NLDB'01. Madrid. June 2001

WebPicker: Knowledge Extraction from Web Resources

Oscar Corcho, Asunción Gómez-Pérez
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo s/n. Boadilla del Monte, 28660. Madrid. Spain.
ocorcho@fi.upm.es, asun@fi.upm.es

Abstract. We show how information distributed in several web resources and represented in different restricted languages can be extracted from its original sources and transformed into a common knowledge model represented in XML using *WebPicker*. This information, which has been built to cover different needs and functionalities, can be later imported into *WebODE*, integrated, enriched and exported into different representation formats using *WebODE* specific modules. We show a case study in the e-commerce domain, using products and services standards from several organizations and/or joint initiatives of industrial and services companies, and a product catalogue from an e-commerce platform.

1 Introduction

The popularity of Internet and the huge growth of new Internet technologies have led in the last years to the creation of a great amount of e-commerce applications ([Fe00] [Be99]). Technology is not the unique key factor for the development of current e-applications: the context of e-commerce, and especially the context of B2B (Business to Business) applications, requires also an effective communication between machines.

Two extremely relevant factors contribute to semantic interoperability between machines: (1) a common language in which resources implied in the communication are specified, and (2) a shared knowledge model and vocabulary between systems that are present in the whole process. We call them the *syntactic* and *semantic* dimensions.

The first dimension has led to the creation of varied representation languages for the specification of web resources (XOL, SHOE, OML, RDF, RDF Schema, OIL and DAML+OIL). A comparative study of their expressiveness and reasoning mechanisms can be found in [CG00].

The semantic dimension is related to the knowledge model and vocabulary used by systems involved in the communication. In fact, the use of a shared knowledge model and vocabulary (an ontology) increases the semantic interoperability among information systems. Several standards and initiatives[1] came up in the previous years to ease the information exchange between customers and suppliers, and between different suppliers, by providing frameworks to identify products and services in global markets. However, the proliferation of initiatives reveals that B2B markets have not reached a consensus on coding systems, level of detail, granularity, etc. These issues are obstacles for the interoperability of applications that follow different standards (an application that uses UNSPSC codes cannot *understand* an application that uses e-cl@ss codes).

[1] UNSPSC (http://www.unspsc.org/), RosettaNet (http://www.rosettanet.org/), e-cl@ss (http://www.eclass.de/), NAICS (http://www.census.gov/epcd/www/naics.html), SCTG (http://www.bts.gov/programs/cfs/sctg/welcome.htm), etc.

Building large and consensuated knowledge models for e-commerce applications from scratch is both difficult and expensive. In this paper, we will focus on the process of importing three classifications (UNSPSC, e-cl@ss and RosettaNet) and an e-commerce catalogue into the *WebODE* workbench [Ar01]. This process is identified in a method of integration and enrichment of existing standards through the use of ontological mappings, which is described in [CG01] and summarized in section 2.

The sources of information used in this paper are represented in different restricted, structured languages. Relevant information can be retrieved from them, transformed into XML syntax [BPS98] and integrated using the common XML knowledge model of *WebODE*. These processes must be performed automatically, considering the low stability of these sources of information (they evolve fast), and have been implemented in a service called *WebPicker*. Their automatic treatment and configuration management capabilities ensure that information will be always maintained up-to-date.

Additionally, each standard or initiative has been created aiming to cover different needs and functionality. UNSPSC, e-cl@ss and NAICS are intended to cover all the products and services that can be offered by a provider, although they will be used in different markets – worldwide, German and North American-; RosettaNet covers the electronic equipment domain, and SCTG deals with goods that can be transported. This leads to the problem of integrating varied knowledge models in a common architecture, allowing the intra-operability of vertical markets in specialized domains and the inter-operability between different vertical markets (known as horizontal markets). Integration issues are out of the scope of this paper: a deep analysis can be found in [CG01].

Not only integration is important for our task, but also the enrichment of the resulting ontologies with additional information, by adding attributes, disjoint knowledge, relationships, etc., to the components of the ontologies. Finally, ontologies are exported into different formats and languages, so that they can be tractable by other systems. These issues are also out of the scope of this paper, and can be found in [CG01].

