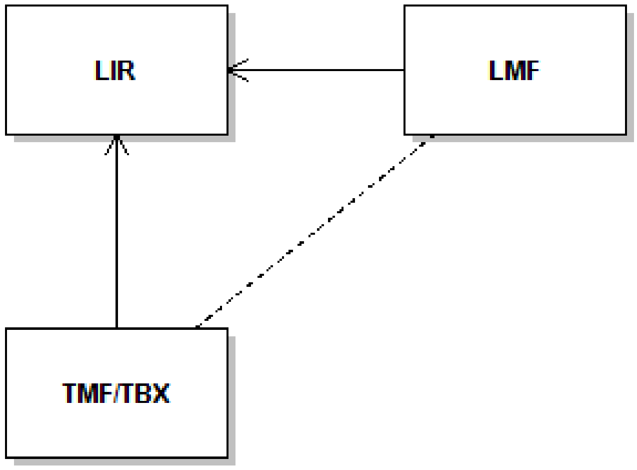


Figure 18. The LIR as a partial hub

If we want the LIR to maintain its hub function for linguistic/terminological description, it will need to cover an increasing number of data categories. This can be done in two ways:

1. Extend its coverage by incorporating all available (standard) data categories, e.g. the ISO data category registry, other standard descriptions, or, if necessary, non-standard resource specific descriptions. These can be incorporated as individual concepts, or as modules consisting of related concepts from particular ontologies.

2. Use links to external ontologies rather than incorporation. The links should describe at least the C-Owl relations equivalence, hypernymy, hyponymy and partial overlap [2]. Further, a more fine-grained mapping language can be used such as the one proposed by [27] and [8].

5.2 Some examples of possible refinements of the LIR

a. Part of Speech

In LIR, the data property partOfSpeech is associated with LexicalEntry, and denotes only major part of speech types: "noun, verb, adjective, adverb, proper noun, other".

The general classes are suitable for the morphosyntactic description of lexical entries, because these are underspecified for fine-grained morphosyntactic description such as inflection and conjugation.

This general typology, however, cannot accommodate the more specific part of speech types actually provided by certain lexical resources, such as NN and NNS in the Penn Treebank[34], representing singular nouns and plural nouns respectively. Within the LIR, these phenomena are captured at the Lexicalization level.

If we want to capture these morphosyntactically specific tags such as 'plural noun' and 'verb imperfect tense', we need to promote the LIR data attribute partOfSpeech to a disjoint class, with

[34] http://www.comp.leeds.ac.uk/amalgam/tagsets/upenn.html

2006–2008 © Copyright lies with the respective authors and their institutions.

as many subclasses of partOfSpeech as necessary. The number of subclasses can be constrained by adhering to a standard list of morphosyntactic labels, e.g. EAGLES[35].

b. Semantic relations between word senses

LIR property no.14 *isRelatedTo* associates Senses with each other.

This property denotes a general notion of relatedness, and is therefore incapable of capturing more fine-grained semantic relations, such as the thesaural relations "NARROWER TERM", "BROADER TERM" and "RELATED TERM", or the relations from EuroWordNet, such as:

role_agent

role_instrument

role_patient

role_location

role_direction

role_manner

role_result

in_state

state_of

Incorporation through any of the two mechanisms described above is desirable, in order to enhance the LIR's lexical semantic coverage.

c. UsageContext

A more fine-grained typology of this class is expected within ISO, with subcategories such as the following additional data attribute *contextType* (included in the TBX specification[36]):

The characterization of a context according to a set of theoretical or

pragmatic types.

The attribute value of *contextType* must be one of the following:

• definingContext

• explanatoryContext

• associativeContext

• linguisticContext

• metalinguisticContext

• translatedContext

---

35 http://www.ilc.cnr.it/EAGLES/annotate/annotate.html

36 http://www.lisa.org/TBX-Specification.33.0.html

6. Second Version of the LabelTranslator NeOn plug-in

This section describes the features and design aspects of the second version of the LabelTranslator NeOn plug-in which improves research reported in [18]. LabelTranslator is a NeOn plug-in that suggests translations of ontology labels among English, Spanish and German, with the purpose of localizing ontologies [6,7]. In this section, the main characteristic of the first version of LabelTranslator are summarized highlighting why this first version does not fulfil some of the characteristics that, in our opinion, an ontology localization system should have in the context of the Semantic Web.

6.1 Main characteristics of the first version of the LabelTranslator NeOn plug-in

This section describes the features and design aspects of the first version of LabelTranslator, which offers users a set of functionalities for linguistically enriching the labels of an ontology. These characteristics can be summarized as follows:

- **Ontology importation**. LabelTranslator supports the localization of F-Logic, (subsets of) RDF(S) and OWL ontologies. Moreover, this plug-in localizes ontologies of different domain whose labels are described in English, Spanish or German.

- **Translation of ontology label(s)**. LabelTranslator supports automatic translations among the languages above described. To search for translations and senses of each ontology label, the system accesses different resources: 1) remote lexical databases as EuroWordNet, 2) multilingual dictionaries as IATE, and 3) translation services as GoogleTranslate, Wiktionary, Babelfish, and FreeTranslation.

- **Discovery of translation senses**. LabelTranslator uses EuroWordNet in order to discover the different senses of each translated label. The senses of each translated label are used by LabelTranslator to disambiguate candidate translations.

- **Translation selection**. LabelTranslator uses a ranking method which sorts out each translation sense according to similarity with its lexical and semantic context. The ranking method relies on a measure of disambiguation based on glosses. Once all translation senses are ranked, the user can either confirm the translation proposed by the ranking method, or select the translation that better fits the ontology context.

- **Ontology enrichment**. The system updates the ontology model with the selected translation(s). In the first version, the link established between the ontology concepts and their associated translations is characterized by simple references between concepts and labels (as offered by the standard `owl:comment` and `rdfs:label` properties)

6.1.1 Limitations of the first version of LabelTranslator

Once the main characteristics have been described, we want to point out some limitations this version has in order to perform ontology localization in the context of the Semantic Web.

- **Problems in handling translation specificities**. A key issue for any localization system is to handle some traditional problems in translation: for instance, some concepts from the source language have no equivalents in other languages, polysemous words and homographs, quasi-synonyms, regional differences within the same language, etc. However, the first version of LabelTranslator did not deal with these problems in a fully satisfactory way.

2006–2008 © Copyright lies with the respective authors and their institutions.

- **Lack of domain independence to discover translation senses.** If we consider applying our system to heterogeneous ontologies, we need a system which is able to locate ontologies without any pre-defined assumption about the ontological domain. However, this first version is limited only to EuroWordNet as knowledge source to discover the senses of each translated label. Although using a huge lexical database, such as EuroWordNet, provides many benefits, a dependence on a single lexical resource is not desirable due to different reasons: we depend on the availability of senses, and some may not appear in it. This approach works well in the localization of ontologies of general domain covered by EuroWordNet.

- **Lack of a modular approach to support linguistic information.** Deliverable 2.4.1 "Multilingual Ontology Support" suggests the suitability of keeping ontology knowledge and linguistic (multilingual) knowledge separately and independently. The current trend in the integration of multilinguality in ontologies also follows this modular approach. The main advantages of modelling modality in this way are related to the enormous range of possibilities that multilingual information integrated in ontologies can offer to applications in the Semantic Web. However, the first version of LabelTranslator followed a non-modular approach, in which multilingual information was embedded in the ontology by means of the RDFS/OWL predicates. This approach has important limitations related to the restricted amount of linguistic information that can be attached to ontology concepts. Furthermore, multilingual information is limited to strings without information about senses in their respective languages, and without provenance of the information, which made concept localization to different natural languages quite difficult.

In the next section we describe the innovations incorporated in the second version of the plug-in in order to solve the limitations above described.

6.2 Innovations of the second version of LabelTranslator

First, we use a three layered architecture in order to compare and contrast both versions. In Figure 19, the high level architecture of both versions is shown. The figure on the left shows the components of the first version, while the figure on the right represents the enhanced architecture of the second version.

The first layer encapsulates the graphical user interfaces that permit interaction with the user. While the first version uses the current NeOn ToolKit for storing the multilingual information related to a specific ontology label, the second version of LabelTranslator adds support to the new linguistic information model, LIR, described in section 3. Moreover, we have implemented a *LinguisticView* (LinguisticReposService in Figure 19), which contains a set of fields for editing the linguistic information associated now to each ontology element.

In the second layer, i.e. the business functionality of the system, the second version of the prototype adds a new algorithm used to support semi-automatic translations of ontology elements and applying the characteristics of the LIR.

Finally, in the third layer, which is used to store the multilingual information, unlike the previous version the new prototype adds a new repository (*LinguisticReposService*) to store the linguistic information associated to each ontology element. Consequently, the linguistic information is stored in two places at the same time.

First Prototype	***Second Prototype***

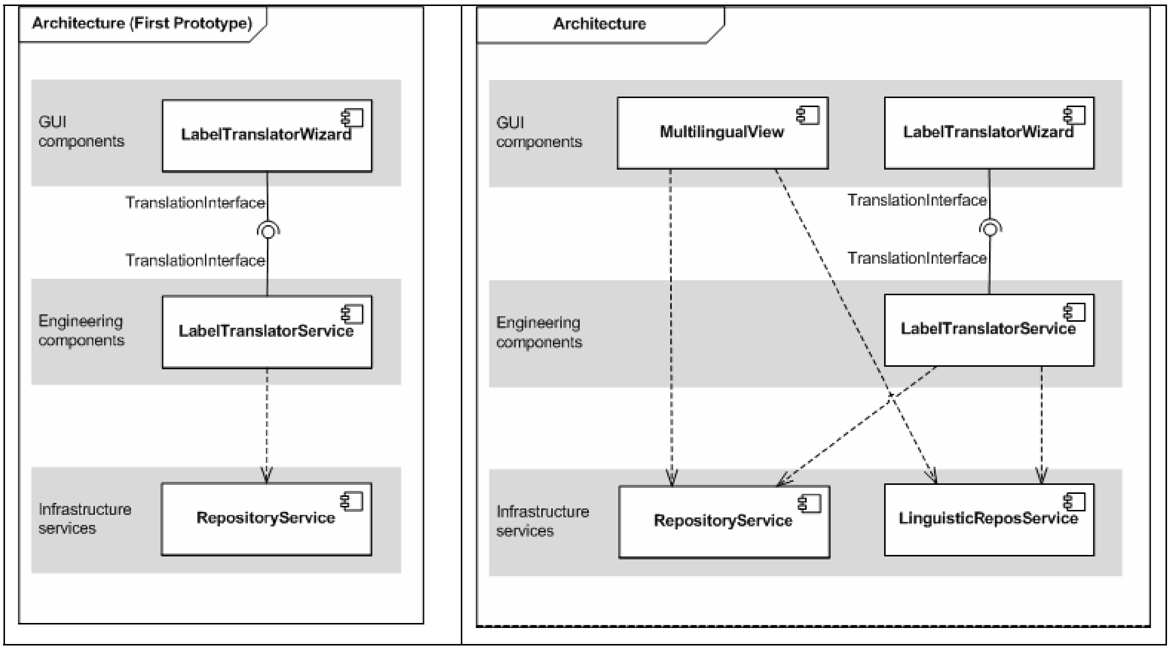


Figure 19. Three layered architecture of the first and second versions of LabelTranslator

