MULECO – Multilingual Upper-Level Electronic Commerce Ontology

Overview

This CEN/ISSS Electronic Commerce Workshop project will research the most efficient means of developing a multilingual upper-level ontology for describing and identifying the relationships between electronic commerce applications and the ontologies used to describe them. In particular it will investigate how information related to business processes can be integrated with existing techniques for classifying businesses, their products and services.

There are many existing and proposed "electronic commerce ontologies". The vast majority have been defined monolingually, or in at most three or four languages, often from the same language group. The problem is that different trading partners tend to use different ontologies, and tend to prefer ontologies developed in their native language or in a "neutral" language, which is often English. It is, therefore, difficult to identify points of overlap between ontologies, and it is also difficult for people to find relevant terms in ontologies using their native language.

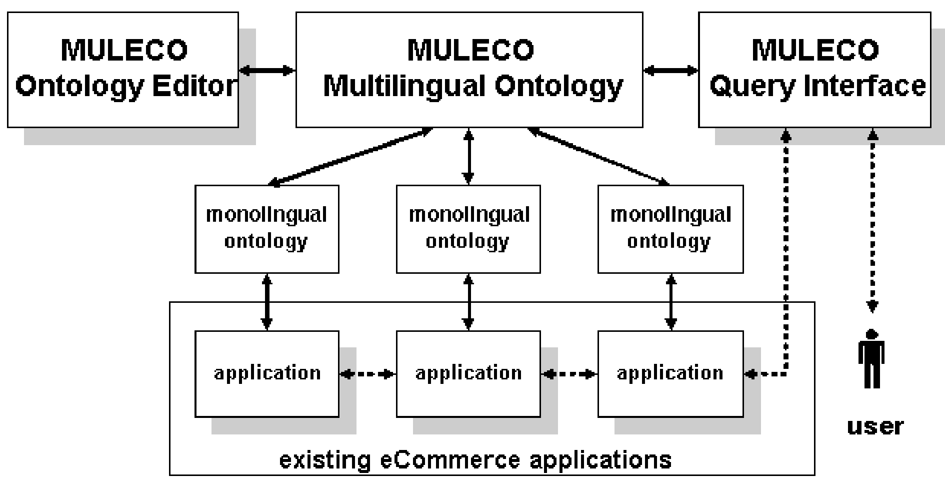


Figure 1: The relationship of MULECO to eCommerce Applications

The aim of MULECO is to develop a mechanism that will allow existing ontologies to identify their inter-relationships by identifying the relationships between themselves and a set of terms defined in a multilingual ontology that has been designed specifically to allow people to find terms using their native language. We realise that it is not possible, or desirable, to create and maintain a multilingual ontology that covers all terms used in all business applications in all European languages. What is needed is a way of classifying entries at the upper-most levels of existing ontologies in a form that takes account of the sort of terms used by people when they are trying to locate the term(s) they wish to use. To do this we need to extend existing business classification schemes to take account of things like business processes, variant names within different user communities, exclusion properties (e.g. no peanuts), etc. Such extensions need to be based on a well documented model that is based on properly researched linguistic characteristics, such as that provided by the Expert Advisory Group in Language Engineering Standards in The EAGLES Guidelines for Lexical Semantic Standards provided in Chapter 6 of *EAGLES LE3-4244: Preliminary Recommendations on Lexical Semantic Encoding -- Final Report* (http://www.ilc.pi.cnr.it/EAGLES96/EAGLESLE.PDF).

The MULECO project will develop an upper-level ontology, expressed as an extended network of industry descriptors, commercial terms and business roles, that will be recorded in a way that allows each entry to be addressed from other ontologies and applications by means of a Uniform Resource Identifier or an XML Path/Query. The project will also show how such an upper-level ontology can be used by open source tools to allow integration of ontologies and data dictionaries used within diverse electronic commerce applications.

The upper-level ontology will take as its start point existing standardized industry and process classification schemes, such as the International Standard for Industrial Classification (ISIC) used as the basis for the NACE classification of European business. The project will take note of the work being done by the IST CLAMOUR project to formally define such classification schemes. In particular it will extend currently used techniques for data classification, based on hierarchical classification of terms into broader and narrower meanings, by allowing for more complex relationships, in particular those relating to the relationships of wholes and parts which are vital to the mapping of the relationships between business processes. By defining a set of business relevant relationships between terms the project will allow classification hierarchies to become a controlled network of related words that forms an ontology rather than a classification scheme.

The ontology will be expressed in a language that provides the following functionality not currently found in electronic commerce ontologies based on languages such as RDF, OIL, KIF, etc, which is felt to be needed in order to model different kinds of relationships between multilingual electronic commerce ontologies:

1. The ability to uniquely identify the domain (e.g. industry sector) in which each term is employed
2. The ability to formally record the meaning of the term within a particular domain
3. The ability to identify other domains in which the same meaning applies
4. The ability to record alternative terms that have the same meaning within the original domain
5. The ability to identify alternative terms used for the same meaning in other domains
6. The ability to identify an exactly equivalent term used in a different language
7. The ability to identify a nearly equivalent term used in a different language
8. The ability to identify terms that form a part of an object defined by a term
9. The ability to identify wholes that a term forms a part of
10. The ability to identify an opposite term or property (e.g. water-resistant/water-soluble)
11. The ability to record relationships between terms or properties
12. The ability to identify opposite relationships (e.g. isMother/isChild)
13. The ability to declare properties that record measurements
14. The ability to declare properties that record times

15. The ability to associate terms with specific points in process chains

Monolingual ontologies that are linked to the multilingual ontology will be able to make use of equivalences expressed in the multilingual ontology to extend their search potential. This will allow companies that have developed electronic commerce applications for a single country/language to extend their applications to other European countries and beyond without having to change their underlying data dictionaries. With the forthcoming extension of the European Single Market into Eastern Europe and the Mediterranean there will be an increasing need for tools that allow the creation and maintenance of complex multilingual business ontologies of the type to be developed by this project. The project will evaluate the problems associated with developing multilingual ontologies, methodologies and techniques for overcoming them and the advantages to be gained from their use.

