Computers in Industry 72 (2015) 1–13

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Managing technological knowledge of patents: HCOntology, a semantic approach

Maria Bermudez-Edo [a,*], María V. Hurtado [b], Manuel Noguera [b], Nuria Hurtado-Torres [c]

[a] University of Surrey, Institute for Communication Systems (ICS), Guildford, United Kingdom
[b] University of Granada, Department of Software Engineering, E.T.S.I.I., c/Saucedo Aranda s/n, 18071 Granada, Spain
[c] University of Granada, Department of Management, Business School, Campus Cartuja s/n, 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 9 March 2015
Accepted 13 March 2015
Available online 15 May 2015

Keywords:
Ontology
OWL
Patent
IPC classification

ABSTRACT

Patent data provide technological information essential to define strategies and decisions in the context of firm innovative processes. At present, information regarding patents is usually represented and stored in large databases. Information from these databases is commonly retrieved in the form of files with a CSV- or XML-based codification but with little semantics that enable the inference of further relationships among patents. In these databases, each patent is associated with a technological field by a code. Although the codes assume a hierarchical classification approach, inclusion/subsumption relationships are not explicitly specified such that computers can process them automatically. In recent years, ontologies have been proven to facilitate the exchange of information between people and systems. In this context, the Web Ontology Language (OWL), whose formal semantics are based on description logics, has become the most widely used language for the representation of ontologies. Certain patent ontologies have already been developed in OWL to benefit from the semantics of patent information. However, none have fully exploited the information that can be derived from the formal representation of patent code classification hierarchies through description-logics-based reasoning. This paper presents an approach to automatically translate the hierarchies found in the patent classification codes into concept hierarchies. This proposal also enables the automatic inference of implicit knowledge based on reclassification techniques and relationships between different application domains without changing the applications that make use of patent information. Several examples are presented to illustrate the applicability of the proposal and how it can assist firms in patent information management.

1. Introduction

Innovation is an important driver of both firm success and national economic growth. In the context of a growing necessity for investments that develop innovation, patents are relevant tools for protection of the firm's innovative ideas. Patents are legal documents that protect the rights of the inventor of an industrial property. A patent document provides different information about the innovation: the applicant (or patenting firm), geographic location, date, the technological field, and some information about the other patents that it cites, among others. All of these data that describe or are related to patent documents are called patent metadata [1], and used with other data they are particularly useful to analyze dimensions of innovation that are of policy interest, such as the relationship between patents and economic performance, entrepreneurship, etc. [2,3].

Patent documents are usually stored in large databases of patent offices (e.g. those maintained by the European Patent Office (EPO) and those under supervision of the United States Patent and Trademark Office (USPTO)). These databases often use different data structures that make it difficult to interoperate or automatically and efficiently process the information contained therein. This difficulty is even more important in the context of the different objectives (informative and infringement searches, etc.) of the agents using these databases (e.g. patent lawyers, managers of start-up firms, scientists, and managers of competitors in the industry) [4].

In this context, ontologies have proved to be useful for sharing information by providing formal, uniform and shareable representations about a domain. Several pieces of work have proposed

* Corresponding author. Tel.: +44 1483 689760.
E-mail address: m.bermudez@surrey.ec.uk (M. Bermudez-Edo).

http://dx.doi.org/10.1016/j.compind.2015.03.010
0166-3615/

the use of an ontology-based approach to represent patent metadata using the Web Ontology Language (OWL) (e.g. [1,5]). The main objective of these works is to provide a semantically, well-defined and homogeneous representation of the main types of patent metadata. The use of ontologies enables the representation of knowledge and the identification of context and dependency information more easily than using database-centric structures and interfaces [1]. Likewise, ontologies are increasingly being used in several domains related to innovation and patent registration (Ahmet et al., 2011), including economic and management business areas such as logistics [6], knowledge-commerce [7], e-commerce [8], team organization [9], manufacturing [10] or business processes [11–13]. Furthermore, the power of reasoning in ontologies allows different domains, represented through ontologies, to be integrated, such as patents associated with economics and management studies [14].

The technological field of a patent document represents relevant metadata information about the patent. Patent databases use codes for technological fields that assume a hierarchical classification that delimits the categories a patent may pertain to or be associated with. This field is widely used in searches in databases to determine the field(s) in which a firm may infringe upon another company's industrial rights or where there exists a gap in the technology in which a company could innovate. However, in the patent databases, this hierarchy is not explicitly described and cannot be automatically processed by computers.