Figure 20 shows the three main components of the new version of our system (GUI component, Ontology Localization component, and Repository component) and illustrates the process for enriching an ontology with linguistic information using the LIR model. Let us see them in greater detail:

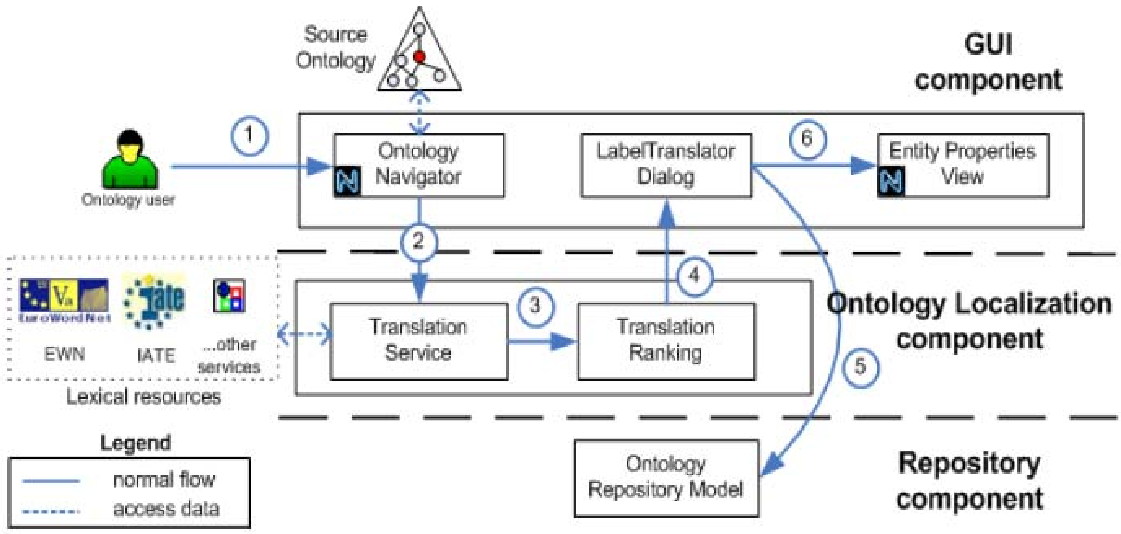


Figure 20. Main components of the second version of LabelTranslator

1) **GUI component**: This component controls the GUI in order to show the multilingual results appropriately. Once invoked, LabelTranslator uses some views[37] of the Neon ToolKit to

[37] In the NeOn ToolKit a view is typically used to navigate a hierarchy of information, open an editor, or display properties for the active editor.

2006–2008 © Copyright lies with the respective authors and their institutions.

load the ontology and to store the multilingual results respectively. A more detailed description of the functionalities of the new interface implemented to support the LIR model can be found in section 5.3.

2) **Ontology Localization component**: This component is responsible for obtaining the most appropriate translation for each ontology label. It relies on two advanced modules. The first one, *Translation Service*, automatically obtains the different possible translations of an ontology label by accessing different linguistic resources. This service also uses a compositional method in order to translate compound labels (multi-word labels). A more detailed description of the Translation Service can be found in Section 5.4.1. The second module, *Translation Ranking*, sorts the different translations according to the similarity with its lexical and semantic context. The method relies on a relatedness measure based on glosses to disambiguate translations. This is done by comparing the senses associated to each possible translation and their context. More details about the ranking method can be found in Section 5.4.2.

3) **Repository component**: This component captures all the linguistic information associated with ontology elements. LabelTranslator supports the new version of the linguistic model, the LIR, described in section 3, and designed for representing multilingual information in ontologies. A comprehensive solution to the problems of managing the conceptual knowledge and the linguistic knowledge by means of synchronization techniques can be found in Section 5.5.1.

In the next sections we describe in more detail the improvements implemented in each layer of the architecture above described.

6.3 LabelTranslator GUI component

The main effort in this component has been put in the implementation of a new interface that supports the editing of the linguistic information associated to each ontology element. When the ontology editor user imports a new FLogic, RDFS or OWL ontology in Neon, the LabelTranslator system automatically builds an empty linguistic model associated to the ontology under consideration. This model is used by our plug-in to store the linguistic information associated with each ontology element.

In order to show the new linguistic interface, the user has to select a frame (class or property, for example) in the Ontology Navigator (see Figure 21), then (s)he chooses the *Linguistic Information* page shown in the Entity Properties View (see bottom of Figure 22). All fields and tables that show linguistic information correspond to the new version of the LIR model described in section 3.

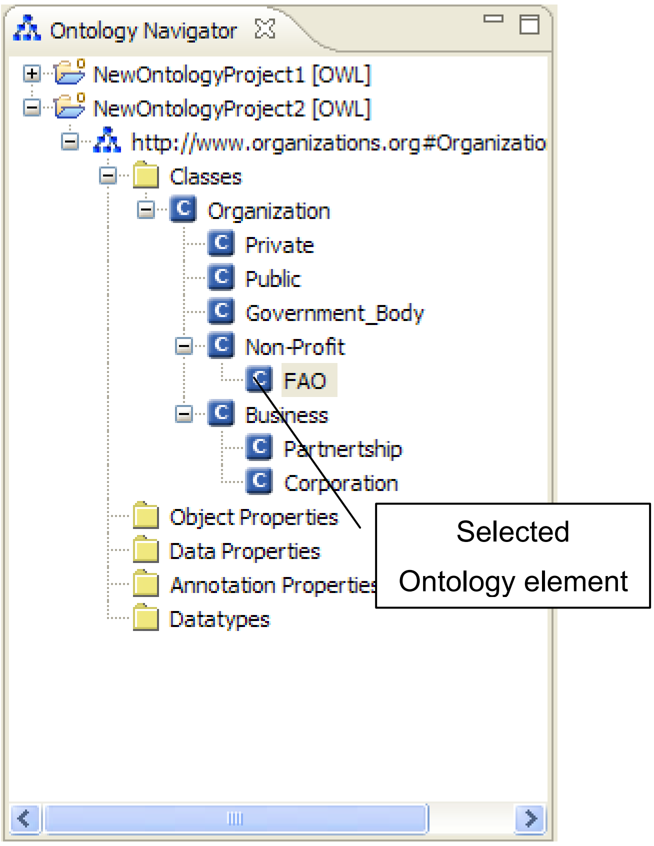


Figure 21. Ontology Navigator with a selected ontology element

Figure 22 shows the *Linguistic Information* page associated to the sample ontology element *FAO*. Initially, the *Linguistic Information* page shows five sections that correspond to the lexical entries of the selected ontology element (FAO in our example) and the associated information of each lexical entry: *lexicalizations*, *senses, usage contexts* and *sources*. For instance, in this case the concept FAO has three lexical entries, two in English and one in Spanish.

The Lexical Entries section represents the master (in a master/detail model) from which it is possible to deploy the related information. The information shown in the different lexical entry sections depends on the selected lexical entry (LexicalEntry_2 in our case).

2006–2008 © Copyright lies with the respective authors and their institutions.

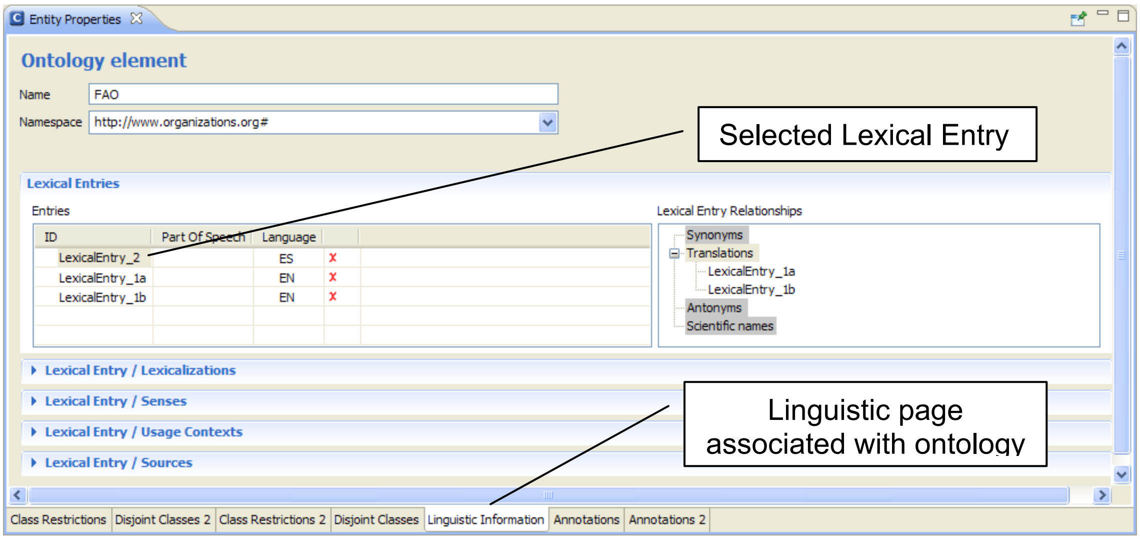


Figure 22. Linguistic Information page with data of the concept FAO

Of course, every time that the user chooses a new entry, the interface automatically displays the information correlated in the different sections. For example, Figure 23 shows the lexicalizations associated with the selected lexical entry, shown in Figure 22.