This project will incorporate and build on the concepts currently being developed to introduce monolingual ontologies into the Semantic Web. It will introduce such concepts into electronic commerce applications that are aimed at improving the flow of information between businesses within different language communities. At present most of the development work on the Semantic Web is postulated on the basis of using English language terms to identify the relationship between web resources and ontologies. Existing tools for applying the World Wide Web Consortium (W3C)'s Resource Description Framework (RDF) to the identification of related resources are generally postulated on the manual indexation of resources. Business applications require that this work be automated so that resource relationships can be identified automatically in a timely manner as part of business processes, without any human intervention. To be able to do this in a multilingual environment requires the use of a new generation of methodologies and tools. The project will seek to develop methodologies and tools for the creation and maintenance of multilingual ontologies, and for the querying of such ontologies.

The project will:

1. Develop a methodology for expressing a general-purpose ontology for describing the full gamut of electronic commerce applications
2. Develop an open source tool to support the development and maintenance of the upper-level ontology
3. Populate an ontology with Internet-addressable terms for describing electronic commerce applications and services, and the relationships between them
4. Identify a set of existing electronic commerce ontologies and associate them with relevant terms in the upper-level ontology.
5. Input draft specifications into the European and international standardization process.

The results of the project will be reviewed by members of the CEN/ISSS Electronic Commerce Workshop and other relevant standardization organizations.

Existing Techniques

The following techniques have been studied in depth as possible bases for MULECO:

- The EAGLES Guidelines
- Techniques for the Definition of Ontologies
  - IEEE Standard Upper-level Ontology (SUO)
  - DAML+OIL
- XML Representation of ISO 13250 Topic Maps
- Unified Modeling Language (UML)
- The International Standard for Industrial Classification (ISIC)

The EAGLES Guidelines

The EAGLES project was concerned with Natural Language Processing (NLP). As such it had a very wide theme, and needed to cater for the large number of circumstances in which text is used. Many of its features were concerned with word disambiguation in different contexts that are not directly applicable to the more limited applications for which business semantics are required. This paper, therefore, only discusses those features of the EAGLES Guidelines that are directly relevant for the description of business semantics.

The EAGLES Guidelines for Lexical Semantic Standards provided in Chapter 6 of *EAGLES LE3-4244: Preliminary Recommendations on Lexical Semantic Encoding -- Final Report* (http://www.ilc.pi.cnr.it/EAGLES96/EAGLESLE.PDF) points out that:

> "Hierarchical networks [describing hyperonym/hyponym relationships] are very powerful structures because classifications at the top can be inherited to large numbers of word meanings that are directly or indirectly related to these top levels."

and

> "to achieve consistency in encoding hyponymy relations, the best approach is to build the hierarchy top down starting from a limited set of tops or unique beginners … Having an overview of the classes, even at a very high level, makes it possible to more systematically check the possible classes. Furthermore, a systematized top level makes it easier to compare and merge different ontologies."

Business semantics will need someone to develop a top level hierarchy suitable for business uses if they are to be able to interoperate.

As is pointed out in the EAGLES Guidelines, many thesauri cluster words that are related in an unstructured way. For example, the standardized medical thesaurus MESH contains the following entries related to transportation:

```
Transportation
... Aviation
... ... Aircraft
... ... ... Air Ambulances
... ... Space Flight
... ... ... Extravehicular Activity
... ... ... Spacecraft
```

The terms `Space Flight` and `Extravehicular Activity` do not represent subclasses of transportation vehicles but are, rather, types of activities related to certain vehicles. Because of this, MESH can only be used to globally extract words that are related; it cannot be used to make inferences such as: all the things that can be used to transport people, goods, etc.

Words can have different meanings in different contexts. A term that has more than one meaning is said to exhibit polysemy. Words that share the same meaning within a particular context are synonyms. Synonyms should be able to replace each other in stated contexts. If their replacement is not always possible they are referred to as near-synonyms. Near-synonyms have meanings that partially overlap each other. Terms that share the same parent hyperonym but do not overlap in meaning are known as co-homonyms.

Word-sense disambiguation is an important subtask for Information Retrieval, Information Extraction or Machine Translation. One of the key factors in disambiguation is the identification of the domain with which the relevant text is concerned. If you have identified the domains in which each meaning of a term applies you can disambiguate meanings by utilizing information relating to the domains of discourse within a resource.

While hyperonym/homonym relationships work for nouns they are not so useful for other parts of speech, which are generally harder to disambiguate. For most business related classification schemes, however, verbs and other parts of speech are of relatively low importance in identifying meaning. (Verbs identify relationships or actions: they can be useful to identify the role played by particular agents on particular objects. Roles can be classified to create thematic roles. Adjectives are used to describe properties of nouns, e.g. brown gloves. Adverbs, prepositions, conjunctions, etc, are not widely used in electronic business messages. Of key importance to business, however, are terms used for the quantification of measurements and for defining time.)