Previous patent ontologies have also represented technological fields by using codes. However, they do not fully exploit the formal representation of patent classification hierarchies of these technological fields, and basically mirror the technological patent codes of patent offices databases without leveraging further reasoning capabilities.

In summary, the motivations of this paper in relation to patent information and ontologies are:

- Technological patent codes are defined in patent databases according to hierarchical classifications of technological patent fields, but these hierarchies are not explicitly specified for automatic processing by computers.
- Previous pieces of work on patent metadata ontologies do not represent the hierarchies of the technological patent codes either.
- In current patent ontologies it is not possible to infer transversal relationships between the different metadata that describe a patent document through the hierarchies of concepts of technological patent codes.

This paper draws on the well-known Noy and McGuinness's methodology [15] to develop the proposed ontology by automating some of the steps and focusing on hierarchical codes. Specifically, this paper proposes a method to automatically build and populate patent metadata ontologies by indexing hierarchical codes, which can be retrieved from different patent repositories, by defining ontological categories which enrich the information retrieval process with new relationships, properties and enable the inference of new knowledge. In particular, the paper first studies the characteristics of the hierarchical codes, such as the structure they follow. Second, we propose a representation of these hierarchical codes in OWL. And third, the hierarchy of the technological patent codes is translated from XML into the formal Web Ontology Language OWL 2 using XSLT (eXtensible Stylesheets Transformation Language) [16], according to the characteristics of the hierarchical codes, and the representation proposed.

Additionally, we show the power of reasoning that OWL offers in combination with the hierarchy of concepts (OWL classes) that we have created to infer new information in different fields through transversal relations among concepts from different (but related) domains. The aim is to benefit from linking patent information with external information by providing several methods to infer knowledge in different fields and to connect different knowledge sources. We use OWL-DL, which is the sublanguage of OWL that support maximum expressiveness without losing computational completeness. We use the methodology proposed in Bermudez-Edo et al. [17] for inferring new knowledge with existing reasoning tools.

To illustrate the applicability of our proposal, this paper shows how new information can be inferred from the hierarchy of patent metadata concepts. In particular, the first case study demonstrates how a suitable ontological representation of patent metadata enables the automatic reclassification of patents when a new technological patent code appears. In the second case, a proposal is introduced to link two different knowledge domains by specifying new relationships between the representation of patent classifications and an external classification, the classification of industrial sectors provided by the United Nations Statistics Division (UNSD),[1] which can help in the elaboration of economics studies that evaluate the innovation level of firms based on patent indicators.

The remainder of the paper is organized as follows. Section 2 describes the related work. Section 3 explains our proposal on how to translate the hierarchical technological patent codes into hierarchies of concepts and accomplish the population of the ontology. Section 4 presents motivating examples of the reasoning and the inference of new information from the hierarchical codes within the patent domain and in multiple domains. Section 5 discusses the contributions of the research. Finally, Section 6 concludes with a short summary and proposals for future research.

2. Patent ontology related works

Several patent ontologies have been proposed so far for a semantically well-defined and homogeneous representation for the major types of patent metadata. The most prominent examples are the ontology created within the European *Patexpert* project [1,18] and the *PatentOntology* from Stanford University [5].

Patexpert was created to homogeneously represent different patent information from several EPO databases and to provide it with semantic meaning. However, Patexpert does not merge information retrieved from different patent offices. The patent metadata ontology has been populated by XSLT stylesheets. Unfortunately, the public version of this ontology[2] is not populated, but to the best of our knowledge this ontology does not automatically represent the semantics of the hierarchy of technological patent codes.

PatentOntology was developed to avoid the limitations of Patexpert when integrating heterogeneous domains [19]. PatentOntology merges information from USPTO patent documents retrieved from the USPTO database with information from patent courts of USPTO from the LexisNexis database [20]. Although this ontology has been populated with a parser, it does not automatically retrieve the semantics of the hierarchy of technological patent codes and does not merge information from different patent offices. The technological patent codes classify innovations into fields of activity to facilitate the searches of interested agents. This technological field is widely used to delimit the scope of the searches and is one of the most used items of patent metadata [4].