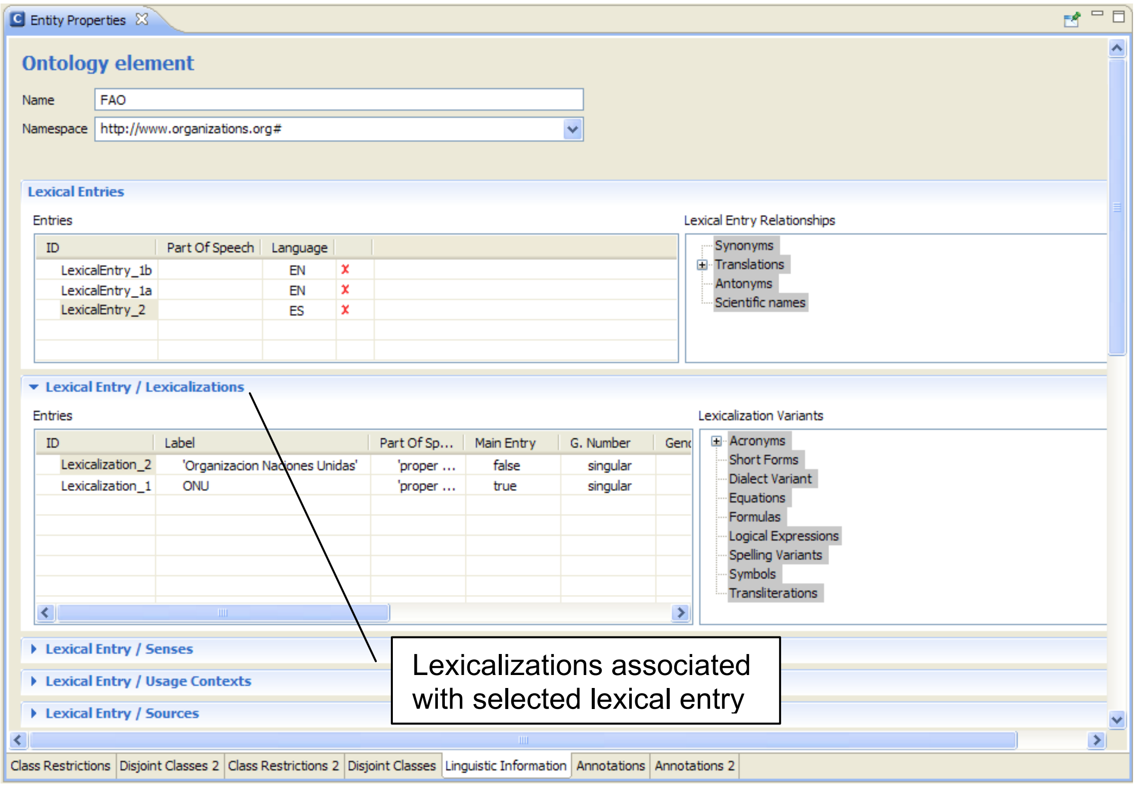


Figure 23. Lexicalizations associated with the selected LexicalEntry associated with the concept *FAO*

Additionally, if the user selects an entry in a section different to the master section (Lexical Entries section), then LabelTranslator automatically reconfigures the sections in order to show the related information with the last selected entry. For example, supposing that the user has selected an entry of the section Lexicalizations, the new sections associated to that entry will be shown to the user, as Figure 24 displays (see bottom of the figure).

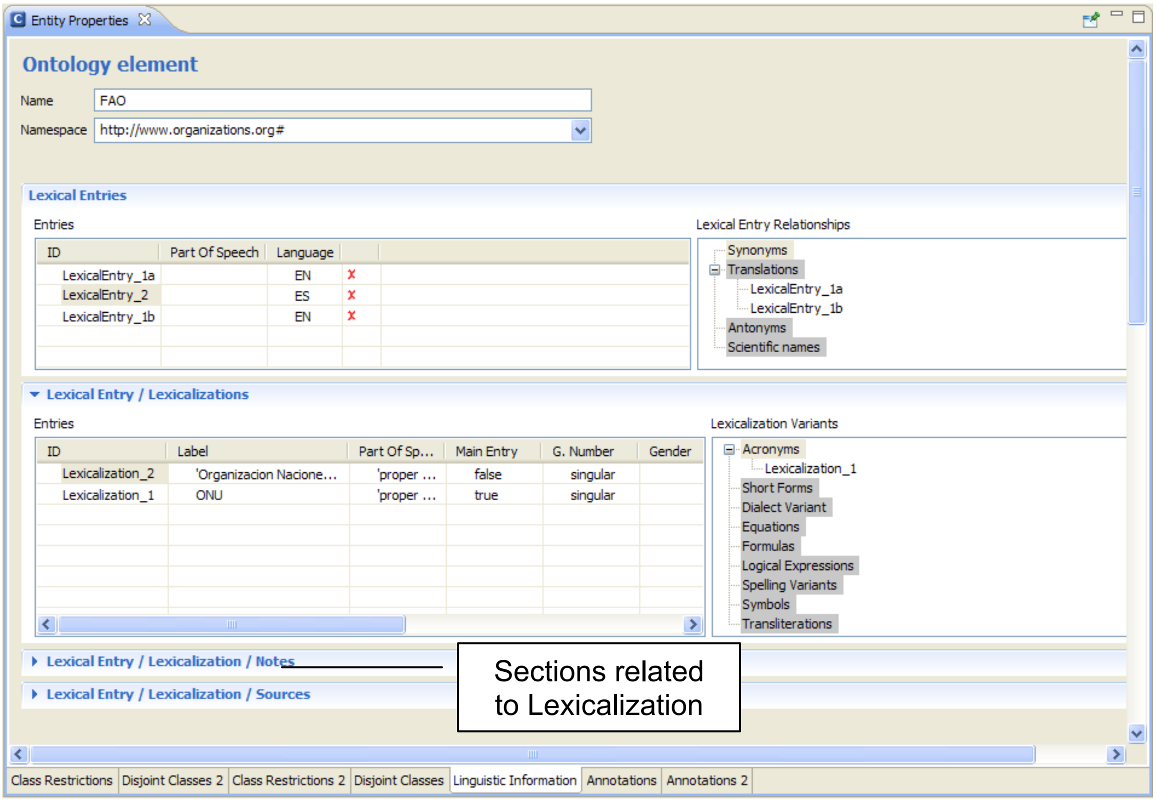


Figure 24. Sections associated with the Lexicalization class

6.4 LabelTranslator Ontology Localization Component –translation algorithm

This component is responsible for obtaining the most suitable translation for each ontology label. It should be noted here that we only focus on the issues which are specific to the new version of LabelTranslator and have not been tackled by the first version. As an illustrative example, let us consider the extract of the sample *University Ontology* shown in Figure 25. Let us suppose that the user wants to translate the term *chair* from English into Spanish. According to the domain of the sample ontology, the correct translation of the selected term should be in the sense of the position of a professor, rather than in the sense of a place where a person can sit, or an instrument of execution by electrocution, etc.

2006–2008 © Copyright lies with the respective authors and their institutions.

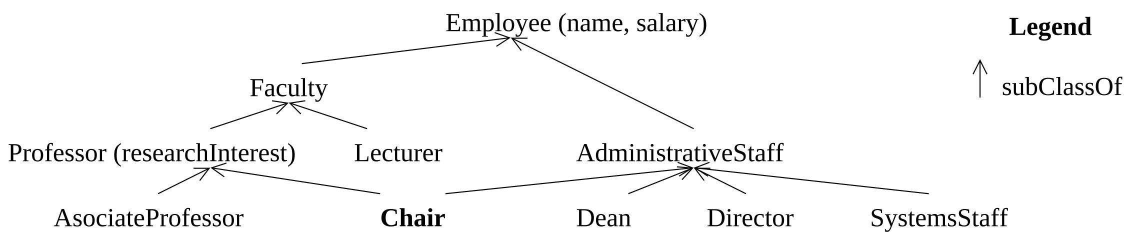


Figure 25. Extract of the sample University Ontology

6.4.1 Translation Service

In this section we provide details about how the system obtains the different translations of an ontology label (which can name different kinds of ontology terms: concepts, properties or relations) by using different linguistic resources.

The *Translation Service* takes as input an ontology label *l* described in a source language and returns a set of possible translations $T = \{t_1, t_2, ..., t_n\}$ in a target language. The current prototype supports translations in English, Spanish, and German. In order to search for the translations of each ontology label, the system accesses different lexical resources: 1) remote lexical databases as EuroWordNet, 2) multilingual dictionaries as IATE[38], and 3) translation services as GoogleTranslate[39], Wiktionary[40], Babelfish[41], and FreeTranslation[42]. A *cache* stores previous translations to avoid accessing the same data twice.

The algorithm used by the Translation Service can be summarized as follows: 1) If the selected ontology label is already available in the target language in our cache, then LabelTranslator just displays it, with all the relevant available information; 2) If the translation is not stored locally, then it accesses remote repositories to retrieve possible translations. A compositional method may be needed to translate compound labels (explained in **b) Compositional Method to Translate Compound Labels** below). If no results are obtained from the two previous steps, then the user can enter his/her own translation (together with a definition).

In our approach, the translation of an ontology label denoted by *t*, is a tuple ⟨*trs, senses*⟩, where *trs* is a translated label in a specific target language, and *senses* is a list of semantic senses extracted from different knowledge pools. In the following we briefly describe the task of automatically retrieving the possible semantic senses of a translated label.

a) Semantically Representing a Sense

In order to discover the senses of each translated label (t_i), we have considered the approach proposed in a previous work [31]. Our system takes as input a list of words or labels (each t_i), it checks their meaning in run-time, and obtains a list of senses extracted from different ontology pools; next, it deals with the possible semantic overlapping among senses. We summarize here the key characteristics of the sense discovering process:

---

[38] http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load

[39] http://www.google.com/translate_t

[40] http://en.wiktionary.org/wiki/

[41] http://babelfish.altavista.com/

[42] http://ets.freetranslation.com

1. To discover the semantics, i.e. the meaning, of the input words, the system relies on a pool of ontologies instead of just a single ontology.
2. The system builds a sense (meaning) with the information retrieved from matching terms in the ontology pool.
3. Each sense is represented as a tuple $s_k = \langle s, grph, descr \rangle$, where *s* is the list of synonyms[43] of a word *k*, *grph* describes the sense s_k by means of the hierarchical graph of hypernyms and hyponyms of synonym terms found in one or more ontologies, and *descr* is a description in natural language of such a sense.
4. As matching terms could be ontology classes, properties or individuals, three lists of possible senses are associated with each word *k*: S_k^{class}, S_k^{prop} and S_k^{indv}
5. Each word sense is enhanced incrementally with synonym senses (which also searches in the ontology pool).
6. A sense alignment process integrates the word sense with those synonym senses representing the same semantics.

A more detailed description of this process can be found in [31]. In order to perform cross-language sense translations, external resources are limited to those resources that have multilingual information like EuroWordNet. However, other resources can be also used. For example, the AGROVOC[44] Thesaurus of the Food and Agricultural Organization, which could cover the vocabulary missing in EuroWordNet. The multilingual retrieval of a word sense (synset) in EuroWordNet[45] is done by means of the InterlingualIndex (ILI) that serves as a link among the different wordnets. For example, when a synset, e.g. *chair,* with the meaning *professor position*, is retrieved from the English wordnet, its synset ID is mapped through the ILI to the synsets IDs of the same concept in the different language-dependent wordnets,(German, Spanish, etc.) that describe the same concept, but naturally contain the word description in its specific language. A similar retrieval process is used in the case of multilingual ontologies, but using the references between concepts and labels as offered by the standard `owl:comment` and `rdfs:label` properties.

Coming back to the example in section 5.4, in Figure 26 we show the translations of the ontology label *chair* from English into Spanish. Our system finds eight translations, but we only show three. Notice that t_3 has the desired semantics according to the similarity with the lexical and semantic ontology context (see Figure 25 in section 5.4).

[43] The system extracts the synonyms of a term by consulting the synonym relationships defined in the ontology of such a term.

[44] http://www.fao.org/aims/ag download.htm

[45] The EuroWordNet lexicon was also analysed in detail in D2.4.1, section 9 [18]

2006–2008 © Copyright lies with the respective authors and their institutions.

