

# Creating an Ontology for a Multilingual E-Commerce Dictionary

Johannes Schwall, [schwall@uni-muenster.de](mailto:schwall@uni-muenster.de)

Arbeitsbereich Linguistik, Westfälische Wilhelms-Universität Münster, 21.12.2007

## 1 Introduction

In this paper, knowledge of the basic principles of ontologies is assumed, although insight into the historical and philosophical background of ontologies is not needed. For a general introduction to ontologies please refer to online resources such as <http://www.formalontology.it/> or printed material like the introduction by SMITH (2003). In the following, the term *ontology* refers to the concept of an ontology as given by SMITH (2003, p. 158) for the Tower of Babel problem in information science: “a canonical description” of a shared taxonomy of entities; “a dictionary of terms formulated in a canonical syntax and with commonly accepted definitions designed to yield a lexical or taxonomical framework for knowledge-representation which can be shared by different information systems communities.”<sup>1</sup> More precisely, we use the term ‘ontology’ in GRUBER’s sense as “an explicit specification of a conceptualization”<sup>2</sup>.

## 2 The Ontology in the Context of the E-Commerce Dictionary

Taking part in the *Languages for E-Commerce*<sup>3</sup> project, the Arbeitsbereich Linguistik<sup>4</sup> has developed a multilingual dictionary for electronic commerce with ontological support. The dictionary was designed with European languages in focus and comprises data

for English, Finnish, German, Polish, Portuguese and Spanish<sup>5</sup> but it is open to the addition of further languages.

The dictionary consists of separate files for each language in XML format and stores for each entry the lemma, a definition, grammatical information, synonymy, acronyms and collocations. The translation between the languages is done via linking of internal IDs. Up to this point the construct is no more and no less than a standard dictionary and a basic interface allowed for simple automatic linking: in cases where a string within a lemma’s definition or collocations matched another lemma, a link would lead the user to the respective entry.

To improve on this, an ontology was created which introduced sensible and useful named links between dictionary entries; its aim: to provide means for querying the knowledge base for both implemented and inferred knowledge as well as to provide means for disambiguation.

In accordance with the project’s aims, the target groups for the dictionary are students and teachers in higher education as well as employees of small and medium enterprises involved in any kind of electronic business. Therefore, the terminology covered in the dictionary were developed in close cooperation with the teams working on course material in the “Languages for E-Commerce” project.

In this paper, the dissociation of the ontology’s topic (e-commerce) is presented in section 3, fol-

<sup>1</sup>SMITH Blackwell Guide to the Philosophy of Computing and Information 2003 p. 158.

<sup>2</sup>GRUBER (1993, p. 1)

<sup>3</sup>The *Languages for E-Commerce* was a project funded by the EU’s Leonardo da Vinci program, running from 2004 to 2007; <http://www.languages-for-ecommerce.com/>, <http://lfe.uni-muenster.de/>.

<sup>4</sup>The Arbeitsbereich Linguistik is part of the Department for English at the University of Münster, Germany: <http://santana.uni-muenster.de/>, <http://www.uni-muenster.de/>. The project team at the Arbeitsbereich Linguistik consisted of Annabelle Koppen, Jan Lehnardt, Thorsten Merse, Prof. Dr. Wolf Paprotté, Martin Pyka and Johannes Schwall.

<sup>5</sup>English, German and Spanish as project’s the main languages contain 1.000 entries each, Portuguese is present with more than 800 entries, Finnish and Polish with more than 200 each.

lowed by a survey of the creation of the e-commerce ontology (part 4). Some specific problems and an outlook on further improvements are presented as conclusions in section 5.

### 3 Dissociating the Ontology's Topic

It has to be made perfectly clear that “there is no single correct ontology for any domain”<sup>6</sup>. Modelling an ontology always means modelling a certain view on the field observed. Each singular decision made in the process of ontology development means a possible deviation from a path some other developer might have taken. Thus, any class or property included in an ontology might have found another place or form of existence if another view had been expressed in the model. Nevertheless, in designing the basic taxonomical structures of our ontology, we tried to take a point of view as practical to the possible users' approaches as possible. Thus, questions posed to the ontology are the guidelines for the basic layout: The first thing to be considered in creating a domain-centered ontology is to try to define the ontology's topic as clearly as possible. With the manpower available it is not realistic to expect full coverage of any too comprehensive field, thus limiting the size and the scope of the ontology is essential.<sup>7</sup> For this, in their 7-step-manual for creating an ontology, NOY/MCGUINNESS propose starting with competency questions, mainly to act as a “litmus test” to indicate whether the ontology is as complete as necessary, but also as a thread to indicate the general direction for the ontological design.<sup>8</sup> For the e-commerce domain, we decided to take a more general approach including the separate fields of technology, business and legislation.