There are also patent ontologies based only on keywords found in patents [14,21,22]. These ontologies allow retrieving information from the text of the patent documents, however they do not

[1] http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27.
[2] http://mklab.iti.gr/project/patexpert.

represent patent metadata and, hence, they do not exploit the potential of reasoning using the technological patent codes.

3. An approach for the analysis of hierarchical codes using patent ontologies

In this section, we present a patent description and analysis method based on an ontological representation of technological patent codes. In practical terms, developing an ontology includes: defining classes, organizing the classes in a taxonomy, defining slots and values, facets of the slots and creating individual instances of the classes. Our proposal draws on the methodological proposal of Noy and McGuinness [15] to develop ontologies, which is the most popular methodology in the semantics literature. Noy and McGuinness propose 6 steps: ‘‘Determine the domain and scope of the ontology’’ (Step 1), ‘‘Consider reusing existing ontologies’’ (Step 2), ‘‘Enumerate important terms in the ontology’’ (Step 3), ‘‘Define the classes and the class hierarchy’’ (Step 4), ‘‘Define the properties of classes’’ (Step 5), ‘‘Define the facets of slots’’ (Step 6) and ‘‘Create instances’’ (Step 7). Our method proposes to automate steps 4 and 7 with special focus on hierarchical codes. To this end we will need to redefine step 3 with new terms. In step 1 we have already spot the patent domain and in step 2 we have decide to reuse ‘‘PatentOntology’’. This ontology has already implemented steps 3–7, except that it does not take into account the hierarchy of the technological codes. We will automate step 4 and 7 for the inclusion of the hierarchies of the technological codes into the ontology ‘‘PatentOntology’’, and redefine some concepts of step 3, when needed.

In order to automate step 4, our method begins by splitting the codes of the technological patent fields into their constituent parts (parts of the code) that aid in inferring a hierarchy of concepts as a final output; as far as we know, previous patent ontologies have not fully exploited the potential to automatically create a hierarchy of concepts in an ontology from the hierarchy represented in hierarchical codes.

To understand the technological codes, we have to redefine some concepts in step 3 and study the codes. Different classifications of technological patent codes exist, such as the CPC (Cooperative Patent Classification), which has been recently created as the union of US classification and the European ECLA (European CLAssification) and ICO (In Computer Only) classifications. The World Intellectual Property Organization (WIPO) has also defined the International Patent Classification (IPC), and most of the patent offices classify patents with their own classification as well as with IPC. This paper will use the IPC codes for general application, and the CPC (the old ICO codes part), which are based on the IPC codes and have the same underlying structure, for a specific classification of ‘‘green patent technology’’. We will explain this term in Section 4.1.

IPC codes are divided into 5 parts:

1. Section, which is represented by a letter (e.g. *A*, meaning *human necessities*, to *H*, for *electricity*).
2. Class, which is represented by 2 digits.
3. Subclass, which is represented by one letter.
4. Main group, represented by 1–3 digits.
5. Group, represented by at least 2 digits.

For example, the IPC code *H04L25/02* is composed of section (*H*), class (*04*), subclass (*L*), main group (*25*) and group (*02*). As such, the complete code *H04L25/02* means:

H	*electricity*
H04	*electric communication techniques*
H04L	*transmission of digital information*
H04L25	*baseband systems and*
H04L25/02	*shaping networks.*

To automatically create a classification hierarchy of concepts in a patent ontology and accomplish the automation of steps 4 and 7, this method includes the following phases:

1. The first phase is to study the technological patent codes (IPC) and identify certain characteristics; this study considers two important characteristics of technological patent codes, which seem to have been disregarded in previous patent ontologies literature. This phase is detailed in Section 3.1 and will be a necessary study to find important terms of step 3 in Noy and McGuinness methodology.
2. The second phase is to provide a patent ontology for the analysis of the hierarchical technological patent codes, addressing the features found in phase 1, which is covered in Section 3.2. This corresponds with step 3 of Noy and McGuinness methodology.
3. The third step is to propose an automatic mechanism to populate the patent ontology, which is discussed in Section 3.3. This corresponds with the automation of step 4 and 7 of Noy and McGuinness methodology.