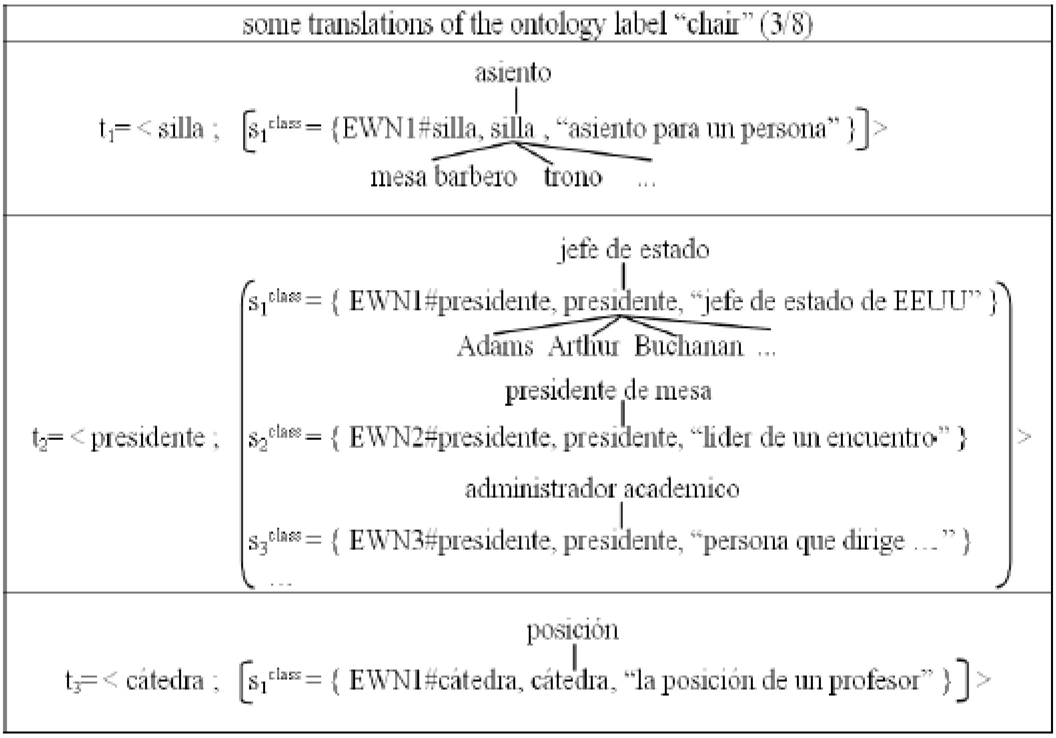


Figure 26. Some translations of the ontology label "chair" into Spanish.

b) Compositional Method to Translate Compound Labels

Compound labels which have an entry in linguistic ontologies such as EuroWordNet (for example, *jet lag*, *travel agent* and *bed and breakfast*) are treated in our system as single words. Others like *railroad transportation*, which have no entry in the previous resources as they are not fully lexicalized, are translated by using a compositional method. This method splits the label into its corresponding tokens (*railroad* and *transportation* in the example). The individual components are translated and then combined into a compound label in the target language. Care is taken to combine the components maintaining the word order of the target language. A set of lexical templates derived from different ontologies are used to control the order of translation. The main steps of the algorithm are the following:

1. The compound label is normalized, e.g., rewriting in lowercase, hyphens are removed, it is split into tokens, etc.
2. A set of possible translations is obtained for each token of the compound label using the translation service.
3. Since translations between languages do not keep the same word order, the algorithm creates candidate translations in the target language using lexical templates[46]. Each lexical template contains at least a pair of patterns, namely 'source' and 'target' patterns. A source pattern is a template to be compared with the *tagged compound label*[47], described in the source language, while the target pattern is used to generate the label in the target language. If no applicable template is found, the compound label is translated using the translation service directly.
4. All candidate labels that fulfill the target pattern requirements are returned as candidate translations of the compound label.

---

[46] The notion of lexical template proposed in this paper refers to text correlations found between a pair of languages.

[47] We use TreeTagger [3] in order to annotate the compound labels with part-of-speech and lemma information.

Senses of each candidate translation are discovered by using the sense discovering process described in **a) Semantically Representing a Sense**. If no results are obtained, the method tries to discover the senses of each token separately.

In the following we describe the process to learn the lexical templates used to control the order of translation of compound labels.

c) Learning Lexical Templates from Ontological Labels

The lexical templates used to translate compound labels are necessary to produce high quality translations because (1) they guarantee grammatical output and, (2) they make sure that the structural source language meaning is preserved. In our approach, we used a semi-automatic process to obtain lexical templates. As explained before, each lexical template is composed of source and target patterns. On the one hand, the ontology labels used to learn the source patterns were extracted from different domain ontologies expressed in English, German, or Spanish. Each label was tokenized and tagged using the language independent part-of-speech tagger proposed in [29]. On the other hand, the labels used to learn the target patterns were extracted either from the multilingual information associated with each ontological term or by means of a manual translation process. The same process used for part-of-speech annotation of the labels in the source patterns was also used to annotate the labels of the target patterns. The empirical results collected during the learning of lexical templates are briefly described below:

- *Existing ontologies share the same lexical patterns*. For instance, approximately 60% of the labels that describe an ontological concept include an adjective followed by a noun (e.g. spatial region, industrial product, natural hazard, etc.). Other labels have as lexical pattern (≈30%) a noun followed by another noun (e.g., transport vehicle, knowledge domain, etc.).
- *Ontology labels usually have less than four tokens*. Approximately 85% of the labels fulfill this requirement. Thus, for the current prototype we only focus on the definition of lexical templates for compound labels of two or three tokens.

A repository is used to store all the lexical templates obtained for each pair of languages. In Table 1 we show only a sample list of the lexical templates learned to translate compound labels from English into Spanish.

Table 1. Some lexical templates to translate compound labels from English into Spanish

Templates (4/25)	*Samples of source and target patterns*	
	English	*Spanish*
[J1 N2]en→[N2 J1]es	spatial region→ industrial product→ natural hazard→	región especial producto industrial peligro natural
[N1 N2]en→[N2 ⟨pre⟩ N1]es	transport vehicle→ knowledge domain→ research exploration→	vehiculo de transporte dominio del conocimiento exploración de la investigación
[J1 VB2]en→[VB2 ⟨pre⟩ J1]es	remote sensing→	detección remota; detección a distancia

2006–2008 © Copyright lies with the respective authors and their institutions.

[J1 N2 N3]en→[N2 〈pre〉 N3 J1]es	associated knowledge domain→	dominio de conocimiento asociado

J: adjective; N: noun; VB: verb

By way of illustration of the compositional method, in Figure 27 we show the steps of the algorithm when collecting Spanish translations for the English compound label *AssociateProfessor*, given in our example (see Figure 25). Our system finds ten translations for the token *associate* and one for *professor* (normalized in the first step). In the next step, our tool searches a lexical template (in our repository) to create candidate translations. In the template found, *[J1 N2]en* represents the source pattern in English whilst *[N2 J1]es* represents the target pattern in Spanish. In both cases, numbers represent the position of each token of the compound label. Notice that, in the last step, the candidate translations *profesor socio* (professor member) and *profesor compañero* (professor mat) are discarded because they do not fit the target pattern.

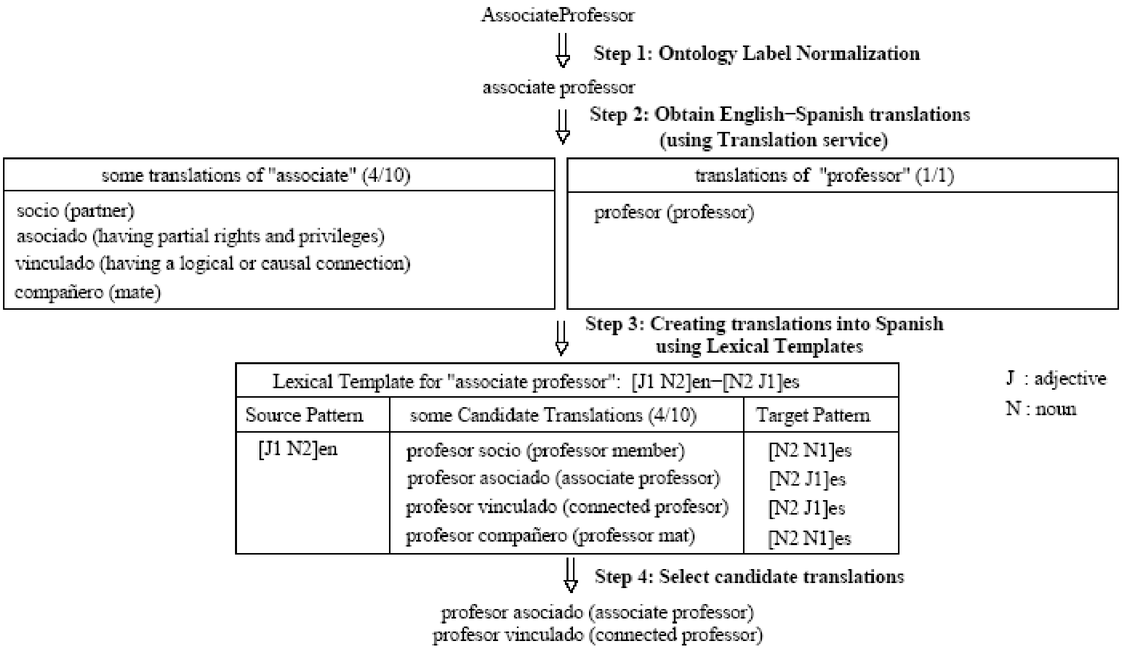


Figure 27. Algorithm to translate the compound label *AssociateProfessor* into Spanish.

6.4.2 Translation Ranking Method

In this section we explain the *ranking method*, which sorts out the list of translations according to similarity with the context of the label to be translated. The method takes as input the set of translations *T* obtained in the previous step. From this set of translations, the ranking method uses a disambiguation algorithm described in [23] to sort the translations. Once all translations are ranked, the method allows two operation modes:

- Semi-automatic mode: It shows a list with all possible translations sorted out decreasingly. The method proposes the most relevant translation to be selected first although the user can change this default selection.
- Automatic mode: It automatically selects the translation with the highest score.