Many lexicons permit multiple hyperonyms to be associated with a homonym. Three types of hyperonym have been identified within the EAGLES project: exclusive, conjunctive and non-exclusive. For exclusive hyperonyms one of a choice of meanings must be determined by context. Conjunctive hyperonyms allow more than one meaning to be associated with a given context. If either multiple meanings or a single meaning can apply in a given context the hyperonym is deemed to be non-exclusive.

The EAGLES-based EuroWordNet distinguishes between Entities, Concepts, Events and States. Each of these is further divided, with up to 5 levels of subdivision. A typical EuroWordNet entry has the form:

```
[ -ORTHOGRAPHY : horse
  -WORD-SENSE-ID : horse_1
  -BASE-TYPE-INFO : [ BASE-TYPE: ANIMAL
                      LX-RELATION: LX-HYPONYM]
                    [ BASE-TYPE: OBJECT
                      LX-RELATION: LX-HYPONYM]
  SYNONYMS : Equus_caballus_1
  HYPERONYMS : [HYP-TYPE: conjunctive
                HYP-ID: animal_1]
               [HYP-TYPE: conjunctive
```

```
                HYP-ID: equid_1]
               [HYP-TYPE: non-exclusive
                HYP-ID: pet_1]
               [HYP-TYPE: non-exclusive
                HYP-ID: draught_animal_1]
  HYPONYMS : [HYP-TYPE: disjunctive
              HYP-ID: mare_1]
             [HYP-TYPE: disjunctive
              HYP-ID: stallion_1]]
```

Meronymy is defined as a lexical part-whole relationship between elements. A good example is provided by human body parts. "Finger" is a meronym of "hand" which is a meronym of "arm" which is a meronym of "body". The "inverse relation" is called holonymy. “Body" is the holonym of "arm" which is the holonym of "hand" which is the holonym of "finger". The co-meronymy relationship is one between lexical items defining sister parts (arm, leg, head are co-meronyms of body). Meronymy is different from taxonomy because it does not classify elements by class. That is to say, the hierarchical structuring of meronymy does not originate in a hierarchy of classes (toes, fingers, heads, legs, etc, are not hierarchically related).

Not all meronyms are related to a single holonym. For example, "nail" is more general than its holonym "toes" as it can also be part of a finger as well. Cruse introduced the notions of super-meronym ("nail" is a super-meronym of "toes") and hypo-holonym ("toes" is a hypo-holonym of "nail") to allow for this.

The EAGLES paper recommends that "any lexical semantic standard should record a simple binary relation of antonymy where possible between [opposite] word senses". For example, "north" is the antonym of "south", and vice versa.

The on-going work, within the ISLE project for the development of International Standards for Language Engineering (http://www.ilc.pi.cnr.it/EAGLES96/isle/), on a Multilingual ISLE Lexical Entry (MILE) will extend the EAGLES Guidelines to cover the relationships between entries in different languages.

Techniques for the Definition of Ontologies

An ontology is a particular system of categories that provides a certain vision of the world. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships (e.g. lower-level terms meet the criteria set for higher-level terms). An ontology is the general framework within which catalogues, taxonomies, terminologies, etc, may be organized.

The key ingredients that make up an ontology are a vocabulary of terms and a precise specification of what those terms mean. But ontologies also analyse the fundamental categories of *objects, their current state,* and whether they form a *part* or the *whole* of something else, as well as the relations between parts and the whole and their laws of dependence.

A formal ontology is the result of combining the intuitive, informal method of classical ontology analysis with the formal, mathematical method of modern symbolic logic. Over the years a wide range of formal ontologies have been proposed. To make it possible for ontologies to exchange data a number of "knowledge representation languages" have been developed, including KIF, Ontolingua, SNePS, HOL and Conceptual Graphs. Of these the most influential seems to have been the Knowledge Interchange Format (KIF). The basis for the semantics of KIF is a conceptualization of the world in terms of objects and relations among those objects. There are nine types of terms in KIF -- individual variables, constants, character references, character strings, character blocks, functional terms, list terms, quotations, and logical terms.

IEEE Standard Upper-level Ontology (SUO)

KIF, which is in the process of being published as a US standard by ANSI (see http://logic.stanford.edu/kif/dpans.html), has been chosen by IEEE as the basis for a Standard Upper-level Ontology (SUO). This upper ontology is limited to concepts that are meta, generic, abstract and philosophical, and therefore are general enough to address (at a high level) a broad range of domain areas. As well as very high level constructs such Independent Entity and Relative Entity SUO will cover such things as Agents, Persons and Organizations, using KIF definitions of the form:

```
(subclass-of Agent Object)
(subclass-of Person Agent)
(subclass-of Organization Agent)
(subclass-of Publisher Organization)
(subclass-of University Organization)
(disjoint Person Organization)
(subclass-of LegalObligation InstitutionalObligation)
```

and constructs for basic business functions, such as:

```
(subclass-of Quantity SpatialForm)
(subclass-of Weight Quantity)
(subclass-of Arrangement Schema)
(subclass-of Number Arrangement)
(subclass-of Set Arrangement)
```

SUO will also define instances of particular relationships, using formulations such as:

```
(instance-of hasAnnotation BinaryRelation)
(nth-domain hasAnnotation 1 Object)
(nth-domain hasAnnotation 2 TextObject)
```

and

```
(instance-of subProcess BinaryRelation)
(nth-domain subProcess 1 Process)
(nth-domain subProcess 2 Process)
```

Definitions can be assigned to SUO concepts using documentation statement of the form:

```
(documentation Agent "An active animate entity that voluntarily
initiates an action.")

(documentation Arrangement "Mathematical structures that do not have
spatial dimensions: numbers, sets, lists, algebras, grammars, and the
data structure of computer science. Arrangement includes the
subclasses whose names are derived from _taxis_, the Greek word for
"arrangement", including taxonomies and syntax.  All the syntactic
```

forms in natural languages, programming languages, and versions of symbolic logic are included under Arrangement.")

