The Case for a Multilingual Upper-Level Electronic Commerce Ontology

Martin Bryan, The SGML Centre

The classification of business information exchanged via the Internet is of vital importance for the long-term understanding of business processes. This paper looks at how business information can be classified, and the role that ontologies and lexical semantic standards should have in this process. The discussion will illustrate some of the limitations of the current approaches being adopted for the Semantic Web based on the use of the Resource Description Framework (RDF) and the XML representation of Topic Maps (XTM), and suggest ways in which the EAGLES Guidelines for Lexical Semantic Standards and proposed ontology standards such as the Knowledge Interchange Format (KIF) and the DARPA Agent Markup Language Ontology Inference Layer (DAML+OIL) might help to improve existing techniques.

Overview of the EAGLES Guidelines

The EAGLES project is concerned with Natural Language Processing (NLP). As such it has a very wide theme, and needs to cater for the large number of circumstances in which text is used. Many of its features are 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 relatedto 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.

The formal definitions for the EAGLES Guidelines for Lexical Semantic Standards are provided in Annex 1. 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 taxonymy 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.

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.

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 the upper level 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.

The role of 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="#hasFather"/>
    </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>
```

[1] The classes used in programming are 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.

```
  <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:

```
<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 abstract subjects, which are called *topics*. Topics have their characteristics within *scopes*: i.e. the limited contexts within which the names and resources are regarded as their name, resource, and relationship characteristics. 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, scope establishes the context 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.

Defining Business Semantic Hierarchies

All of the methodologies listed above are intended for the general description and use of semantics over very broad subject areas. This means that their starting points are related to much broader categories of information than those currently used in business applications.

There are a number of key points, however, which any proposals for the development of business semantic sets can learn from the above examples, including:

1. The need for multiple inheritance
   All of the above interchange formats allow a subclass to be subordinate to more than one superclass. As such they represent hierarchical networks rather than simple hierarchies.
2. The need for user-defined properties
   With the exception of XTM (but not the underlying ISO 13250 Topic Map standard) the interchange formats allow user-defined properties to be associated with classes/categories.
3. The need to recognize that whole/part relationships are intrinsically different from class/subclass relationships
4. The need to identify opposites in both terms (antonyms) and relationships

In addition there is a key need, not covered in the above descriptions, to be able to identify equivalences between terms in one language with those in another language. In this respect the difference between exact synonyms and near-synonyms identified by the EAGLES guidelines needs to be recognized.

None of the approaches listed above meets all of the needs of a multilingual upper-level electronic commerce ontology. What are the requirements of such an ontology? It must include:

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 identify the 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.

Any business-oriented upper level ontology must take as its start point a set of subjects that is relevant to businesses, rather tha n a general-purpose description of "things". In particular it needs to take as its starting point a well-recognized set of industrial classifications, such as the International Standard for Industrial Classification (ISIC). This starting point will need to be supplemented with terms that will allow industries to be further subdivided in respect to geographical and legislative regions, business processes and business roles.

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 two languages, English and Spanish. Translations into other languages would be needed to provide a truly multilingual classification scheme.

No one organization can expect to produce an ontology that describes all requirements of all businesses. (Most multinational businesses do not have enough knowledge of their own working environments to describe their own requirements in a fully multilingual form.) Therefore a mechanism is needed whereby different organizations, and different parts of the same organization, can define the semantics they use and can relate these to the semantics used in other sectors from the fact that their ontologies are linked to a shared point in an upper level ontology.

If business semantics are to be used to categorize resources available on the Internet they must be defined in a way that can be referenced using a Uniform Resource Identifier (URI). This suggests that they should be coded as a hierarchically structured object whose path or unique identifier can be used as a reference to the semantic. Use of the Extensible Markup Language (XML) as the encoding mechanism for defining the upper level ontology would be the best way to ensure that the ontology could be widely referenced over the Internet.

A possible representation for such a language, based on the EAGLES framework, with alternative forms of element names based on non-technical renditions of technical terms (e.g. BroaderTerm in place of Hypernym), might take the form:

```
<Ontology
 Region="http://www.iso.org/ISO639/EU"
 Industry="http://esa.un.org/unsd/registry/ISIC/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">Adresse</Name>
  <SubjectDomain xlink:href="#id('CommerceDomain')"/>
  <SubjectDomain xlink:href="#id('CorrespondenceDomain')"/>
  <Synonym xml:lang="EN" RecordedBy="M Li" WhenRecorded="1999-09-22"
   SubjectDomain="#id('TransportationDomain')">
   Deliver To</Synonym>
  <NearSynonym RecordedBy="Gerhard Heine" WhenRecorded="1999-10-02"
   xml-lang="DE" SubjectDomain="#id('CorrespondenceDomain')">
   Anschrift</NearSynonym>
  <BroaderTerm xlink:href="GenericConcepts.xml#id('Location')">
   Location</BroaderTerm>
  <Hypernym xlink:href="GeographicConcepts.xml#id('Place')">
   Place</Hypernym>
  <NarrowerTerm xlink:href="GenericConcepts.xml#id('Post')">
   Postal Address</NarrowerTerm>
  <Hyponym xlink:href="TransportConcepts.xml#id('DeliveryPoint')">
   Delivery Address</Hyponym>
  <Holonym xlink:href="PersonnelConcepts.xml#id('PrivateAddress')">
   Personnel Details</Holonym>
  <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>
  <Meronym xlink:href="GeographicConcepts.xml#id('Country')">
   Country</Meronym>
 </Term>
...
</Ontology>
```

This example is not complete, and does not illustrate all the points listed above as requirements for a Multilingual Upper-Level Electronic Commerce Ontology, but it does provide a starting point from which such an ontology could be developed and fully described once a suitable body for undertaking this work has been identified.

Acknowledgements

The following people have assisted in reviewing and extending this document:
Bernard Vatant, Mondeca

Annex 1: The EAGLES Guidelines for Lexical Semantic Standards

The basic information unit is a word sense. Obligatory attributes are preceded by a dash (-). The star and plus signs (*, +) are used in the usual way to indicate expansion of types (e.g. subject-domain), into 0; :::; n and 1; :::; n tokens (music, dance). The vertical bar (|) indicates disjunction.

```
word-sense-entry -->
        [ -ORTHOGRAPHY : string
          -WORD-SENSE-ID : word-sense-id
          -BASE-TYPE-INFO : base-type-info*
          SUBJECT-DOMAIN : subject-domain*
          SYNONYMS : word-sense-id*
          NEAR-SYNONYMS : word-sense-id*
          HYPONYMS : hyponym*
          HYPERONYMS : hyperonym*
          ANTONYMS : antonym*
          MERONYMS : meronym*
          HOLONYMS : holonym*
          QUANTIFICATION : quantification
          COLLOCATIONS : collocation*
          SEMANTIC-FRAME : sem-frame
          ACTIONALITY : actionality ]
```

A word sense identifier is an integer which refers to a WordNet synset.

```
word-sense-id --> integer
```

Base type information provide a specification of the conceptual entities germane to the word sense in question, chosen from the list of base types, and the relation(s) that each conceptual entity bears to the word sense (e.g. LX-synonym, LX-near-synonym)

```
base-type-info -->
        [ BASE-TYPE : base-type
          LX-RELATION : lx-relation+ ]
base-type --> (entity | animate | ...)
lx-relation --> (lx-synomym | lx-near-synonym | lx-hyponym
                 | lx-hyperonym | lx-holonym | lx-meronym
                 | lx-subevent)
```

Subject domain information is encoded in terms of defined categories and the subsumption relations among them.

```
subject-domain -->(sports-games-pastimes|history-heraldry|...)
sports-games-pastimes --> hunting-fishing
hunting-fishing --> (hunting | fishing)
...
```

Information about hyponymy, antonymy and meronymy involves specification of the specific type of lexical semantic relation involved.

Note: \non-exclusive" is the default value for 'HYP -TYPE'.