Next, we first describe how the system obtains the context of each ontology label, and then we explain the disambiguation algorithm used to sort the translations according to similarity with their context.

a) Determining the Context of an Ontology Term

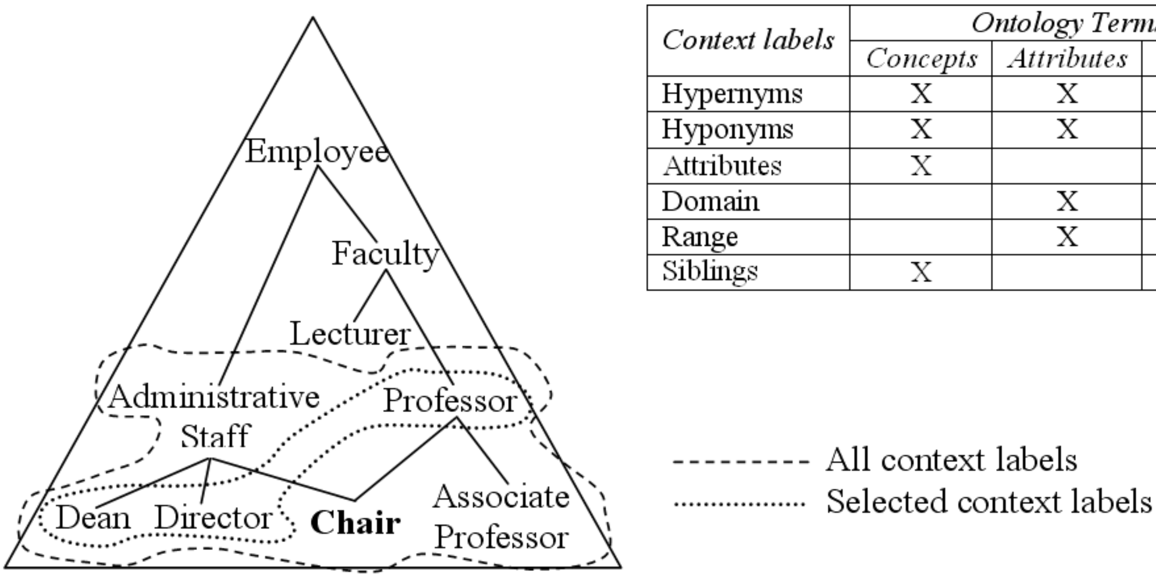
For the purposes of this research, we understand context in its more general sense of being "the part of a discourse that surrounds a word or passage and can throw light on its meaning"[48]. More specifically, in D3.1.1 [10] of this project, and regarding ontology networks, context has been defined as "a set of all circumstances, properties and facts within which the ontology has the desired semantics". In our approach, we could substitute the word "ontology" by "ontology term" in the previous definition, and it would also be valid. Hence, the context of an ontology term is used to disambiguate its lexical meaning. To determine the context of an ontology term, the system retrieves the labels of the set of terms associated with the term under consideration. The list of context labels, denoted by *C*, comprises a set of names which can be direct label names and/or attributes label names, depending on the type of term that is being translated.

In order to mitigate risks associated with system performance, the *ranking method* limits the number of context labels used to disambiguate the translated label. Every context label $c \in C$ is compared with the ontology label *l* using a measure based on Normalized Google Distance [4] (NGD). NGD measures the semantic relatedness between any two terms, considering the relative frequency in which two terms appear in the Web within the same documents. Those labels with the higher values of similarity are chosen (maximum 3). To discover the senses of each context label (denoted by *Sc*), the system performs the same process used to discover the senses of each translated label (as explained in the previous section).

In Figure 28, on the left, the dashed area represents all the context labels found for the ontology label *chair*. Our prototype finds five labels, but it only selects three (see the dotted area) to disambiguate the term. In the table on the right, we show the context labels that could be extracted for each type of the ontology terms (concept, attribute, or relation). For instance, for the concept *chair* the system retrieves its hypernyms, hyponyms, attributes, and sibling concepts.

[48] http://www.merriam-webster.com/dictionary/context

2006–2008 © Copyright lies with the respective authors and their institutions.



Context labels	Ontology Terms		
	Concepts	Attributes	Relations
Hypernyms	X	X	
Hyponyms	X	X	
Attributes	X		
Domain		X	X
Range		X	X
Siblings	X		

Figure 28. Context of the ontology label *chair*

b) Disambiguating Senses of Translations

In some works [23,21] glosses are considered as a very promising means of measuring relatedness, since they can be used: (1) to make comparisons between concepts semantically different, and (2) to discover relations of which no trace is present in the resource they come from. In this version of the system, the ranking method relies on a measure based on glosses proposed in [23] to sort out the translations according to their context. However, we recognize that glosses are necessarily short and may not provide sufficient information on their own to make judgments about relatedness. Therefore, we make use of the hierarchical graph of the sense to extend the gloss with the relatedness glosses of their ontological terms.

We carry out disambiguation in relation to the senses of each translated label and the senses of the context labels. In the following we describe the method: let us suppose that the ontology label *l*, after executing the translation process, has yielded *n* translations: $T = \{t_1, t_2, ..., t_n\}$. For each translation the system retrieves its corresponding senses, for example the first translated label (t_1) to be disambiguated has *n* senses $S_{t1} = \{s_{t1}^1, s_{t1}^2, ..., s_{t1}^n\}$. We use the notation *TSC* (translation sense collection) in order to group the senses of all translated labels.

$$TSC = \{S_{t1} \cup S_{t2} \cup ... \cup S_{tn}\}$$

where S_{t_j}, $tj \in T$, represents all senses corresponding to j^{th} translated label.

Now, suppose that the ontology label *l* has the context *C* which comprises several labels: c_1, c_2, c_3. Each of these context labels has a list of corresponding senses, for instance, c_i has *m* senses: $S_{ci} = \{s_{ci}^1, s_{ci}^2, ..., s_{cm}^n\}$. We use the notation *CSC* (context sense collection) in order to group the senses of each context label.

$$CSC = \{S_{c1} \cup S_{c2} \cup ... \cup S_{c3}\}$$

where S_{c_j}, $cj \in C$, represents all senses corresponding to j^{th} context label.

The goal of the disambiguation algorithm is to select one of the senses from the set *TSC* as the most appropriate sense of the translation of label *l*. The algorithm performs word sense disambiguation by using a measure of semantic relatedness that is given by:

$$\max_{j=1}^{|TSC|}\left(SenseScore\left(TSC_j, CSC\right)\right)$$

where TSC_j is the representation of one of the senses of each translated label. The selected sense is the one with the greater value of SenseScore, defined as:

$$SenseScore\left(TSC_j, CSC\right) = \Sigma_{k=1}^{|CSC|}\left(Similarity\left(TSC_j, CSC_k\right)\right)$$

where CSC_k is the representation of each sense of the different context labels.

In order to compute the similarity between the senses of each context and the translated label, the method applies an overlap scoring mechanism. Details about this process are available in [23].

In our example, *cátedra* (chair or professorship) in the sense of the *position of professor* is ranked as first translation of the ontology label *chair*. Once the right sense has been selected, the system updates the linguistic information of the corresponding ontological term.

6.5 Repository Component

The modular approach followed by the second version of LabelTranslator relies on the combination of two independent modules, the ontological and the linguistic one. In order to keep both models, the ontology model (OM) and lexical model (LM), *synchronized* we first need to find out exactly what has been changed in the ontology model, then find the equivalent places in the linguistic model and only then start the updating.

In this section we present our solution to the problems of managing changes in the ontology model and we state how to propagate those changes to the linguistic model which manages information in multiple languages. By processing the original ontology, the original translated version and the updated ontology label(s), we show that it is much easier to deal with the re-translation process.

Main Problems

- The first problem encountered when translating updates of a source ontology is how to identify where changes have been made. Even if change tracking facilities are used to update the ontology, identifying where each change has occurred, and ensuring that a matching change is made to the linguistic model, is a task that can be prone to error.

- The second problem is to minimize re-translation of already translated labels. First, our system needs to identify the term in the ontology model where the change occurred. Then, it needs to ensure that any altered label in the ontology is accurately replaced by the relevant translation.

Addition of new terms in the ontology, or deletion of an existing term can be controlled by some mechanism of change synchronization. In the NeOn ToolKit, an advanced change tracking, based on Resource Delta[49], is able to capture changes even when ontological terms have changed their position within the ontology model. By adopting this feature, our system can accurately identify the

[49] A resource delta represents changes in the state of a resource tree between two discrete points in time.

2006–2008 © Copyright lies with the respective authors and their institutions.

minimal set of changes needed to adjust the structure of the linguistic model, a critical first step to ensure that a matching change is made in the localized ontology. To correctly update the linguistic model, the system needs to identify:

1. all ontology terms in the original ontology whose labels have changed in the updated ontology.
2. any ontology term that has been added to the updated ontology.
3. any ontology term which has been removed from the original ontology.
4. any ontology term whose position in the updated ontology differs from that in the original ontology.

Finding where a translation is required is only part of the problem. We also need to ensure that changes in the ontology structure are accurately propagated to the linguistic information. This requires that elements whose structure need to be updated are clearly flagged in the linguistic model, and that the relevant structural changes are indicated in a form that turns updating the translation into a simple process, thus involving minimal work on the part of the linguist user or domain expert.

6.5.1 Simplifying Localization Management by means of Synchronization

LabelTranslator provides a model where sets of ontology terms and linguistic information associated (in different languages) are separately stored. Therefore, it would be very difficult for a person to update all the linguistic definitions and information associated with a particular concept. We believe that this process will be done by different people at different times and in different countries. Thus, the maintenance cycle for each language will often be separated. Figure 29 illustrates the localization management used in our system to synchronize the conceptual and linguistic information. In the following we analyze the process in more detail, describing the actions performed by each actor of our scenario.

- *Ontology expert.* (S)he is responsible for editing the changes in the ontology model. All the changes executed in each user session are stored in a repository as a new version. The types of changes that our system can manage are the following: changes of the label content (e.g., ontology label rename) and ontology structure changes (e.g., delete or add operations). For each case, LabelTranslator stores the type of operation executed and its additional information (e.g., the name of the renamed label). This information is used in our system to synchronize the conceptual and linguistic information.

- *Linguist expert(s).* The linguist expert in a specific target language is responsible for performing the localization process. Notice that this process always uses the last version of an ontology. When the linguist needs to update the linguistic model (LM), our system tries to synchronize both models, performing the following actions: (1) obtaining the current version of the LM to be updated, (2) extracting the last version of the changes in the ontology model (OM) from which the last localization was taken (normally the one with the same number as the LM), (3) performing all the actions of the file of changes in the LM, and (4) updating the LM version in the repository.