As was the case with the all-encompassing lexical approach proposed by EAGLES, one of the major problems with the proposed Standard Upper-level Ontology is that it is designed to cover all knowledge, and therefore starts with concepts that are at much too high a level for the integration of business processes. It would be more correct to call it the Standard Top-level Ontology as it is designed to encompass all ontologies, rather than provide an upper level for a set of ontologies that cover specific areas, of the type proposed for the Multilingual Upper-Level Electronic Commerce Language.

Note: MULECO is not designed to integrate all existing ontologies, or to provide a meta-schema for describing ontologies. It is strictly limited to providing a means of identifying the relationsh ips between existing ontologies by providing them with a set of addressable shared terms that they can link their top-levels to.

DAML+OIL

The Ontology Inference Language (OIL) that has been adopted as part of the DARPA Agent Markup Language (DAML) is an application of the W3C Resource Description Framework (RDF). DAML+OIL (http://www.daml.org/2001/03/reference.html) divides the world up into objects, which are elements of DAML classes, and datatype values, i.e., values that come from XML Schema datatypes, like the integer 4.

In DAML+OIL an ontology is recorded using a set of definitions that define classes, subclasses, properties that connect classes and individual instances. Classes have names, descriptive documentation, statements of which class it creates a subclass of, and one or more constraining facets. Classes are allowed to have multiple superclasses, which are deemed to be conjunctive unless specifically defined as being disjoint. DAML+OIL properties are divided into two sorts, those that relate objects to other objects and those that relate objects to datatype values. The former belong to `daml:ObjectProperty` and the latter belong to `daml:DatatypeProperty`. Properties are defined as having ranges of permitted values. Multiple ranges can be applied to a property but then the value of the property must satisfy all range statements (they are conjunctive rather than disjoint, with only the intersection of all the statements being valid). Properties, but not their values, can be defined as being the inverse of each other

DAML Class definitions can be defined in multiple statements, as the following parts of a March 2001 DAML Class definition example illustrate:

```
<daml:Class rdf:ID="Person">
  <rdfs:subClassOf rdf:resource="#Animal"/>
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#hasParent"/>
      <daml:toClass rdf:resource="#Person"/>
    </daml:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <daml:Restriction daml:cardinality="1">
      <daml:onProperty rdf:resource="#age"/>
    </daml:Restriction>
  </rdfs:subClassOf>
```

```
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#shoesize"/>
      <daml:minCardinality>1</daml:minCardinality>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
<daml:Class rdf:about="#Person">
  <rdfs:comment>Every person is a man or a woman</rdfs:comment>
  <daml:disjointUnionOf rdf:parseType="daml:collection">
    <daml:Class rdf:about="#Man"/>
    <daml:Class rdf:about="#Woman"/>
  </daml:disjointUnionOf>
</daml:Class>
...
<daml:Class rdf:about="#Person">
  <rdfs:subClassOf>
    <daml:Restriction daml:maxCardinality="1">
      <daml:onProperty rdf:resource="#hasSpouse"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
...
<daml:Class rdf:about="#Person">
  <rdfs:subClassOf>
    <daml:Restriction daml:maxCardinalityQ="1">
      <daml:onProperty rdf:resource="#hasOccupation"/>
      <daml:hasClassQ rdf:resource="#FullTimeOccupation"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
```

DAML classes are a subset of the RDF Schema (RDFS) `Class` construct. The `rdfs:SubclassOf` element that forms its first level contents is extended by the use of the `daml:Restriction` definition. Whilst this leads to a more detailed definition of DAML classes it does mean that there is a confusion between classes of the type used for defining schemas in RDF and the types of categorization used to define an ontology.[1]

An instance of the DAML Class shown above might take the form:

```
<Person rdf:ID="Peter">
  <rdfs:comment>
  Peter is an instance of Person. Peter has shoesize 9.5 and age 46
  </rdfs:comment>
  <shoesize>9.5</shoesize>
  <age><xsd:integer rdf:value="46"></age>
</Person>
```

Each DAML ontology can have associated with it metadata that identifies what the ontology is about, the version of DAML being used, and other information relevant to the management of the ontology. Ontologies can import part or all of another ontology.