Next, a starting point for designing the ontology had to be found and the need to quickly get to a comprehensive system which would cover the most important knowledge fields let us center the first nodes around “webshop” concept. The scope of a small ontological network centered around “web-

shop” would thus contain terms from technology (e. g. *browser*, *web server*), business (e. g. *marketing*, *vendor*) and legislation (e. g. *license*, *general terms and conditions*). A practical competency question, which could act as a test for the ontology later, would accordingly be: “What are the parts of a webshop?” To get a qualified answer, detailed information from all fields had to be entered into the domain model and relations between different nodes and subdomains had to be created.

As stated above, the e-commerce ontology was designed first to support the dictionary as a knowledge base. Additional uses like synonymy and translation that may be implemented in other ontologies have not been implemented here, because both were available through means of the dictionary which already contained this information. It might be prudent to develop ontologies of both synonymy and translation at a later date. In the frame of this project, it was decided to refrain from further complications because of budget and time restrictions.

Of course, the search for synonyms is possible within the e-commerce dictionary, but does only work within the dictionary's own scope and directly implemented knowledge. Synonyms are not stored as separate entries and cannot carry additional information themselves.<sup>9</sup> The multilingual facet of the dictionary connects terms across languages and covers the basic needs of translation. However, the concept currently comprised does not allow for gradual correspondency – neither in translation nor in synonymy.

In another aspect of simplicity, the e-commerce ontology currently does not provide means to include similar or analogue concepts: Within the intersection of the internet culture and the European Union we assume that people have the exact same meaning for every concept represented in the ontology, independent of their origin – a *webshop* is a *webshop*; *Amazon* is *Amazon*, although they may have subsidiaries in several European countries. Thus, concepts differing to some degree are not allowed for within the e-commerce ontology; at this point entries have either

<sup>6</sup>NOY/MCGUINNESS p. 23.

<sup>7</sup>Cai Ziegler states that most ontologies are being created in cumbersome manual labour (ZIEGLER p. 56), but have proven to be valuable in well-defined fields of knowledge (ZIEGLER p. 59).

<sup>8</sup>NOY/MCGUINNESS p. 5.

<sup>9</sup>The concept of synonyms was integrated into the project at a date where any step towards a more future-proof design was unrealistic. But as synonyms would normally have at least their own grammatical information and collocations, they might be implemented as separate entries, perhaps even with their own definitions if the synonymy of two entries is correct only to a certain degree.

to be entirely different or completely synonymous.

## 4 Exemplary Design of an Ontology for E-Commerce

The ontology was designed with the Protégé editor<sup>10</sup> developed at Stanford University. Protégé stores data in OWL<sup>11</sup> format and allows for transformation to other formats. The ontology used for the e-commerce dictionary is actually the second attempt of supporting the dictionary with ontological knowledge. A first ontology proved to be technically too intertwined with the dictionary itself after some development time and did not allow for newly arising needs for the tools within the “Languages for E-Commerce” project. The following thus concentrates on decisions made for the second ontology which actually made it into the final product.

The e-commerce ontology was developed in a practical approach, i. e. when questions arrived which technical or theoretical concept to follow, the solution which promised the best outcome within the project frame was chosen.

Depending on the approach chosen to create the ontological hierarchy, one can either start with very general nodes (top-down approach), or with the most specific ontological concepts as leaves (bottom-up) or one could take a mixed approach, centering the development around some given nodes which seem to be essential to answer at least some of the competency questions prepared before.<sup>12</sup> As indicated above, we decided to start designing the ontology around the concept<sup>13</sup> “webshop”, entering more and more nodes as needed and creating an appropriate hierarchy along the way, thus, taking the mixed approach.

The question to be answered first was where to put the webshop node in the (not yet existing) hierarchy. Out of the dictionary’s expert knowledge the concept was identified as a specialisation of both shop and website. Thus, these two were inserted as hyperonyms to webshop (which in turn would then be a hyponym to shop and website). As

more nodes were added, the webshop finally found its place in the chain Division → Shop → Webshop for the business perspective and ICT → Software → Website → Webshop for the technical view. This way, a network with a strong hierarchical influence was developed (cf. Fig. 1).