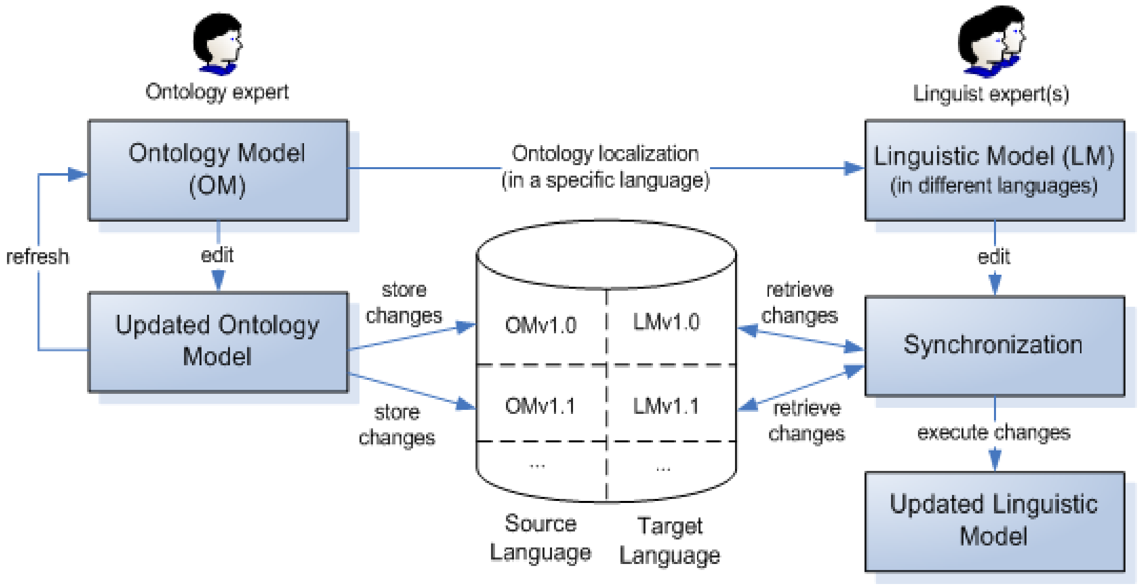


Figure 29. Synchronization of ontology and linguistic model

7. Conclusions

Taking into account the importance given to multilinguality within the NeOn project, as well as its relevance in the development of the Semantic Web of the future, we have, in this deliverable, concluded the task initiated in D2.4.1 [18] to propose (1) a model for providing multilinguality to NeOn ontologies, the *Linguistic Information Repository* (LIR), and (2) the NeOn plug-in that will support it, *LabelTranslator*. After having introduced preliminary versions in D2.4.1, in the present contribution we have presented enhanced versions of both the LIR and the LabelTranslator NeOn plug-in.

In the first place, we have described the different options we have identified for modelling multilinguality in ontologies, and have discussed the suitability of associating the ontology meta-model to an external multilingual linguistic model, which is the approach followed by the LIR model. This approach implies that the Ontology Localization Activity is mainly carried out at the ontology terminological layer, although it also allows localization at its conceptual layer by means of ontology modules, for example, if necessary.

The second major contribution of this deliverable has been a detailed description of the LIR classes and relations. Initial evaluation results have shown that the LIR model covers the needs expressed by the FAO, which was one of the main priorities of this task. We also plan to run some tests with resources from the Pharmaceutical Industry, in order to check if the LIR also covers all their needs. However, being aware that further refinements to the LIR could be required in order to improve more complex operations with linguistic data, we have discussed some possible enhancements that will be considered in future work.

Thirdly, we have focused on a complete description of the three main components of the LabelTranslator NeOn plug-in, namely, the GUI component, the Ontology Localization component and the Repository component. A great emphasis has been given to the benefits of the translation algorithm used by this second version of the LabelTranslator plug-in, which allows an automatic localization of ontology terms.

For the final deliverable of this task, D2.4.3, we will concentrate on the embedding of the LIR within a network of ontologies that capture linguistic, terminological and ontological knowledge. This will

2006–2008 © Copyright lies with the respective authors and their institutions.

allow us to refine the knowledge covered by the LIR, and systematically relate knowledge from different domains at various levels of descriptive granularity.

The next release of the LabelTranslator NeOn plug-in is scheduled for November 2008. Then, we will start the process of evaluating the algorithm used by LabelTranslator, which tries to select the most appropriate translation for each ontology label. In particular, we will evaluate two aspects of the algorithm: the obtained output using the automatic operation mode and the quality of the translation. We will define different scenarios to perform ontology localization using ontologies from multiple domains. Additionally, we will check the impact of the resources used during the process of localization. Results from this evaluation process are expected to be included in D5.6.2.

References

1. J. Barrasa. *Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales*. PhD Thesis. Universidad Politécnica de Madrid, Madrid, Spain, 2007.
2. P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. *Contextualizing Ontologies*. Journal of Web Semantics 1 (4), 2004. Available at: http://www.websemanticsjournal.org/papers/20041031/document4.pdf.
3. P. Buitelaar, M. Sintek, M. Kiesel. *A Multilingual/Multimedia Lexicon Model for Ontologies*. In Proceedings of the ESWC06 in Budva, Montenegro. 2006.
4. R. L. Cilibrasi and P. M. Vitnyi. *The Google Similarity Distance*. IEEE Transactions on Knowledge and Data Engineering, 19(3):370-383, March 2007.
5. P. Cimiano, P. Haase, M. Herold, M. Mantel, and P. Buitelaar. *LexOnto: A Model for Ontology Lexicons for Ontology-based NLP*. In Proceedings of the OntoLex07 Workshop, co-located at the ISWC 2007 Conference in Busan, South Corea. 2007.
6. M. Espinoza, A. Gómez-Pérez, and E. Mena. *Enriching an Ontology with Multilingual Information*. In Proceedings of 5th European Semantic Web Conference (ESWC'08), Tenerife (Spain), Springer Verlag LNCS, ISBN 978-3-540-68233-2, ISSN-0302-9743, pp. 333-347, June 2008.
7. M. Espinoza, A. Gómez-Pérez and E. Mena, LabelTranslator - A Tool to Automatically Localize an Ontology. In Proceedings of 5th European Semantic Web Conference (ESWC'08), Tenerife (Spain), Springer Verlag LNCS, ISBN 978-3-540-68233-2, ISSN-0302-9743, pp. 792-796, June 2008. demo paper.
8. J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag Berlin Heidelberg, 2007.
9. C. Fellbaum, J Grabowski, and S. Landes. “Performance and confidence in a semantic annotation task”. In C. Fellbaum (ed.) *WordNet: An Electronic Lexical Database*. Cambridge: The MIT Press. 1998.
10. P. Haase, P. Hitzler, S. Rudolph, G. Qi, M. Grobelnik, I. Mozetič, D. Bojadžiev (cords.), J. Euzenat, M. d’Aquin, A. Gangemi, and C. Catenacci. *NeOn Deliverable D3.1.1 Context Languages-State of the Art*. NeOn Project http://www.neon-project.org. September 2006.
11. G. Hirst. Ontology and the Lexicon. In S. Staab and R. Studer (eds.) *Handbook on Ontologies and Information Systems*. Springer, Berlin. 2004.
12. N. Ide, L. Romary. *A Registry of Standard Data Categories for Linguistic Annotation*. Proceedings of the 4th Language Resources and Evaluation Conference (LREC 2004), Lisbon, 2004.

13. M. Kemps-Snijders, M. Windhouwer, P. Wittenburg, and S.E. Wright. *ISOcat: Corralling Data Categories in the Wild*. Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008) Marrakech (Morocco). May 2008.
14. H. Kitakami, Y. Mori, M. Arikawa. *An Intelligent System for Integrating Autonomous Nomenclature Databases in Semantic Heterogeneity*. Database and Expert System Applications, DEXA'96, Zürich, Switzerland, pp. 187-196. 1996.
15. A.C. Liang, B. Lauser, M. Sini, J. Keizer, S. Katz. *From AGROVOC to the Agricultural Ontology Service / Concep Server. An OWL model for managing ontologies in the agricultural domain*. 2008
16. G. Miller, R. Beckwith, C. Fellbaum, D. Gross, K.J. Miller. "Introduction to WordNet: an On-line Lexical Database" *International Journal of Lexicography*, 3 (4), 235-244. 1990 (Revised in 1993).
17. E. Montiel-Ponsoda, G. Aguado de Cea, A. Gómez-Pérez, and W. Peters. *Modelling Multilinguality in ontologies*. In Proceedings of the Coling08 Conference, in Manchester, UK. August 2008.
18. E. Montiel-Ponsoda, W. Peters, G. Aguado de Cea, M. Espinoza (coords.), I. Álvarez de Mon, A. Gangemi, A. Gómez-Pérez, Ó. Muñoz, R. Palma, J.A. Ramos Gargantilla, M. Sini, M.C. Suárez-Figueroa. *NeOn Deliverable D2.4.1. Multilingual Ontology Support*. NeOn Project http://www.neon-project.org. August 2007.
19. C.W. Morris. *Foundations of the Theory of Signs*, International Encyclopedia of Unified Science, vol. 1, no 2, Chicago. 1938.
20. C.K. Ogden and I.A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul. 1923.
21. M.T. Pazienza and A. Stellato. *Clustering of terms from translation dictionaries and synonyms lists to automatically build more structured Linguistic Resources*. *6th Language Resources and Evaluation Conference* (LREC 2008). Marrakech (Morocco). May 2008.
22. M.T. Pazienza and A. Stellato. *Exploiting linguistic resources for building linguistically motivated ontologies in the semantic web*. In Second Workshop onInterfacing Ontologies and Lexical Resources for Semantic Web Technologies (OntoLex2006), co-located at LREC2006, May 24-26, 2006, Genoa, (Italy), 2006.
23. T. Pedersen, S. Banerjee, and S. Patwardhan. *Maximizing Semantic Relatedness to Perform Word Sense Disambiguation*. Research Report UMSI 2005/25, University of Minnesota Supercomputing Institute, March 2005.
24. W. Peters, E. Montiel-Ponsoda, G. Aguado de Cea. *Localizing Ontologies in OWL*. In Proceedings of the OntoLex07 Workshop, co-located at the ISWC 2007 Conference in Busan, South Corea. 2007.
25. J. Pustejovsky. *The Generative Lexicon*. The MIT Press: Cambridge. 1995.
26. Z. Saloni, S. Szpakowicz, and M. Swidzinski. "The Design of a Universal Basic Dictionary of Contemporary Polish", *International Journal of Lexicography*, vol. 3 no. 1, Oxford University Press. 1990.
27. F. Scharffe, J. Euzenat, and D. Fensel. *Towards design patterns for ontology alignment*. In R.L. Wainwright and H. Haddad (eds.): Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Fortaleza, Ceara, Brazil, March 2008: 2321-2325.
28. M.C. Suárez-Figueroa, A. Gómez-Pérez. *First Attempt towards a Standard Glossary of Ontology Engineering Terminology*. 8th International Conference on Terminology and Knowledge Engineering (TKE2008), Copenhagen, August 2008.

2006–2008 © Copyright lies with the respective authors and their institutions.

29. M.C. Suárez-Figueroa, A. Gómez-Pérez. *Towards a Glossary of Activities in the Ontology Engineering Field* 6th Language Resources and Evaluation Conference (LREC 2008). Marrakech (Morocco). May 2008.

30. TreeTagger. 1997. http://www.ims.uni-stuttgart.de/projekte/corplex/

31. R. Trillo, J. Gracia, M. Espinoza, and E. Mena. “Discovering the semantics of user keywords”. *Journal on Universal Computer Science. Special Issue: Ontologies and their Applications, ISSN 0948-695X*, 2007.

32. P. Vossen. “EuroWordNet”: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index”. *Semi-special issue on multilingual databases*. IJL 17/2. June 2004.