3.1. *Characteristics of the technological patent codes*

3.1.1. *Importance of the parts of the codes*

The first important characteristic that IPC codes exhibit is related to the meaning of each part of the codes. When dividing the code into its parts, the meaning of the individual parts can be different even when their representations are the same. For example, the IPC code *H03K3/03* means:

H	*electricity*
H03	*basic electronic circuitry*
H03K	*pulse technique in basic electronic circuitry*
H03K3	*circuits for generating electric pulses*
H03K3/03	*astable circuits for generating electric pulses*

This code is defined as ‘‘*dealing with astable circuits for generating electric pulses*’’. Dividing it into its parts – section (*H*), class (*03*), subclass (*K*), main group (*3*), subgroup (*03*) – we can see that the class and the subgroup have the same value (*03*). Although they have the same representation (*03*), they have different meanings because of their respective position in the code. One represents the class ‘‘*basic electronic circuits*’’, and the other one represents the subgroup ‘‘*astable circuits for generating electric pulses*’’.

Therefore, when translating IPC codes into a hierarchy of concepts, one characteristic is that the same individual (in the example, *03*) has different meaning depending on the part of the code to which it belongs (in the example, the class or the subgroup).

3.1.2. *Importance of the context of the codes*

Another important characteristic found in this study of IPC codes is that the meaning of the representation of one part of the code depends on the previous parts.

For example, lets consider the IPC code *H02K3/02*, where:

H	*electricity*
H02	*generation, conversion or distribution of electric power*
H02K	*dynamo electric machines*
H02K3	*details of winding in dynamo machines*
H02K3/02	*winding characterized by the conductor material*

By splitting the code *H02K3/02* into its parts – section (*H*), class (*02*), subclass (*K*), main group (*3*), and subgroup (*02*) – and comparing the codes in both examples, (*H03K3/03*) and (*H02K3/02*), we see that both of them have the subclass (*K*), but the meaning of the subclass (*K*) is different in each case because its meaning depends on the previous parts of the code.

Therefore, we have found another characteristic that must be taken into account when translating IPC codes into a hierarchy of concepts; that is, the need to put the code into context to understand the meaning of each part. In particular, it is necessary to consider not only the value of a particular part of the code, but also the values of the previous parts.

Hence, to create a hierarchy of concepts in the ontology, it is crucial to make an accurate translation of the hierarchical codes.

3.2. HCOntology

This subsection provides a translation of the IPC hierarchy of the technological patent codes into a hierarchy of concepts in an ontology.

The concepts of the ontology are used to describe the semantics of the parts of the codes and to make their meaning explicit. Noy and McGuinness [15] describe two approaches to define the hierarchy of the concepts in the ontology, bottom-up and top-down approach. We use the top-down approach due to the characteristics of the codes, that define the terms for more general to more specific.

First, a hierarchical classification in the ontology has been designed taking into account all of the IPC sections (*IPC_A*, *IPC_B*, *IPC_C*, *IPC_D*, *IPC_E*, *IPC_F*, *IPC_G* and *IPC_H*). Next, for each IPC section, the subclasses for all IPC classes of the corresponding section (e.g. *IPC_H03*) have been created. In the same way, the IPC subclasses (e.g. *IPC_H03K*) and main groups (e.g. *IPC_H03K03*) have been created, and the subgroups are the individuals (e.g. *H03K03_03*), creating a large hierarchy of classes and individuals (from now on referred to as HCOntology – Hierarchical Codes Ontology; see Fig. 1). The figures of the ontologies in the examples have been represented by using the ontology editing tool Protégé [23–25]. This approach allows us to make explicit the hierarchy found in the technological patent codes and enable automatic processing, filling the gaps in the previous patent metadata ontologies.

This design is intended to address the two characteristics mentioned in the previous section. The hierarchy of classes in HCOntology considers the first characteristic (Section 3.1.1) because one part of the code is a child class belonging to the previous part of the code, and therefore, the meaning of both parts are different. In the example of Section 3.1.1, the individual *H03K03_03* represents the subgroup (*03*), and *IPC_H03* represents the class (*03*). However, because the individual *H03K03_03* inherits *IPC_H03* and they belong to different classes, the meanings of both are well differentiated.

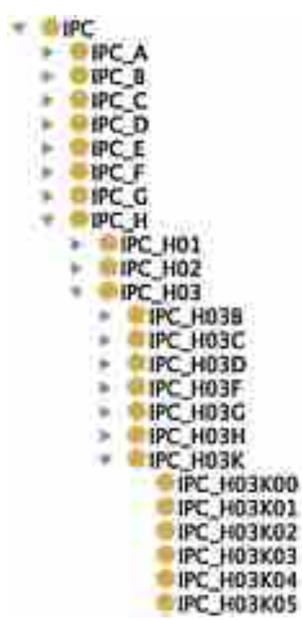

Fig. 1. HCOntology in Protégé.