To really be able to call this network an ontology and to answer more questions as indicated above, it needed more and different types of relations. Thus, in the next step whole-part relations (holonymy / meronymy) were introduced; so the shop received the concepts of product catalogue and checkout as parts. Over time, more whole-part relations as well as other types (e. g. uses / used by) were added all over the network. Although the ontology comprises all dictionary entries as nodes, not all possible triples of relation / source node / target node have been added due to restrictions to time and budget.

It has to be said here, that when designing this second ontology it became clear that only a minor part of the features provided by Protégé and OWL would be put to use. The dictionary was the main aspect of development and most project resources were put into completing it to specifications. The ontology remained a supporting element. It did receive some additional functions like cardinalities for relations and inheritance of attributes, not all of which were later used.

Nevertheless, the ontology was designed to provide some additional features for accessing the dictionary. As the courses were developed for use in the Moodle learning environment<sup>14</sup> it was planned to somehow integrate the dictionary into the system. This proved to be more or less unsuccessful, but the dictionaries’ contents were at least transformed to the Moodle glossary format which unfortunately does not allow for the extended dictionary features,<sup>15</sup> let alone ontological information.

Still, some kind of visualisation had to be found which would make the ontological information easily accessible. Of course, several tools for ontology visualisation exist and Protégé even comes with dif-

<sup>10</sup><http://protege.stanford.edu/>

<sup>11</sup>Web Ontology Language, <http://www.w3.org/TR/owl-ref/>

<sup>12</sup>NOY/MCGUINNESS pp. 6 ff.

<sup>13</sup>For more information on concepts, see SMITH (2004).

<sup>14</sup>Moodle is an open source course management system, for more information see [www.moodle.org](http://www.moodle.org).

<sup>15</sup>For this purpose, an XSLT script was engineered which integrated the textual information (i. e. synonyms, collocations, grammatical information) into the entry’s definition. This way, the information was contained, although not cleanly separated in a technical sense.

ferent plugins for this purpose. But all these solutions would have meant additional development efforts to integrate the dictionary and the ontology plus the necessity of a software installation on user side. It was therefore decided to take another approach: a web interface was created to replace the self-made DictionaryEditor which was used for work on the dictionary files and up to then had been the only tool to visualise the dictionary itself. This new tool – the DictionaryViewer – would not only show the dictionaries’ contents, provide links to the translations as well as a rudimentary search function but also display information from the ontology for each entry (cf. Fig. 2).

Here, for the first time inferred knowledge is being used – albeit in a somewhat rudimentary manner: attributes that have been given to nodes in the ontology are being passed down to the node’s hyponyms and evaluated for inference. To stick with the original example, `website` has an attribute `online = true`. This is passed on to the `webshop` node through inheritance. Additionally, from its “business side” `webshop` inherits the information that it uses a `payment system` (because each shop uses such a system and a `webshop` is a special kind of shop). This, combined with the information that there is a special version of a `payment system` – the `electronic payment system` which again has the `online` attribute set to `true` (inherited from `online system`) – can be used to infer that a `webshop` not only uses a `payment system`, but it’s special variant `electronic payment system` (cf. Fig. 3).<sup>16</sup>

## 5 Conclusions & Outlook

Although within strong limits of time and budget, the Arbeitsbereich Linguistik managed to develop a multilingual dictionary for e-commerce and not only demonstrate the use of an ontology to add further knowledge but also implement a basic viewing tool.

The developed ontology turned out to be entirely different from the first design which was based on the top-down approach. Although still strongly hierarchical due to the original development steps, the ontology soon turned into a network in which the relevance of a topmost node was rapidly declining.

During the ontology development, classes – sometimes even whole subtrees – have been shifted from one place to another as nodes on higher levels were being added. Fortunately, software tools can take care of all changes that unfold from this, in accordance with SMITH’s advice to “move gradually closer to the truth via an incremental process of theory construction, criticism, testing, and amendment.”<sup>17</sup> Thus, as with most ontologies, the e-commerce ontology is under constant construction and will remain far from being complete.

However, it was completed in respect to the size envisioned at the beginning of the project and thus contains now more than 1.000 nodes which can be used language-independently. The ontological network has a strong basis but needs to be enriched with additional relations to unfold its full potential.

Today the ontology is mainly used to support the dictionary. In the future it may also act as a knowledge base that can be integrated with other systems and ontologies.