The paper is organized as follows: Section 2 provides a global view of the processes needed for extracting and enriching the information provided by standards, initiatives and catalogues. Section 3 describes the standards and initiatives we have selected for this study (UNSPSC, RosettaNet and e-cl@ss) and an e-commerce catalogue where the products that are sold can be linked to these classifications. It also describes the representation languages in which our sources of information are available. Section 4 presents *WebODE's* knowledge model and its XML syntax, together with the processes of automatic extraction of knowledge that have been implemented in *WebPicker*. Finally, section 5 presents the main conclusions extracted from this work.

2 A method for reusing standards and initiatives to create e-commerce ontologies

In this section, we explain the main steps of the method where the process of ontology acquisition and integration from web environments can be placed. The rationale of this method, its environment and its justification are presented with more detail in [CG01].

1. **Selection of standards, joint initiatives, laws, etc., of classification of products and services**. Relevant sources of information are selected from existing global or specific agreements on classifications of products and services. They usually provide a commonly agreed product taxonomy, whose depth goes from 2 to 5 levels.

2. **Knowledge models extraction**. *WebPicker* automates the process of knowledge extraction from the selected sources of information. They are adapted to the XML syntax of *WebODE*'s knowledge model and imported into the *WebODE* workbench.
3. **Design of a multi-layered knowledge architecture**. Relationships between components in the different taxonomies are identified, taking into account the main features of the selected sources of information (covering, globality, specificity, etc).
4. **Integration of knowledge models**. All the knowledge models previously imported into the *WebODE* platform are integrated in a layered architecture, using the relationships identified at the design phase.
5. **Enrichment of the integrated ontology**. Current standards do not include attributes for products, relations between products, disjoints nor exhaustive knowledge, functions, axioms, etc. They just represent taxonomies of concepts. Hence, they can be enriched with this extra information when possible.
6. **Ontology exportation**. The whole ontology or specific parts of it can be exported into different languages, so that they can be tractable by other systems.

This paper explains deeply steps 1 and 2 from this method, which will be presented in detail in sections 3 and 4.

3 E-commerce standards as knowledge models

In this section, we present three different proposals for the classification of products in the e-commerce domain: UNSPSC, e-cl@ss and RosettaNet. Although other similar approaches also exist and are available (NAICS, SCTG, etc), we have just selected the ones enumerated before to show the adequacy of our work in this context. We will also present an electronic catalogue from an e-commerce platform that can be fitted in the overall ontology architecture.

3.1 UNSPSC (Universal Standard Products and Services Classification Code)

UNSPSC is a general classification of products and services. Its coding system is organised as a five-level taxonomy of products, each level containing a two-character numerical value and a textual description. These levels are defined as follows:

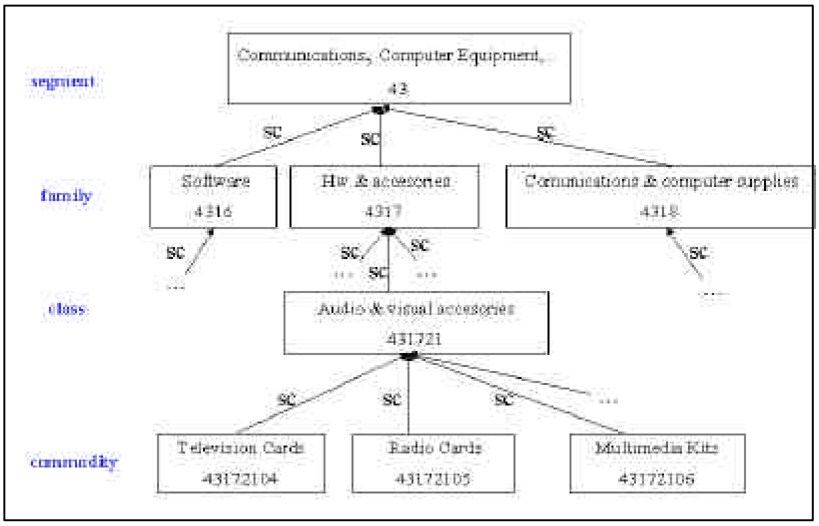

Fig. 1. Part of the classification of UNSPSC for computer equipment.