The proposed HCOntology also addresses the second characteristic. In the example of Section 3.1.2, the class *IPC_H03K* inherits the properties of the superclass *IPC_H03*, and another class, *IPCH02K*, inherits the meaning of the superclass *IPC_H02*. Therefore, the meanings of the two codes are well differentiated because the hierarchy of classes takes into account the values of the previous parts of the code.

Furthermore, HCOntology allows reasoning and using the classes and the semantics of their hierarchy. For example, the individual *H03K03_03* inherits all of the properties from its parent classes, *IPC_H03K03*, *IPC_H03K*, *IPC_H03* and *IPC_H*. Hence, this individual will be found in the searches of IPC codes *H03* because it belongs to this parent class.

After introducing the new hierarchy of classes, in the next section, we will implement an automatic mechanism that populates the ontology (the automated step 7 in our methodology) from query responses of the patent databases, and automatically creates the hierarchy of classes described in Fig. 1 (our automated step 4).

3.3. Populating the patent ontology and HCOntology

Patent databases provide their results to queries in different formats, one of which is XML. We took advantage of this result format to create and populate a patent ontology with the contents of a database using stylesheets.

Stylesheets are XSL documents that with the help of an XSLT processor can convert XML files into OWL files. This section presents a technical solution that automates the processes of creating and populating the ontology, parsing the query results from different databases in XML format. The XML resulting files are converted into OWL by means of the corresponding XSL files and an XSLT processor. Fig. 2 shows an overview of the system proposed. The first step of this process takes a XML file from the results of a search in the patent database. In the second step, a stylesheet to translate the XML tags into ontology classes and properties, including the IPC classes and subclasses, is implemented. In the third step, with the help of an XSLT processor, the XML individuals are translated into OWL individuals using our XSL file.

The second step requires to address the correspondence between the XML tags and the OWL classes using a stylesheet [18], and to extract the individuals used to populate the ontology. In the case of HCOntology, we will also need XPATH (XML Path language) [26] to point at the different parts of the code. Additionally, this step allows the extraction of each part of the code one-by-one, navigates inside HCOntology, and inserts the IPC code in the class to which it belongs. We will describe this stylesheet in detail in next Section 3.3.1, and provide and example of stylesheet in Section 3.3.2.

3.3.1. Customization of stylesheets for hierarchical codes

The resulting stylesheets will allow the automatic translation of XML instance documents (where the instances are hierarchical codes) into an OWL ontology model and instances with the process shown in Fig. 2. Previous work has created XSD documents for translating XML instances into OWL models and/or instances [1,5]. However, none of these proposals has studied the hierarchical codes. While other work translates one XML instance into one OWL instance (individual) [27,28], our work, on the other hand, proposes the translation of one XML instance into several

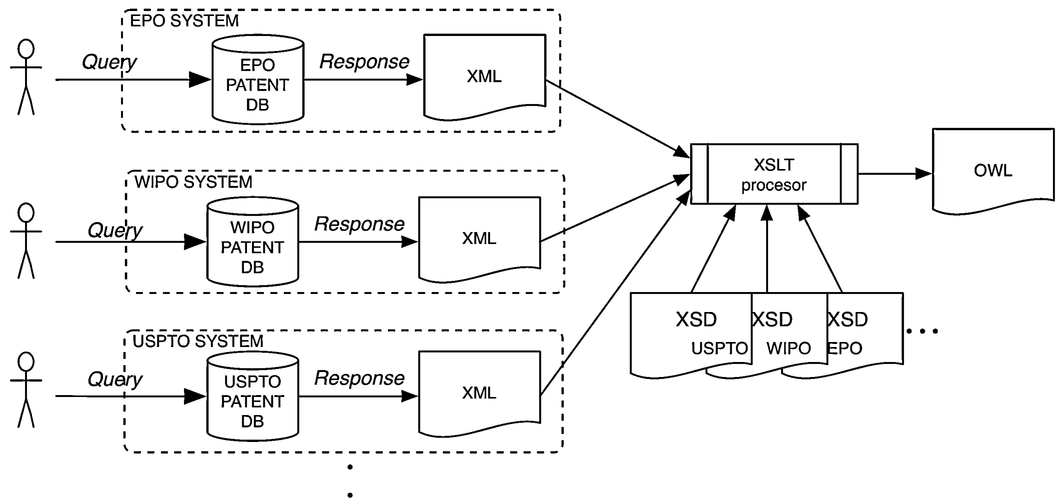

Fig. 2. System overview.