A typical DAML+OIL header takes the form:

---

[1] The classes used in programming are typically additive in nature, properties at a lower level being added to those at higher levels. Categories in ontologies, in contrast, are restrictive in nature, the properties at one level distinguishing subsets of the properties applicable at a higher level.

```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
 xmlns:xsd ="http://www.w3.org/2000/10/XMLSchema#"
 xmlns:dex ="http://www.daml.org/2001/03/daml+oil-ex#"
 xmlns:exd ="http://www.daml.org/2001/03/daml+oil-ex-dt#"
 xmlns     ="http://www.daml.org/2001/03/daml+oil-ex#" >
<daml:Ontology rdf:about="">
 <daml:versionInfo>$Id: daml+oil-ex.daml,v 1.9 2001/05/03 16:38:38
                   mdean Exp $</daml:versionInfo>
  <rdfs:comment>
    An example ontology, with data types taken from XML Schema
  </rdfs:comment>
  <daml:imports rdf:resource="http://www.daml.org/2001/03/daml+oil"/>
</daml:Ontology>
```

XML Representation of ISO 13250 Topic Maps

The XML Topic Maps (XTM) specification provides a model and grammar for representing the structure of information resources used to define *topics*, and the associations (relationships) between topics. Names, resources, and relationships are said to be *characteristics* of topics. Topics can have their characteristics defined within *scopes* that limit the contexts within which the names and resources are regarded as meaningful. One or more interrelated documents employing this grammar is called a “topic map”.

A minimal topic, consisting of a base name and a single resource identified as an occurrence of the topic, could be defined as:

```
  <topic id="hamlet">
    <instanceOf><topicRef xlink:href="#play"/></instanceOf>
    <baseName>
      <baseNameString>Hamlet, Prince of Denmark</baseNameString>
    </baseName>
    <occurrence>
      <instanceOf>
        <topicRef xlink:href="#plain-text-format"/>
      </instanceOf>
      <resourceRef
       xlink:href="ftp://www.gutenberg.org/pub/1ws2610.txt"/>
    </occurrence>
   </topic>
```

An association representing the relationship between Shakespeare and the play *Hamlet* might look like this:

```
  <association>
    <instanceOf><topicRef xlink:href="#written-by"/></instanceOf>
    <member>
      <roleSpec><topicRef xlink:href="#author"/></roleSpec>
      <topicRef xlink:href="#shakespeare"/>
    </member>
    <member>
      <roleSpec><topicRef xlink:href="#work"/></roleSpec>
      <topicRef xlink:href="#hamlet"/>
    </member>
  </association>
```

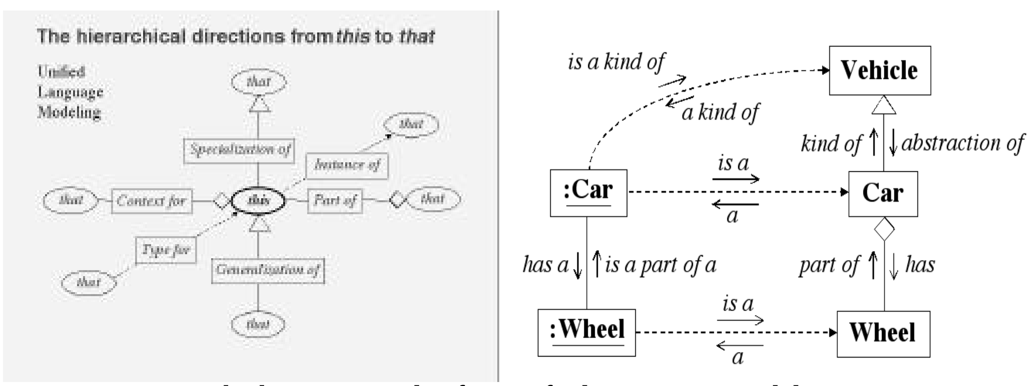
Within topic maps, scopes establishes the contexts in which a name or an occurrence is assigned to a given topic, and the context in which topics are related through associations. Any topics having the same base name in the same scope implicitly refer to the same subject and therefore should be merged.

XTM, unlike the underlying ISO standard, privileges two types of association: class-instance, and superclass-subclass. It fails, however, to follow the ISO standard in permitting the assignment of user-defined facets to provide multi-dimensional views of topic maps.

The Unified Modeling Language (UML)

UML is the main technique used for modelling business processes. It forms the basis of the UN/CEFACT Modeling Methodology (UMM), Version 10 of which can be found at http://www.gefeg.com/tmwg/n090r10.htm). UMM forms the basis for the modelling of business processes within the ebXML/ebTWG initiative to establish a new generation of business messaging services that is compatible with XML.

The Centre for User-oriented IT Design (CID) at the Swedish Royal Institute of Technology (KTH) have developed a technique for generalizing UML models to provide Unified Language Modeling (ULM) that allows formal models to be expressed in terms that are easily understood by businesses. The following diagrams summarize this technique:



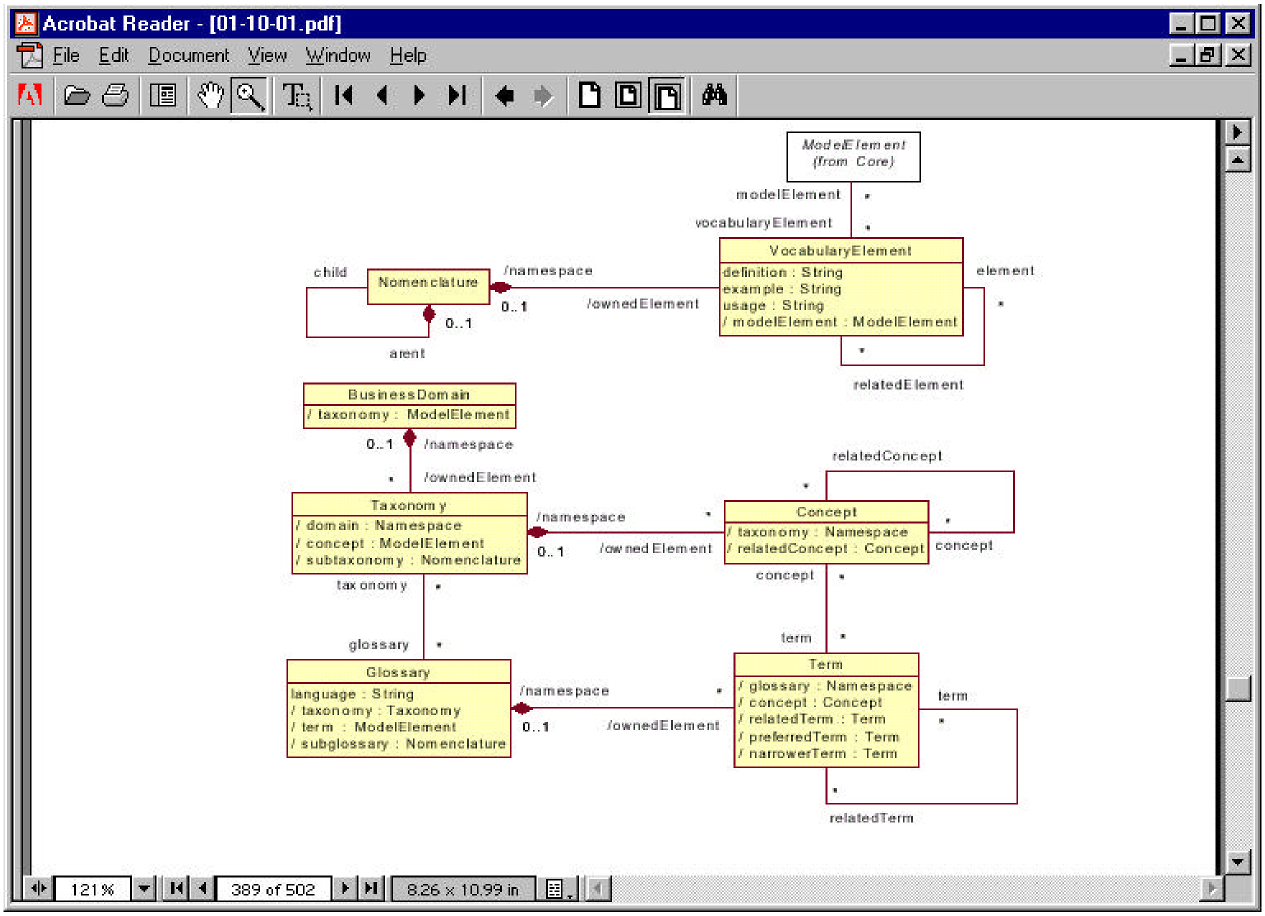
The basic principles for Unified Language Modeling

Using this technique you can understand that:

- The concept called car represents kind of vehicle
- The concept called vehicle is an abstraction of the concept called car
- The concept called wheel forms a part of a car
- A car has one or more wheels
- A specific car (:car) is an instance of the car concept
- A specific wheel (:wheel) is an instance of the wheel concept
- A specific wheel is a part of a specific car
- A specific car is a kind of vehicle

The Common Warehouse Metamodel (CWM) Business Nomenclature Package

The following diagram summarizes the parts of the Open Management Group's Common Warehouse Metamodel (OMG CWM) Business Nomenclature Package:



The OMG model considers Taxonomies as consisting of a number of Concepts, which may or may not have Related Concepts. A Taxonomy may be related to a Glossary, which contains one or more Term, which may have a number of Related Term, one Preferred Term and one or more Narr ower Terms. Terms can be related to Concepts in a Taxonomy.

The International Standard for Industrial Classification (ISIC)

ISIC Version 3.0 (ISIC3) is the primary scheme used by governments throughout the world to classify business activity. It forms the basis of the Euopean NACE classification of EU economic activity. ISIC uses the following top level hierarchy:

- A - Agriculture, hunting and forestry
- B - Fishing
- C - Mining and quarrying
- D - Manufacturing
- E - Electricity, gas and water supply
- F - Construction

- G - Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
- H - Hotels and restaurants
- I - Transport, storage and communications
- J - Financial intermediation
- K - Real estate, renting and business activities
- L - Public administration and defence; compulsory social security
- M - Education
- N - Health and social work
- O - Other community, social and personal service activities
- P - Private households with employed persons
- Q - Extra-territorial organizations and bodies

Each of these subdivisions is further subdivided. For example, the Manufacturing subdivision is further subdivided into:

- 15 - Manufacture of food products and beverages
- 16 - Manufacture of tobacco products
- 17 - Manufacture of textiles
- 18 - Manufacture of wearing apparel; dressing and dyeing of fur
- 19 - Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
- 20 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
- 21 - Manufacture of paper and paper products
- 22 - Publishing, printing and reproduction of recorded media
- 23 - Manufacture of coke, refined petroleum products and nuclear fuel
- 24 - Manufacture of chemicals and chemical products
- 25 - Manufacture of rubber and plastics products
- 26 - Manufacture of other non-metallic mineral products
- 27 - Manufacture of basic metals
- 28 - Manufacture of fabricated metal products, except machinery and equipment
- 29 - Manufacture of machinery and equipment
- 30 - Manufacture of office, accounting and computing machinery
- 31 - Manufacture of electrical machinery and apparatus

- 32 - Manufacture of radio, television and communication equipment and apparatus
- 33 - Manufacture of medical, precision and optical instruments, watches and clocks
- 34 - Manufacture of motor vehicles, trailers and semi-trailers
- 35 - Manufacture of other transport equipment
- 36 - Manufacture of furniture
- 37 - Recycling

It should be noted that the ISIC listing is only available in three languages, English, French and Spanish. Translations into other languages would be needed to provide a truly multilingual classification scheme.