As this paper describes a practical approach to creating an ontology, we do not consider creating a general top-level ontology a necessary task, but the development of the e-commerce ontology has made clear that an ontology, even as specific to a field as presented here, cannot act without representations of basic real-world concepts, e. g. actors in form of entities (humans, companies). These supporting nodes should be accumulated in a separate ontology which then might be used for connecting different ontologies from different fields.

The e-commerce ontology described here has not the most elegant or breathtaking design. It has been created to support the e-commerce dictionary. This it does well.

<sup>16</sup>To implement this kind of inheritance it became necessary to limit the search depth to guarantee a sufficiently short response time. As the DictionaryViewer is considered a proof-of-concept design not much effort was put into program design. A future version might e. g. make use of caching methods not implemented here. Also, the interface capabilities of the DictionaryViewer are fairly limited for now.

<sup>17</sup>SMITH Blackwell Guide to the Philosophy of Computing and Information 2003 p. 163.

## Bibliography

**Gruber, T. R.:** Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In **Guarino, N./Poli, R., editors:** Formal Ontology in Conceptual Analysis and Knowledge Representation. Deventer, The Netherlands: Kluwer Academic Publishers, 1993 (URL: <http://www-ksl.stanford.edu/knowledge-sharing/papers/onto-design.ps>)

**Noy, Natalya F./McGuinness, Deborah L.:** Ontology Development 101: A Guide to Creating Your First Ontology. 2001 (URL: [http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf))

**Smith, Barry; Floridi, Luciano, editor:** Ontology. Oxford: Blackwell, 2003, pp. 155–166

**Smith, Barry:** Beyond Concepts: Ontology as Reality Representation. In **Varzi, Achille/Vieu, Laure, editors:** Proceedings of FOIS 2004, International Conference on Formal Ontology and Information Systems. 2004 (URL: <http://ontology.buffalo.edu/bfo/BeyondConcepts.pdf>)

**Ziegler, Cai:** Smartes Chaos. Web 2.0 versus Semantic Web. iX, 11 2006, pp. 54–59



Figure 1: Hierarchical context for the Webshop node in the e-commerce ontology. Screenshot from Protégé using OWLViz visualisation plugin.



Figure 2: The representation of the ontological network for the node webshop in the DictionaryViewer.

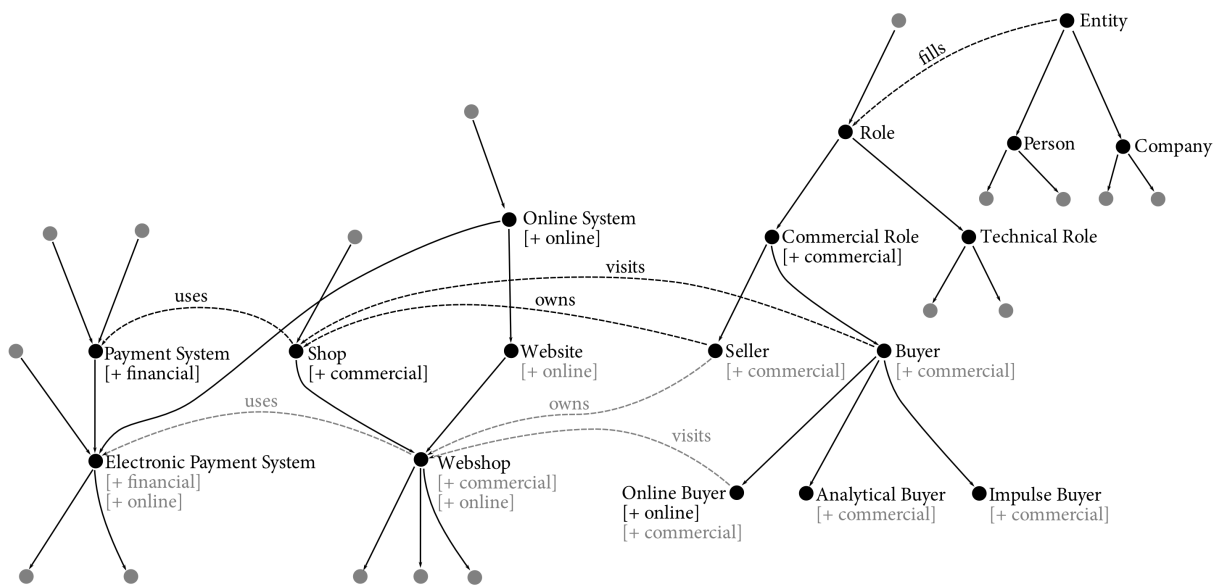


Figure 3: Excerpt of an ontological network with inheritance of attributes and relationships.