- **Segment.** The logical aggregation of families for analytical purposes.
- **Family.** A commonly recognised group of inter-related commodity categories.
- **Class.** A group of commodities sharing a common use or function.
- **Commodity.** A group of substitutable products or services.
- **Business Function.** The function performed by an organisation in support of the commodity. This level is seldom used.

Current version of the UNSPSC classification (version 4.04 on 13/02/2001) contains around 12000 products organized in 54 segments. Figure 1 shows part of segment 43 (in the computer equipment domain), which is visualized in static HTML as follows:

[43] Communications, Computer Equipment, Peripherals, Components and Supplies
 -family-[4316] Software
 -family-[4317] Hardware & Accessories
  -class-[431721] Audio & Visual Accessories
   -commodity-[43172104] Television Cards
   -commodity-[43172105] Radio Cards
   -commodity-[43172106] Multimedia Kits
 -family-[4318] Communications & Computer Supplies

The main drawbacks of UNSPSC are: (a) lack of vertical coverage of products and services that appear in the classification; (b) lack of attributes attached to the concepts that appear in the taxonomy; (c) design of the classification without taking into account the inheritance between the products that are described; (d) the non-providing different views of the classification, taking into account cultural and social differences, where classifications could be made in different ways than the ones presented in this standard.

3.2 e-cl@ss

e-cl@ss is a German initiative (similar to UNSPSC) for classification of material and services.

It consists of four concept levels (called *material classes*), with a UNSPSC-like numbering code (each level has two digits that distinguish it from the other concepts). The four levels are: *Segment, Main group*, *Group* and *Commodity Class*, which are almost equivalent to the first four ones provided in UNSPSC. Finally, we can have several products inside the same commodity class (in this sense, several products may

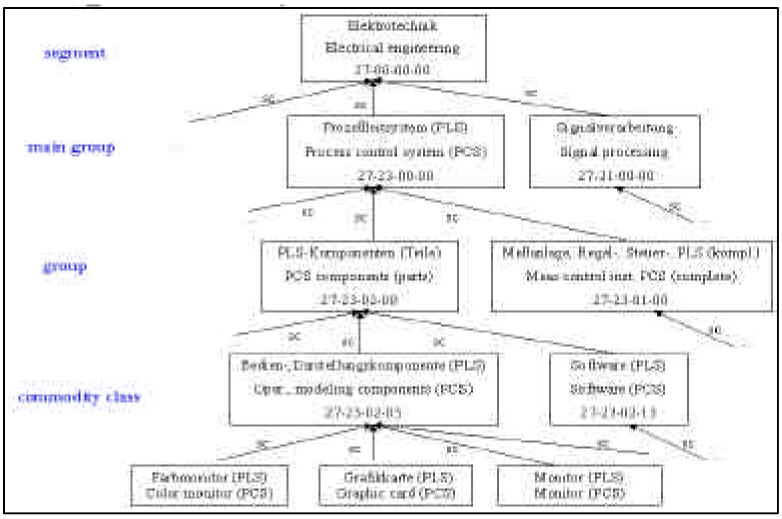

Fig. 2. Part of the classification of e-cl@ss for electrical engineering products (German and English).

share codes. This leads to a fifth level with all of them, as shown in figure 2.

Table 1 shows the source code (in Microsoft Excel format) that corresponds to the classification presented in figure 2. Although contents in the document are ordered and grouped by their codes, this order is not important, as it is implicitly given by the class code. There are no predefined relationships with other standards and names of classes are given both in German and English. Finally, a number is given with the depth of the element in the taxonomy: "S" means that it belongs to the last level and has no children.

Klasse/Class	Klassenbezeichnung	Class description	L
27-00-00-00	Elektrotechnik	Electrical-engineering	1
27-21-00-00	Signalverarbeitung	Signal processing	2
27-23-00-00	Prozeßleitsystem (PLS)	Process control system (PCS)	2
27-23-01-00	Meßanlage, Regel-, Steuer-, PLS (kompl.)	Meas.control inst. PCS (complete)	3
27-23-02-00	PLS-Komponenten (Teile)	PCS components (parts)	3
27-23-02-03	Bedien-,Darstellungskomponente (PLS)	Oper., modeling components (PCS)	4
27-23-02-03	Farbmonitor (PLS)	Color monitor (PCS)	S
27-23-02-03	Grafikkarte (PLS)	Graphic card (PCS)	S
27-23-02-03	Monitor (PLS)	Monitor (PCS)	S
27-23-02-13	Software (PLS)	Software (PCS)	4

Table 1. Source format for e-cl@ss classification of electrical engineering.

e-cl@ss contains around 12000 products organized in 21 segments. Segment 27, which deals with *Electrical Engineering*, contains around 2000 products. Finally, the main group 27-23, which deals with *Process Control Systems*, together with the main groups 24-01 to 24-04, which deal with *Hardware*, *Software*, *Memory* and other computer devices, contain around 400 concepts.