Proposed Approach

The ontology representation language should be expressed in XML so that individual components of it can be referenced as component parts of either a Unique Resource Indicator (URI), XML Path definition or XML Query.

The underlying structure of the XML should be based on the concepts described in the EAGLES framework, but with alternative forms of element names based on typical business renditions of technical terms (e.g. BroaderTerm in place of Hypernym). The terms to be adopted form EAGLES, and their equivalent business terms are shown in the following table:

Linguistic Terminology	Ontological Terminology	Business Terminology
Phrase	Concept	Term/Name
Hypernym	Superclass	Broader Term
Holonym Synonym	Subclass Synonym	Narrower Term Alternative Term
Near-Synonym		Near Equivalent
Holonym		Forms Part Of
Meronym		Has Part/Subprocess
Antonym		Opposite
	Restriction	Constraint

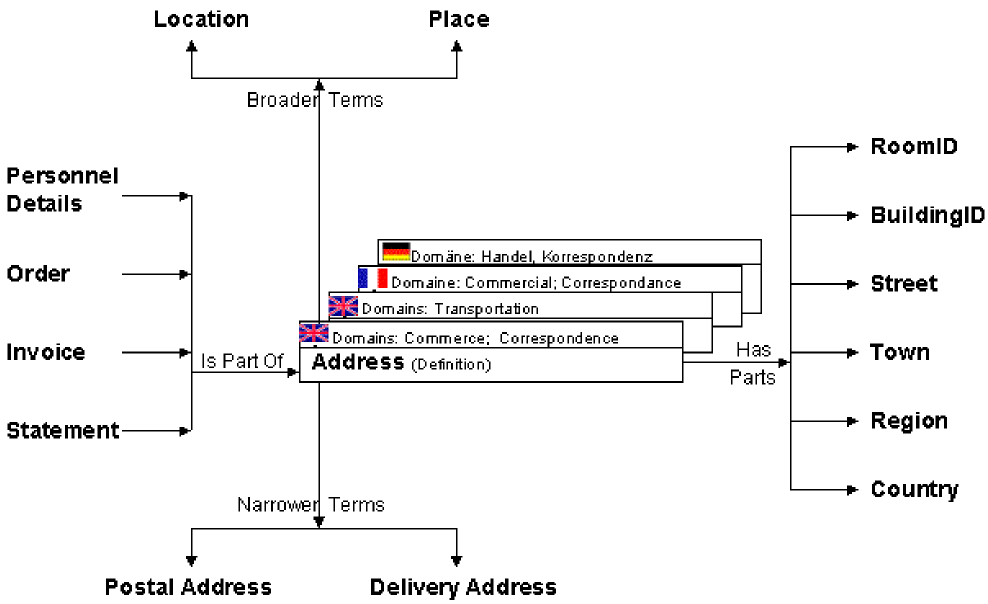
Entries should be provided with metadata which is defined by reference to existing sources of information or by use of standardized metadata descriptors. Each term must be assigned to at least one subject domain, ideally by linking it to a standardized domain identified within ISIC.

A simplified example of the use of these terms might have the following form:

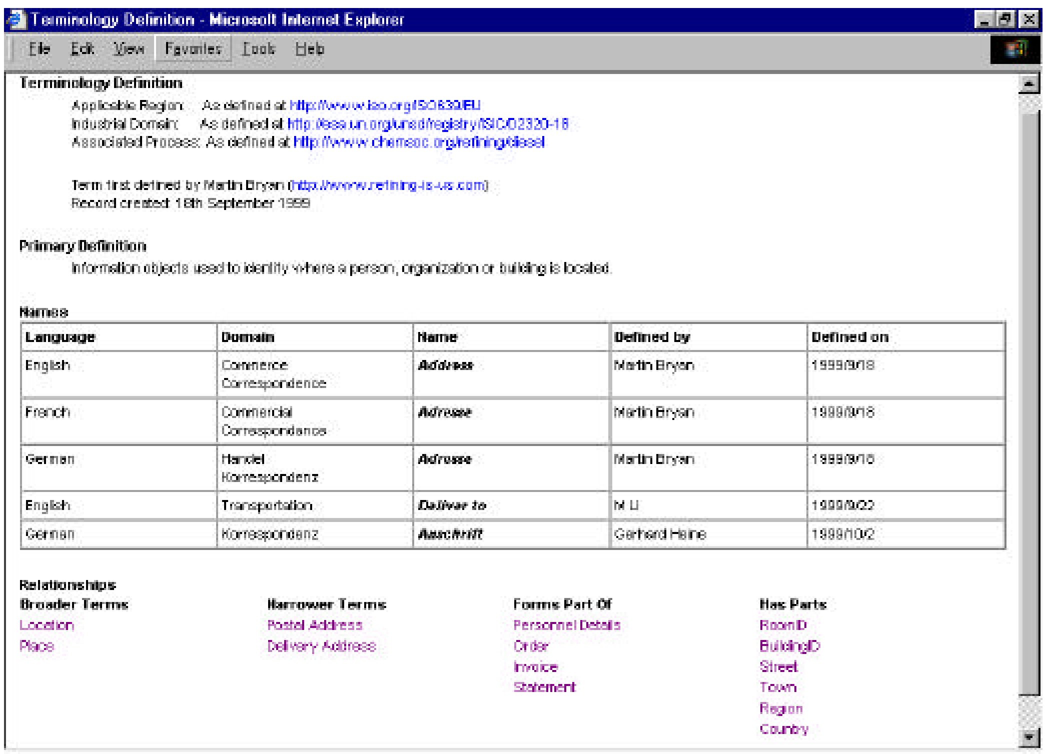
```
<Ontology
 Region="http://www.iso.org/ISO639/EU"
 Industry="http://esa.un.org/unsd/registry/ISIC3/D2320-18"
 Process="http://www.chemsoc.org/refining/diesel">
 <Term
  ID="Address"
  RecordedBy="Martin Bryan"
  Organization="http://www.refining-is-us.com"
  WhenRecorded="1999-09-18">
```

```
  <Definition>Information objects used to identify where a
              person, organization or building is
              located.</Definition>
  <Name xml:lang="EN">Address</Name>
  <Name xml:lang="DE FR">Adresse</Name>
  <SubjectDomain xlink:href="#id('CommerceDomain')"/>
  <SubjectDomain xlink:href="#id('CorrespondenceDomain')"/>
  <Alternative xml:lang="EN" RecordedBy="M Li"
               WhenRecorded="1999-09-22"
               SubjectDomain="#id('TransportationDomain')">
   Deliver To</Alternative>
  <NearEquilavent RecordedBy="Gerhard Heine"
                  WhenRecorded="1999-10-02" xml-lang="DE"
                  SubjectDomain="#id('CorrespondenceDomain')">
   Anschrift</NearEquivalent>
  <BroaderTerm xlink:href="GenericConcepts.xml#id('Location')">
   Location</BroaderTerm>
  <BroaderTerm xlink:href="GeographicConcepts.xml#id('Place')">
   Place</BroaderTerm>
  <NarrowerTerm xlink:href="GenericConcepts.xml#id('Post')">
   Postal Address</NarrowerTerm>
  <NarrowerTerm
   xlink:href="TransportConcepts.xml#id('DeliveryPoint')">
   Delivery Address</NarrowerTerm>
  <FormsPartOf
   xlink:href="PersonnelConcepts.xml#id('PrivateAddress')">
   Personnel Details</FormsPartOf>
  <FormsPartOf
   xlink:href="CommercialConcepts.xml#id('Order')">
   Order</FormsPartOf>
  <FormsPartOf
   xlink:href="CommercialConcepts.xml#id('Invoice')">
   Invoice</FormsPartOf>
  <FormsPartOf
   xlink:href="CommercialConcepts.xml#id('Statement')">
   Statement</FormsPartOf>
  <HasPart
   xlink:href="LocationConcepts.xml#id('RoomID')">
   RoomID</HasPart>
  <HasPart
   xlink:href="LocationConcepts.xml#id('BuildingID')">
   BulidingID</HasPart>
  <HasPart
   xlink:href="LocationConcepts.xml#id('Street')">
   Street</HasPart>
  <HasPart
   xlink:href="LocationConcepts.xml#id('Town')">
   Town or City</HasPart>
  <HasPart
   xlink:href="LocationConcepts.xml#id('Region')">
   Region</HasPart>
  <HasPart
   xlink:href="GeographicConcepts.xml#id('Country')">
   Country</HasPart>
 </Term>
...
</Ontology>
```


Such a file could be presented graphically as:



Alternatively it could be converted, using the XSL Transformation Language, into an HTML file for display on a web browser in the following format:



Terminology Definition - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Terminology Definition

Applicable Region: As defined at http://www.iso.org/ISO639/EU
Industrial Domain: As defined at http://ess.un.org/unsd/registry/ISIC/D2320-18
Associated Process: As defined at http://www.chemsoc.org/refining/diesel

Term first defined by Martin Bryan (http://www.refining-is-us.com)
Record created: 18th September 1999

Primary Definition

Information objects used to identify where a person, organization or building is located.

Names

Language	Domain	Name	Defined by	Defined on
English	Commerce Correspondence	***Address***	Martin Bryan	1999/9/18
French	Commercial Correspondance	***Adresse***	Martin Bryan	1999/9/18
German	Handel Korrespondenz	***Adresse***	Martin Bryan	1999/9/18
English	Transportation	***Deliver to***	M U	1999/9/22
German	Korrespondenz	***Anschrift***	Gerhard Heine	1999/10/2

Relationships

Broader Terms	**Narrower Terms**	**Forms Part Of**	**Has Parts**
Location	Postal Address	Personnel Details	RoomID
Place	Delivery Address	Order	BuildingID
		Invoice	Street
		Statement	Town
			Region
			Country

Current Status

MULECO is an on-going project, and so no formal set of definitions, or accompanying DTD/Schema has been produced. Areas of ongoing study include those currently being undertaken by European research projects such as MILES, CLAMOUR and OntoWeb, and by international e-commerce initiatives such ebXML/ebTWG, related to:

- Formal languages for describing ontologies
- Formal languages for describing multilingual word sets
- Formal models for the maintaining industrial classification schemes
- Formal languages for modelling business processes

Martin Bryan
The SGML Centre
12/12/2001