Annex 1

OWL Version of the LIR model

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:p1="http://owlodm.ontoware.org/OWL1.0#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:FAOlang="http://www.fao.org/aims/aos/languagecode.owl#"
    xmlns:owl11="http://www.w3.org/2006/12/owl11#"
    xmlns:owl11xml="http://www.w3.org/2006/12/owl11-xml#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns="http://gate.ac.uk/gate-extras/neon/ontologies/lir1.7.owl#"
  xml:base="http://gate.ac.uk/gate-extras/neon/ontologies/lir1.7.owl">
  <owl:Ontology rdf:about="">
    <rdfs:comment xml:lang="en">The owl version of the Linguistic Information
Repository model defined in NeON WP2 to handle multilingual and other linguistic
information associated with ontology element labels.</rdfs:comment>
    <owl:versionInfo xml:lang="en">1.6</owl:versionInfo>
  </owl:Ontology>
  <owl:Class rdf:about="http://owlodm.ontoware.org/OWL1.0#OntologyElement">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasLexicalEntry"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Note">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasSource"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="belongsToLanguage"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isNoteOf"/>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
        </owl:onProperty>
        <owl:cardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:cardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">Supplemental information pertaining to any other
element in the ontology.
Use for any kind of note, including usage
notes, explanations, internal instructions, and
so forth. (see also ISO12620: section 08)</rdfs:comment>
  </owl:Class>
  <owl:Class rdf:about="http://www.fao.org/aims/aos/languagecode.owl#ISO639-1">
    <rdfs:subClassOf>
      <owl:Class
rdf:about="http://www.fao.org/aims/aos/languagecode.owl#languageCode"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Definition">
    <rdfs:comment xml:lang="en">A statement that describes a concept and permits
its differentiation from other concepts within a system of concepts. (ISO12620:
section 05.01)</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasSource"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasNote"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:cardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:cardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isDefinitionOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="UsageContext">
    <rdfs:comment xml:lang="en">A text or part of a text in which a term occurs.
```

```
http://www.lisa.org/Term-Base-eXchange.32.0.html</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasNote"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isContextOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasSource"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.fao.org/aims/aos/languagecode.owl#ISO639-2">
    <rdfs:subClassOf>
      <owl:Class
rdf:about="http://www.fao.org/aims/aos/languagecode.owl#languageCode"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Lexicalization">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasAcronym"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isDialectalVariantOf"/>
        </owl:onProperty>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isLogicalExpressionOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasFormula"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isShortFormOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasLogicalExpression"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasSource"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">A word form</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasShortForm"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
```

```
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isLexicalizationOf"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isSymbolOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasDialectalVariant"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasAbbreviation"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isAcronymOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasNote"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"></rdfs:label>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasSymbol"/>
        </owl:onProperty>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >0</owl:minCardinality>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="hasTransliteration"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="hasEquation"/>
      </owl:onProperty>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="isEquationFor"/>
      </owl:onProperty>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >1</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="isAbbreviationFor"/>
      </owl:onProperty>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >1</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >1</owl:minCardinality>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="isFormulaOf"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:SymmetricProperty rdf:ID="hasSpellingVariant"/>
      </owl:onProperty>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >0</owl:minCardinality>
      <owl:onProperty>
```

```
          <owl:ObjectProperty rdf:ID="hasContext"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:TransitiveProperty rdf:ID="hasVariant"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="LexicalEntry">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasAntonym"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"></rdfs:label>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasNote"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasCommonName"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isTranslationOf"/>
        </owl:onProperty>
      </owl:Restriction>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasLexicalization"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isLexicalEntryOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:FunctionalProperty rdf:ID="hasSense"/>
        </owl:onProperty>
        <owl:maxCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">A lexeme, which is an ordered collection of
related word forms, having the same lexical meaning.</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasSource"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasTranslation"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasScientificName"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
```

```
          <owl:SymmetricProperty rdf:ID="hasSynonym"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasContext"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class
rdf:about="http://www.fao.org/aims/aos/languagecode.owl#languageCode">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isCodeOf"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.fao.org/aims/aos/languagecode.owl#language">
    <rdfs:comment xml:lang="en">The language class imported from the FAO
languagecode ontology (http://www.fao.org/aims/aos/languagecode.owl).
In this ontology, each class is associated with ISO639 codes.</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasLinguisticExpression"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasLanguageCode"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
  </owl:Class>
  <owl:Class rdf:ID="Source">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasNote"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="isSourceOf"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">The provenance of the linguistic/terminological
information.</rdfs:comment>
  </owl:Class>
  <owl:Class rdf:ID="Sense">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="hasDefinition"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"></rdfs:label>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasNote"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:cardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:cardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#belongsToLanguage"/>
```

```
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#hasSource"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:SymmetricProperty rdf:ID="isRelatedTo"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">A language-specific unit of intensional lexical
semantic description.</rdfs:comment>
  </owl:Class>
  <owl:ObjectProperty rdf:about="#isSymbolOf">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasTransliteration">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:ID="isTransliterationOf"/>
    </owl:inverseOf>
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasLogicalExpression">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#isLogicalExpressionOf"/>
    </owl:inverseOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isSourceOf">
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#Note"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#Sense"/>
          <owl:Class rdf:about="#UsageContext"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
    <rdfs:domain rdf:resource="#Source"/>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasSource"/>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
    </owl:inverseOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#belongsToLanguage">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasLinguisticExpression"/>
    </owl:inverseOf>
    <rdfs:range
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#language"/>
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#UsageContext"/>
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#Note"/>
          <owl:Class rdf:about="#Sense"/>
          <owl:Class rdf:about="#Source"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isTranslationOf">
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasTranslation"/>
    </owl:inverseOf>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    ></rdfs:comment>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isTransliterationOf">
    <owl:inverseOf rdf:resource="#hasTransliteration"/>
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasShortForm">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#isShortFormOf"/>
    </owl:inverseOf>
    <rdfs:subPropertyOf>
      <owl:ObjectProperty rdf:about="#hasAbbreviation"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isEquationFor">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasEquation"/>
    </owl:inverseOf>
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasFormula">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#isFormulaOf"/>
    </owl:inverseOf>
  </owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#hasContext">
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="#isContextOf"/>
  </owl:inverseOf>
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Lexicalization"/>
        <owl:Class rdf:about="#LexicalEntry"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:domain>
  <rdfs:range rdf:resource="#UsageContext"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isAbbreviationFor">
  <rdfs:subPropertyOf>
    <owl:TransitiveProperty rdf:about="#hasVariant"/>
  </rdfs:subPropertyOf>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="#hasAbbreviation"/>
  </owl:inverseOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isShortFormOf">
  <rdfs:subPropertyOf rdf:resource="#isAbbreviationFor"/>
  <owl:inverseOf rdf:resource="#hasShortForm"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isLexicalizationOf">
  <rdfs:domain rdf:resource="#Lexicalization"/>
  <rdfs:range rdf:resource="#LexicalEntry"/>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="#hasLexicalization"/>
  </owl:inverseOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isContextOf">
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Lexicalization"/>
        <owl:Class rdf:about="#LexicalEntry"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:range>
  <rdfs:domain rdf:resource="#UsageContext"/>
  <owl:inverseOf rdf:resource="#hasContext"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isDialectalVariantOf">
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="#hasDialectalVariant"/>
  </owl:inverseOf>
  <rdfs:subPropertyOf>
    <owl:TransitiveProperty rdf:about="#hasVariant"/>
  </rdfs:subPropertyOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#isLogicalExpressionOf">
  <owl:inverseOf rdf:resource="#hasLogicalExpression"/>
  <rdfs:subPropertyOf>
    <owl:TransitiveProperty rdf:about="#hasVariant"/>
  </rdfs:subPropertyOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#hasSource">
  <rdfs:range rdf:resource="#Source"/>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#UsageContext"/>
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Note"/>
          <owl:Class rdf:about="#Sense"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <owl:inverseOf rdf:resource="#isSourceOf"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasLinguisticExpression">
    <rdfs:domain
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#language"/>
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#Note"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#Sense"/>
          <owl:Class rdf:about="#Source"/>
          <owl:Class rdf:about="#UsageContext"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
    <owl:inverseOf rdf:resource="#belongsToLanguage"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasLexicalEntry">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#isLexicalEntryOf"/>
    </owl:inverseOf>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:domain
rdf:resource="http://owlodm.ontoware.org/OWL1.0#OntologyElement"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasSymbol">
    <rdfs:domain rdf:resource="#Lexicalization"/>
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isDefinitionOf">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasDefinition"/>
    </owl:inverseOf>
    <rdfs:domain rdf:resource="#Definition"/>
    <rdfs:range rdf:resource="#Sense"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasAntonym">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:ID="isAntonymOf"/>
    </owl:inverseOf>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasCommonName">
```

```
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasScientificName"/>
    </owl:inverseOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasAbbreviation">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
    <owl:inverseOf rdf:resource="#isAbbreviationFor"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasEquation">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
    <rdfs:domain rdf:resource="#Lexicalization"/>
    <owl:inverseOf rdf:resource="#isEquationFor"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isCodeOf">
    <rdfs:domain
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#languageCode"/>
    <rdfs:range
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#language"/>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasLanguageCode"/>
    </owl:inverseOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasDialectalVariant">
    <owl:inverseOf rdf:resource="#isDialectalVariantOf"/>
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasTranslation">
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <owl:inverseOf rdf:resource="#isTranslationOf"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasDefinition">
    <owl:inverseOf rdf:resource="#isDefinitionOf"/>
    <rdfs:domain rdf:resource="#Sense"/>
    <rdfs:range rdf:resource="#Definition"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isNoteOf">
    <rdfs:domain rdf:resource="#Note"/>
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Sense"/>
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#UsageContext"/>
          <owl:Class rdf:about="#Source"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasNote"/>
    </owl:inverseOf>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isAcronymOf">
    <owl:inverseOf>
      <owl:ObjectProperty rdf:about="#hasAcronym"/>
    </owl:inverseOf>
    <rdfs:subPropertyOf rdf:resource="#isAbbreviationFor"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasLanguageCode">
    <owl:inverseOf rdf:resource="#isCodeOf"/>
    <rdfs:domain
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#language"/>
    <rdfs:range
rdf:resource="http://www.fao.org/aims/aos/languagecode.owl#languageCode"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isAntonymOf">
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <owl:inverseOf rdf:resource="#hasAntonym"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasNote">
    <rdfs:range rdf:resource="#Note"/>
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Lexicalization"/>
          <owl:Class rdf:about="#Definition"/>
          <owl:Class rdf:about="#LexicalEntry"/>
          <owl:Class rdf:about="#Sense"/>
          <owl:Class rdf:about="#Source"/>
          <owl:Class rdf:about="#UsageContext"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Object properties cannot be within the domain and range of other object
properties. They are therefore not connected to the Note class by means of the
hasNote property, and the supplemental information functionality expressed by
Note and hasNote is taken over by the rdfs:comment attribute.</rdfs:comment>
    <owl:inverseOf rdf:resource="#isNoteOf"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isFormulaOf">
    <rdfs:subPropertyOf>
      <owl:TransitiveProperty rdf:about="#hasVariant"/>
    </rdfs:subPropertyOf>
    <owl:inverseOf rdf:resource="#hasFormula"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#isLexicalEntryOf">
    <rdfs:range
rdf:resource="http://owlodm.ontoware.org/OWL1.0#OntologyElement"/>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <owl:inverseOf rdf:resource="#hasLexicalEntry"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasScientificName">
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >This relation is equivalent to the ISO16220 attribute "scientificName":
a term that is part of an international scientific nomenclature as adopted by an
appropriate scientific body. (ISO12620: section 02.01.04)</rdfs:comment>
    <owl:inverseOf rdf:resource="#hasCommonName"/>
  </owl:ObjectProperty>
```