This classification suffers from the same drawbacks as UNSPSC. In fact, it is a similar approach, although within a smaller social environment: German companies. Finally, an advantage of this classification is the possibility of having terms, and their descriptions, in both English and German.

3.3 RosettaNet Technical Dictionary

RosettaNet Technical Dictionary is a classification of products in the electronic equipment domain. In contrast with UNSPSC, this classification does not use a numbering system: it is just based on product names, and it is also related to UNSPSC by providing the UNSPSC code for each product defined in it.

RosettaNet has just two levels in its taxonomy of concepts:

- **RN Category**. A group of products, such as *Video Products*.
- **RN Product**. A specific product, such as *Television Card*, *Radio Card*, etc.

RosettaNet consists of 14 categories and around 150 products. Figure 3 shows a small part of the RosettaNet classification, related to video products for computer equipment, and table 2 presents it in its original Microsoft Excel format:

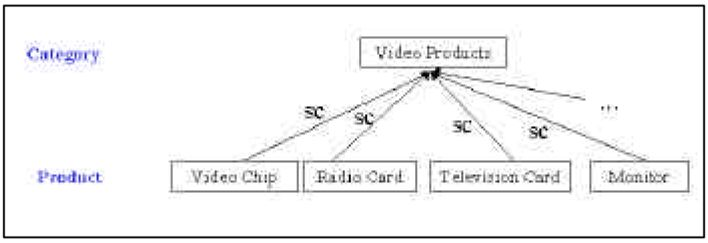

Fig. 3. Part of the classification of video products of the RosettaNet taxonomy.

RN Category Name	RN Product Name	UNSPSC Code	UNSPSC Code Name
Video Products			
	Monitor	43172401	Monitors
	Radio Card	43172105	Radio cards
	Television Card	43172104	Television cards
	Video Chip	321017	Hybrid Integrated Circuits

Table 2. Source format for RosettaNet classification of video products.

In contrast with the previous formats, the order of contents in this format is highly relevant: the relationship between products and the category they belong to is given by the order in which they appear in it. Hence, *Monitor*, *RadioCard*, *TelevisionCard* and *VideoChip* are products from category *VideoProducts*.

Each product in the RosettaNet classification has attached a UNSPSC code but not an e-cl@ss code. This value means that the RosettaNet product is equivalent to the corresponding UNSPSC family, class or commodity.

The main drawback of this taxonomy is that there are only two levels of classification, which implies that the structure of the taxonomy is very simple. Other problems are similar to those of UNSPSC, namely, lack of attributes and design without taking into account inheritance in the taxonomy of concepts. Finally, this classification is more suitable in a vertical market than the others, as it is focused on the specific domain of electronic equipment, although just offering a low level of detail in this domain.

3.4 E-commerce platform catalogue

We have selected a catalogue of products from an existing e-commerce platform that deals with computer equipment and is participating in the MKBEEM IST project.

This catalogue is structured, similarly to RosettaNet, in two kinds of elements: categories and items. Items correspond to actual products offered in the e-commerce platform and have attributes for their main characteristics. Categories are groups of products (items) or groups of other categories. They have no attributes and are just created with the aim of grouping products or other categories. Items can be classified in different categories (due to marketing purposes, functionality proximity, etc). These multiple classifications do not appear in the standards and initiatives presented before.

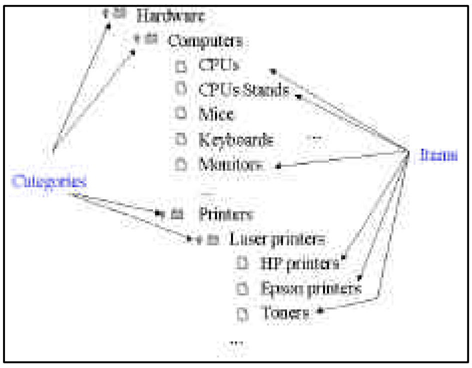

Fig. 4. Some elements in the catalogue.