```
hyponym : [ HYP-TYPE: (exclusive | conjunctive | non-exclusive)
            HYP-ID : word-sense-id ].
hyperonym : [HYP-TYPE: (exclusive | conjunctive | non-exclusive)
             HYP-ID : word-sense-id ].
antonym: [ ANT-TYPE: (complementary | gradable |
                      pseudo-comparative |true-comparative |
```

```
                    antipodal | reversive | relational)
          ANT-ID : word-sense-id ].
meronym : [ MER-TYPE: (member | substance | part)
           HOLS: [HOL-ID: word-sense-id
           REL-TYPE : (exclus | conjunct | non-exclus)]+
           MER-ID : word-sense-id ].
holonym : [ HOL-TYPE: (member | substance | part)
           MERS: [MER-ID: word-sense-id
           REL-TYPE : (exclus | conjunct | non-exclus)]+
           HOL-ID : word-sense-id ].
```

Quantification specifies lexical properties of determiners such as quantification strength and direction of entailment.

```
quantification : [ Q-STRENGTH : (weak | strong)
                   Q-ENTAILMENT : (upward | downward) ]
```

Collocation information includes reference to each collocate, expressed in terms of sets of word senses, its location and upper/lower distance as well as the relevant dependency configuration (e.g. head/dependent to head/dependent).

```
collocation -->
        [ COLLOCATE : word-sense-id+
          DIRECTION : (left | right)
          DISTANCE : [LOWER-LIMIT : integer
                      UPPER-LIMIT : integer]
          DEPENDENCY : (h2d | d2h | d2d | h2h)
          DEP_TYPE : dependency_type
          PROBABILITY : probability-value]
```

Semantic frames include information about the semantic class of a predicate, expressed as either a base-type or a set of word senses, and itsarguments.

```
sem-frame : [-ARG : arg*
             PRED-CLASS : (base-type* | word-sense-id*)
             SEM-REPT : {Some form of semantic representation,
                         with links to arg*}
             QUALIA : {Some encoding of Qualia roles,
                       with links to args*} ]
```

Argument information includes reference to selectional restrictions, collocations and thematic role. No specific guidelines are given for semantic representation and qualia encoding.

```
arg --> [ SELECTIONAL-RESTR : (base-type* | word-sense-id*)
          COLLOCATIONS : collocation*
          THEMATIC-ROLE : th-role ]
```

Both static and dynamic options are give for the encoding of actionality information (CUMULATIVE = +SQA, Quantized = _ SQA)

```
actionality --> (static-actionality | dynamic-actionality)
static-actionality --> [ (STATE | PROCESS | ACHIEVEMENT |
                          ACCOMPLISHMENT) ].
dynamic-actionality -->
         [ACT-TYPE : (DYNAMIC | STATIVE)
          THEME-REF : (CUMULATIVE | QUANTIZED) ]
```

Annex 2: DAML-OIL Language Elements

The March 2001 version of DAML+OIL recognized the following information elements:

- cardinality
- cardinalityQ
- Class
- complementOf
- Datatype
- DatatypeProperty
- DatatypeRestriction
- Datatype value
- differentIndividualFrom
- Disjoint
- disjointUnionOf
- disjointWith
- domain
- equivalentTo
- hasClass
- hasClassQ
- hasValue
- imports
- intersectionOf
- inverseOf
- maxCardinality
- maxCardinalityQ
- minCardinality
- minCardinalityQ
- ObjectClass
- ObjectProperty
- ObjectRestriction
- oneOf
- onProperty
- Ontology
- Property
- range
- Restriction
- sameClassAs
- sameIndividualAs
- samePropertyAs
- subClassOf
- subPropertyOf
- toClass
- TransitiveProperty
- UnambigousProperty
- unionOf
- UniqueProperty
- versionInfo