```
  <owl:ObjectProperty rdf:about="#hasAcronym">
    <owl:inverseOf rdf:resource="#isAcronymOf"/>
    <rdfs:subPropertyOf rdf:resource="#hasAbbreviation"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="#hasLexicalization">
    <owl:inverseOf rdf:resource="#isLexicalizationOf"/>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#Lexicalization"/>
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID="text">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:ID="sourceType"/>
    </rdfs:subPropertyOf>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >e.g. a textual description of the resource, or maybe a unique key into the
resource specific information structure (for instance, in the case of a
dictionary, the composite key lemma, pos and sense number).</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="abbreviation">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:ID="termType"/>
    </rdfs:subPropertyOf>
    <rdfs:comment xml:lang="en">A term resulting from the omission of any part
of the full term while designating the same concept, e.g. adjective vs. adj.
(ISO12620: section 02.01.08)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="transliteration">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#termType"/>
    </rdfs:subPropertyOf>
    <rdfs:comment xml:lang="en">A form of a term resulting from an operation
whereby the characters of an alphabetic writing system are represented by
characters from another alphabetic writing system. (ISO12620: section
02.01.10)</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="grammaticalNumber">
    <rdfs:domain rdf:resource="#Lexicalization"/>
    <rdfs:range>
      <owl:DataRange>
        <owl:oneOf rdf:parseType="Resource">
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >singular</rdf:first>
          <rdf:rest rdf:parseType="Resource">
            <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >plural</rdf:first>
            <rdf:rest rdf:parseType="Resource">
              <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#nil"/>
              <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >other</rdf:first>
            </rdf:rest>
          </rdf:rest>
        </owl:oneOf>
      </owl:DataRange>
    </rdfs:range>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="multiWordExpression">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#termType"/>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
    </rdfs:subPropertyOf>
    <rdfs:comment xml:lang="en">This attribute is equivalent to ISO12620:
Phrase: A phraseological unit containing any group of two or more words that are
frequently expressed together and that comprise more than one concept. The
individual words in a phrase usually function in more than one grammatical
category (part of speech) within the syntax of a sentence.e.g. â€œwork
offlineâ€ ) (ISO12620: section 02.01.18)</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="#termType">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >a mix of types characterizing Lexicalizations, covering orthographic and
cultural aspects.</rdfs:comment>
    <rdfs:domain rdf:resource="#Lexicalization"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="noteText">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:domain rdf:resource="#Note"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="#sourceType">
    <rdfs:domain rdf:resource="#Source"/>
    <rdfs:comment xml:lang="en">The provenance of the linguistic/terminological
information.</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="scientificName">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:comment xml:lang="en">equivalent to ISO-12620
internationalScientificTerm</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="symbol">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:comment xml:lang="en">A designation of a concept by letters, numerals,
pictograms or any combination thereof. (ISO12620: section
02.01.13)</rdfs:comment>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="dialectalVariant">
    <rdfs:comment xml:lang="en">indicates whether a word form originates from a
dialect.</rdfs:comment>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="mainEntry">
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >The concept designation that has been chosen to head a terminological
record.) (ISO12620: section 02.01.01)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="belongsToDialect">
    <rdfs:domain rdf:resource="#Lexicalization"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment xml:lang="en">the dialect name to which the Lexicalization
belongs.</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="nameSpaceIdentifier">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >URL; URI (conform to ISO12620: section 10.21)</rdfs:comment>
    <rdfs:subPropertyOf rdf:resource="#sourceType"/>
  </owl:DatatypeProperty>
```

```
  <owl:DatatypeProperty rdf:ID="sourceIdentifier">
    <rdfs:comment xml:lang="en">The code assigned to a document in a
terminological collection and used as both the identifier for a bibliographic
entry and as a pointer in individual term entries to reference the bibliographic
entry identified with this code. (see ISO12620: section 10.20)</rdfs:comment>
    <rdfs:subPropertyOf rdf:resource="#sourceType"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="context">
    <rdfs:domain rdf:resource="#UsageContext"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="bibliographicReference">
    <rdfs:subPropertyOf rdf:resource="#sourceType"/>
    <rdfs:comment xml:lang="en">bibliographicReference: A complete citation of
the bibliographic information pertaining to a document or other resource.  (see
ISO12620: section 10.19)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="equation">
    <rdfs:comment xml:lang="en">An expression used to represent a concept based
on a statement that two mathematical expressions are, for instance, equal as
identified by the equal sign (=), or assigned to one another by a similar sign.
(ISO12620: section 02.01.15)</rdfs:comment>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="commonName">
    <rdfs:comment xml:lang="en">equivalent to ISO-12620
commonName</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="logicalExpression">
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:comment xml:lang="en">An expression used to represent a concept based
on mathematical or logical relations, such as statements of inequality, set
relationships, boolean operations, and the like. (ISO12620: section
02.01.16)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="gender">
    <rdfs:range>
      <owl:DataRange>
        <owl:oneOf rdf:parseType="Resource">
          <rdf:rest rdf:parseType="Resource">
            <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >feminine</rdf:first>
            <rdf:rest rdf:parseType="Resource">
              <rdf:rest rdf:parseType="Resource">
                <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                >other</rdf:first>
                <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#nil"/>
              </rdf:rest>
              <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >neuter</rdf:first>
            </rdf:rest>
          </rdf:rest>
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >masculine</rdf:first>
        </owl:oneOf>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
      </owl:DataRange>
    </rdfs:range>
    <rdfs:domain rdf:resource="#Lexicalization"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="formula">
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:comment xml:lang="en">Figures, symbols or the like used to express a
concept briefly, such as a mathematical or chemical formula (ISO12620: section
02.01.14)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="fullForm">
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:comment xml:lang="en">The complete representation of a term for which
there is an abbreviated form.</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="partOfSpeech">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >The grammatical class of the LexicalEntry (ISO12620)</rdfs:comment>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range>
      <owl:DataRange>
        <owl:oneOf rdf:parseType="Resource">
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >noun</rdf:first>
          <rdf:rest rdf:parseType="Resource">
            <rdf:rest rdf:parseType="Resource">
              <rdf:rest rdf:parseType="Resource">
                <rdf:rest rdf:parseType="Resource">
                  <rdf:rest rdf:parseType="Resource">
                    <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                    >other</rdf:first>
                    <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
                  </rdf:rest>
                  <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                  >proper noun</rdf:first>
                </rdf:rest>
                <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                >adverb</rdf:first>
              </rdf:rest>
              <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >adjective</rdf:first>
            </rdf:rest>
            <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >verb</rdf:first>
          </rdf:rest>
        </owl:oneOf>
      </owl:DataRange>
    </rdfs:range>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="shortForm">
    <rdfs:comment xml:lang="en">An abbreviated form that includes fewer words
than the full form.
e.g. â€œIntergovernmental Group of Twenty-four on International Monetary
Affairsâ€  vs. â€œGroup of Twenty-fourâ€ . . (ISO12620: section
02.01.08.02)</rdfs:comment>
```

```
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="acronym">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#boolean"/>
    <rdfs:subPropertyOf rdf:resource="#termType"/>
    <rdfs:comment xml:lang="en">An abbreviated form of a term made up of letters
from the full form of a multiword term strung together into a sequence
pronounced only syllabically. (ISO12620: section 02.01.08.04)</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:TransitiveProperty rdf:about="#hasVariant">
    <rdfs:domain rdf:resource="#Lexicalization"/>
    <rdfs:range rdf:resource="#Lexicalization"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#SymmetricProperty"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </owl:TransitiveProperty>
  <owl:SymmetricProperty rdf:about="#isRelatedTo">
    <rdfs:domain rdf:resource="#Sense"/>
    <rdfs:range rdf:resource="#Sense"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </owl:SymmetricProperty>
  <owl:SymmetricProperty rdf:ID="isSynonymOf">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
    <owl:inverseOf rdf:resource="#isSynonymOf"/>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
  </owl:SymmetricProperty>
  <owl:SymmetricProperty rdf:about="#hasSynonym">
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >equivalent to owl:sameAs</rdfs:comment>
    <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >TODO: </owl:versionInfo>
    <owl:inverseOf rdf:resource="#hasSynonym"/>
  </owl:SymmetricProperty>
  <owl:SymmetricProperty rdf:about="#hasSpellingVariant">
    <rdfs:subPropertyOf rdf:resource="#hasVariant"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </owl:SymmetricProperty>
  <owl:FunctionalProperty rdf:ID="definitionText">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
    <rdfs:domain rdf:resource="#Definition"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:FunctionalProperty>
  <owl:FunctionalProperty rdf:about="#hasSense">
    <owl:inverseOf>
      <owl:InverseFunctionalProperty rdf:ID="isSenseOf"/>
    </owl:inverseOf>
    <rdfs:domain rdf:resource="#LexicalEntry"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
    <rdfs:range rdf:resource="#Sense"/>
  </owl:FunctionalProperty>
  <owl:InverseFunctionalProperty rdf:about="#isSenseOf">
    <owl:inverseOf rdf:resource="#hasSense"/>
    <rdfs:range rdf:resource="#LexicalEntry"/>
    <rdfs:domain rdf:resource="#Sense"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </owl:InverseFunctionalProperty>
```

2006–2008 © Copyright lies with the respective authors and their institutions.

```
  <rdf:Description rdf:about="http://gate.ac.uk/gate-
extras/neon/ontologies/lir1.4.owl#__deleted__">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    ></rdfs:comment>
  </rdf:Description>
  <owl:DataRange>
    <owl:oneOf rdf:parseType="Resource">
      <rdf:rest rdf:parseType="Resource">
        <rdf:rest rdf:parseType="Resource">
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >sourcecIdentifier</rdf:first>
          <rdf:rest rdf:parseType="Resource">
            <rdf:rest rdf:parseType="Resource">
              <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >text</rdf:first>
              <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#nil"/>
            </rdf:rest>
            <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >URL</rdf:first>
          </rdf:rest>
        </rdf:rest>
        <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
        >bibliographicReference</rdf:first>
      </rdf:rest>
      <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >nameSpaceIdentifier</rdf:first>
    </owl:oneOf>
  </owl:DataRange>
</rdf:RDF>
```