The selected catalogue contains around 400 items, with 2/3 levels of depth in the hierarchy of categories. Figure 4 shows some elements in the catalogue. The next code presents part of them in XML, using <Category> and <Item> tags for representing categories and items. The XML file contains also classification relationships among categories and among categories and items. Attributes like *price*, *model*, *height*..., and their values are provided just for items.

```
<Category ID="C13423">  <Name>Computers</Name>  ...    </Category>
<Item ID="A23451">   <Name>CPUs</Name>              ...    </Item>
<Item ID="A44356">   <Name>Keyboards</Name>         ...    </Item>
...
```

```
<Belongs origin=”A23451” destination=”C13423”/>
<Belongs origin=”A44356” destination=”C13423”/>
...
<Property ID=”P332”>   <Name>price</Name>   <Type>float</Type>   ...   </Property>
...
<Value item=”A23451” property=”P332” val=”317”/>
```

In contrast with the previous classifications, catalogues cannot be considered as good sources of information for the development of ontologies, as they are not shared by a community nor represent any consensus. They are designed instead as classifications of products and services from the market (not knowledge) point of view. However, catalogues play an important role in the whole e-business process, as they present the set of products offered by each e-commerce application and they are the front-end in the exchange of products in B2C and B2B environments.

4 WebPicker. Discovering and transforming knowledge embedded in structured information

The standards, initiatives and catalogues presented in the previous section are expressed in different representation formats (HTML, XML and Microsoft Excel). They must be transformed into a common format (the XML syntax of WebODE's knowledge model) for their integration and enrichment.

In the first step of the process of transformation the sources of information must be analyzed. Many studies exist in the field of ontology learning where the process of knowledge discovery is performed (semi-)automatically ([MS00], [HS98], [Ki00], [ABS00]). However, instead of trying to automate the whole process, we decided to analyze the structure of information sources manually, identifying relevant information, and just automate the process of knowledge extraction and transformation using *WebPicker*. This decision has been made because the manual analysis is not difficult, as the sources of information are represented in a very restricted, structured language.

4.1 *WebODE*'s knowledge model.

WebODE is a workbench for ontological engineering [Ar01] that covers and gives support to most of the activities involved in ontology development, and provides a great range of services for ontology-based applications. Its knowledge model is extracted from the set of intermediate representations of METHONTOLOGY [Fe99], allowing for the representation of concepts (and their attributes), taxonomies of concepts, disjoint and exhaustive knowledge, ad-hoc binary relations between concepts, properties of relations, constants, axioms and instances.

As already described before, this knowledge model has been selected as a common format for information integration. This decision is partially based on the possibility of expressing conceptual models in its ad-hoc XML syntax, so that they can be imported later into *WebODE* using its import service, which uploads them into its database and allows their further edition and enrichment with the *WebODE's* ontology editor.

In this section, we just focus on concepts, and their attributes, taxonomies of concepts and relations. The rest of components are useful in the enrichment process.

Concepts are represented inside the tag *Concept*. Apart from their name and description, they can include their instance and class attributes, with name, description, type of value, minimum and maximum cardinality, measure unit and precision.

Relations are represented inside the tag *Term-Relation*. They include their name, description, origin and destination, and cardinality.

Taxonomies are mostly built using the built-in relation *subclass-of*. Other taxonomy relations also exist in *WebODE*: disjoint and exhaustive subclass partitions.

Everything that has been presented in the knowledge models of standards, joint initiatives and catalogues can be represented in the *WebODE*'s knowledge model.

4.2 Information transformation from distributed sources

This subsection focuses on the process of translation of the formats presented in the previous section, performed by *WebPicker*. The main conclusion that will be extracted is that very similar processes can be applied for both standards and catalogues to extract their relevant information, no matter what their scope or functionality is. This shows that these processes are general enough to be used for almost any kind of information extraction from structured languages and integration into another common language.

In the end, the XML import service of *WebODE* is used to upload the classifications into *WebODE*, so that we will be able to perform the necessary changes/upgrades to the original classifications using its ontology editor, rather than making changes directly in HTML, XML or Microsoft Excel formats.

UNSPSC. UNSPSC is available in HTML format (one HTML page per segment). We identified relevant pieces of information from this HTML code (code and description of each product/service) and identified the patterns they follow, so that knowledge extraction is automated and a first ad-hoc XML is generated. The resulting XML documents (one per segment) are integrated and transformed into a single document that uses *WebODE's* XML syntax. *WebPicker* uses the Document Object Model (DOM) for this transformation instead of XSLT, because of the complexity of changes.

Step 1

```
<Concept>
  <Code>431721</Code>
  <Name>Audio&Visual Accessories</Name>
</Concept>
<Concept>
    <Code>43172104</Code>
    <Name>Television Cards</Name>
</Concept>
```

Step 2

```
<Concept>
  <Name>431721</Name>
  <Description>Audio&VisualAccessories</Description>
</Concept>
<Concept>
  <Name>43172104</Name>
  <Description>Television Cards</Description>
<Concept>
...
<Term-Relation>
  <Name>Subclass-of</Name>
  <Origin>43172104</Origin>
  <Destination>431721</Destination>
  <Maximum-Cardinality>0</Maximum-Cardinality>
</Term-Relation>
```

The reason for this 2-step transformation resides in the modularity and easy treatment that it provides. Indeed, the first step just extracts the relevant information from the HTML pages, and creates a very simple representation, where just codes and names are included. In the second step, we transform all of them into a single document and add the *subclass-of* relationships that can be extracted from their numbering codes.

e-cl@ss. e-cl@ss is also available in Microsoft Excel format. After exporting the information to a text document, this document has been transformed using *WebPicker*.
Relevant information consists of the numerical code of products (specified in the first column) and two class attributes, *German* and *English name*, whose value is specified in

second and third columns. Finally, we take care of the level attribute in the case its value is “S”, as we have to create a subclass-of relationship between the product and the product whose code is specified in the first column.

```
<Concept>
 <Name>27-00-00-00</Name>
 <Class-Attribute>
  <Name>German name</Name>
  <Type>String</Type>
  <Minimum-Cardinality>1</Minimum-Cardinality>
  <Maximum-Cardinality>1</Maximum-Cardinality>
  <Value>Elektrotechnik</Value>
 </Class-Attribute>
 <Class-Attribute>
  <Name>English name</Name>
  <Type>String</Type>
  <Minimum-Cardinality>1</Minimum-Cardinality>
  <Maximum-Cardinality>1</Maximum-Cardinality>
  <Value>Prozeβleitsystem (PLS)</Value>
 </Class-Attribute>
</Concept>
...
<Term-Relation>
 <Name>Subclass of</Name>
 <Origin>27-23-00-00</Origin>
 <Destination>27-00-00-00</Destination>
 <Maximum-Cardinality>0</Maximum-Cardinality>
</Term-Relation>
```

RosettaNet. RosettaNet is represented in Microsoft Excel format. We have just exported it to text and transformed it into the XML syntax of *WebODE* using *WebPicker*.

Extracted information from this source consists of category/product name, a class attribute called *UNSPSC code*, whose value is specified in the fourth column and which is used in the integration phase, and the taxonomy of concepts, imposed by the order and grouping in which products appear. This leads to a 2-level classification of products.

We show an example of final code with a category and a product:

```
<Concept>
 <Name>Video Chip</Name>
 <Class-Attribute>
  <Name>UNSPSC Code</Name>
  <Type>String</Type>
  <Minimum-Cardinality>1</Minimum-Cardinality>
  <Maximum-Cardinality>1</Maximum-Cardinality>
  <Value>321017</Value>
 </Class-Attribute>
</Concept>
<Concept>
 <Name>Video Products</Name>
</Concept>
...
<Term-Relation>
 <Name>Subclass of</Name>
 <Origin>Video Chip</Origin>
 <Destination>Video Products</Destination>
 <Maximum-Cardinality>0</Maximum-Cardinality>
</Term-Relation>
```

Transforming the catalogue into the *WebODE* knowledge model. This task consists of the transformation between two XML models.

The catalogue provides more relevant information than the rest of sources of information that we have used. In fact, it provides concepts (categories and items), subclass-of relationships (represented with the *Belongs* tag) and class attributes (whose description is given inside tag *Property* and whose value is given inside tag *Value*).

```
<Concept>
 <Name>CPUs</Name>
 <Description>A23451</Description>
 <Class-Attribute>
  <Name>Price</Name>
  <Description>P332</Description>
  <Type>Float</Type>
  <Minimum-Cardinality>1</Minimum-Cardinality>
  <Maximum-Cardinality>1</Maximum-Cardinality>
  <Value>317</Value>
 </Class-Attribute>
</Concept>
<Concept>
 <Name>Computers</Name>
 <Description>C13423</Description>
</Concept>
...
<Term-Relation>
 <Name>Subclass of</Name>
 <Origin>CPUs</Origin>
 <Destination>Computers</Destination>
 <Maximum-Cardinality>0</Maximum-Cardinality>
</Term-Relation>
```

In this case, *WebPicker* uses XSLT to derive the final code from the source code, as transformations to be made are not so complex as the ones performed with UNSPSC.

5 Conclusion

Ontologies play a crucial role on the construction of the Semantic Web, because they provide a shared conceptualization of the knowledge and services available on the web in a machine-readable way, allowing the information sharing between heterogeneous systems. In this paper, we have put our attention onto a specific area of the Semantic Web: the world of e-commerce applications (both B2C and B2B).

We have focused on the process of automatic knowledge acquisition (KA) from resources available on the web, ameliorating the KA bottleneck when building ontologies from scratch. We have chosen as our sources of information several standards and joint initiatives for the classification of products and services (built by different organisms) and a catalogue from an e-commerce platform, and have shown that it is possible to extract the relevant information of the sources of information by:

- Performing a previous analysis of the languages in which they are represented.
- Identifying relevant pieces of information.
- Implementing the necessary conversion mechanisms for transforming it into another single restricted language/model (*WebODE's* knowledge model XML syntax).

Once the information has been uploaded into the *WebODE* workbench, its ontology editor allows its integration, its enrichment and its partial or global exportation into another representation format. This issue is out of the scope of this paper.

Resulting ontologies will be useful for tasks such as product searches in e-commerce platforms, information or product exchange in B2B platforms, catalogue creation, etc.

Acknowledgements

This work is supported by a FPI grant funded by UPM and by the project ContentWeb, funded by MEC. We also want to thank Alberto Cabezas for implementing *WebPicker*.

Bibliography

[Ar01] Arpírez, J.C., Corcho, O., Fernández-López, M., Gómez-Pérez, A. *WebODE: a workbench for ontological engineering*. Submitted to K-CAP 2001.

[ABS00] Aussenac-Gilles, N., Biébow, B., Szulman, S. *Corpus analysis for conceptual modelling*. Workshop on Ontologies and Texts. Juan-Les-Pins. France. October 2000.

[Be99] Berners-Lee, T., Fischetti, M. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. Harper. San Francisco. 1999.

[CG00] Corcho, O., Gómez-Pérez, A. *A RoadMap to Ontology Specification Languages*. EKAW'00. Springer-Verlag. 2000.

[CG01] Corcho, O., Gómez-Pérez, A. Solving integration problems of e-commerce standards and initiatives through ontological mappings. IJCAI'01 Workshop on Ontologies and Information Sharing. Seattle. August, 2001.

[Fe00] Fensel, D. *Ontologies: silver bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag. Berlin. 2000.

[Fe99] Fernández, M.; Gómez-Pérez, A.; Pazos-Sierra, A.; Pazos-Sierra, J. *Building a Chemical Ontology Using Methontology and the Ontology Design Environment*. IEEE Intelligent Systems & their applications. January/February 1999. PP. 37-46.

[HS98] Hahn, U., Schnattinger, K. *Towards Text Knowledge Engineering*. AAAI'98 - 15th National Conference on Artificial Intelligence. Madison. Wisconsin. July 26-30, 1998,

[Ki00] Kietz, J., Maedche, A., Volz, R. *A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet*. Workshop on Ontologies and Texts. Juan-les-Pins. 2000.

[MS00] Maedche, A., Staab, S. *Semi-automatic Engineering of Ontologies from Text*. 12th Intl Conference on Sw Engineering and Knowledge Engineering (SEKE'2000). 2000.