OWL labels (several classes) and one instance (individual), following the implicit hierarchy in the XML instance.

The customization of the stylesheets consists of the following steps:

1. Study the codes and define their parts, identifying the number of characters or digits in each part and whether they have separating characters
2. Study the structure of the labels in the XML and OWL files
   2.1. Location of the hierarchical codes in the XML file
   2.2. Detect the class from which implementing HCOntology in the OWL file
3. Write the XSL file
   3.1. Write the regular header of the OWL files with the namespaces. If HCOntology is implemented on top of an existing ontology, it is also needed to import the existing ontology in the header
   3.2. Clean each label of the XML file, deleting unnecessary spaces, ensuring that each code is only written once in the OWL file, even if the code is repeated in several places in the XML file, etc.
   3.3. Create the hierarchical code structure
   3.3.1. Define the whole code except the last part as a subclass of the previous part of the code
   3.3.2. Repeat Step 3.3.1 for each part of the code until the first part of the code, and define the first part of the code as a subclass of the class of step 2.2
   3.3.3. Insert the individuals (the whole code) in the corresponding class of Step 3.3.1
4. Close the open labels

3.3.2. Example of customization of stylesheets

This section shows the applicability of the customization of the stylesheets proposed in the previous subsection. The EPO Open Patent Services database is used in this case study to download the XML file. We need to use XPATH functions related to substrings to split the ICO codes into its parts and create the corresponding ontology classes, subclasses and individuals for each code. As we mention in Section 2, the ICO codes follows the same structure as the IPC codes. Following the previous section, creating a XSL file to represent the ICO codes needs:

1. Study the ICO codes (for example the ICO instance in the XML file, *Y02E10:20*). The ICO codes have the following structure:
   - Section: one character (*Y*)
   - Class: two digits (*02*)
   - Subclass: one character (*E*)
   - Main group: from one to three digits (*10*)
   - Group: at least two digits (*20*)
   - The mark ":" separates the main group from the group in the XML file
2. Study the structure of the labels in the XML and OWL files
   2.1. In the XML file, the ICO codes are stored under the labels: *<RESULT-LIST> <ROW> <ICO> <p>*
   2.2. In the OWL file, implement HCOntology as subclasses of http://www.semanticweb.org/HCOntology#*ICO*
3. Build the XSL file
   3.1. In the header of the OWL file, write, as usual in OWL files, the RDF namespace envelope and the ontology elements [29]
   3.2. Clean the labels in the XML file that contains the hierarchical code, delete the empty spaces and define each code only once
   3.3. Create the hierarchy of classes of the ICO codes. For example, as shown in Fig. 3, the code *Y02E10:20* should have an OWL instance *ICO_Y02E10-20*, with the previous OWL hierarchy of classes (*ICO_Y*, *ICO_Y02*, *ICO_Y02E*, *ICO_Y02E10*). Listing 1 shows the result of this step in the XSD file
   3.3.1. Define the whole code except the last part of it as a subclass of the previous part of the code. For example, from the XML instance *Y02E10:20*, take the part *Y02E10* and make it a subclass of *Y02E*, thereby creating the class *ICO_Y02E10*. Afterwards, create the previous subclass (*ICO_Y02E*) and then define one as a subclass of the other (*ICO_Y02E10* subclass of *ICO_Y02E*)
   3.3.2. Repeat Step 3.3.1 with the rest of the code using suitable XPATH functions. In the example, it implies the creation of the class *ICO_Y0E* and to define it as a subclass of *ICO_Y02*, to create the class *ICO_Y02* and to define it as a subclass of *ICO_Y*, to create the class *ICO_Y* and to define it as a subclass of *ICO*
   3.3.3. Add the individual (the instance) in the corresponding class. In the example, it implies adding the instance *ICO_Y02E10-20* to the class *ICO_Y02E10*
   3.4. Close all the labels that remain opened

With these stylesheets, the XSLT processor will automatically create a hierarchy of classes in OWL from a single code in XML. Technological patent codes represent technologies, and as the

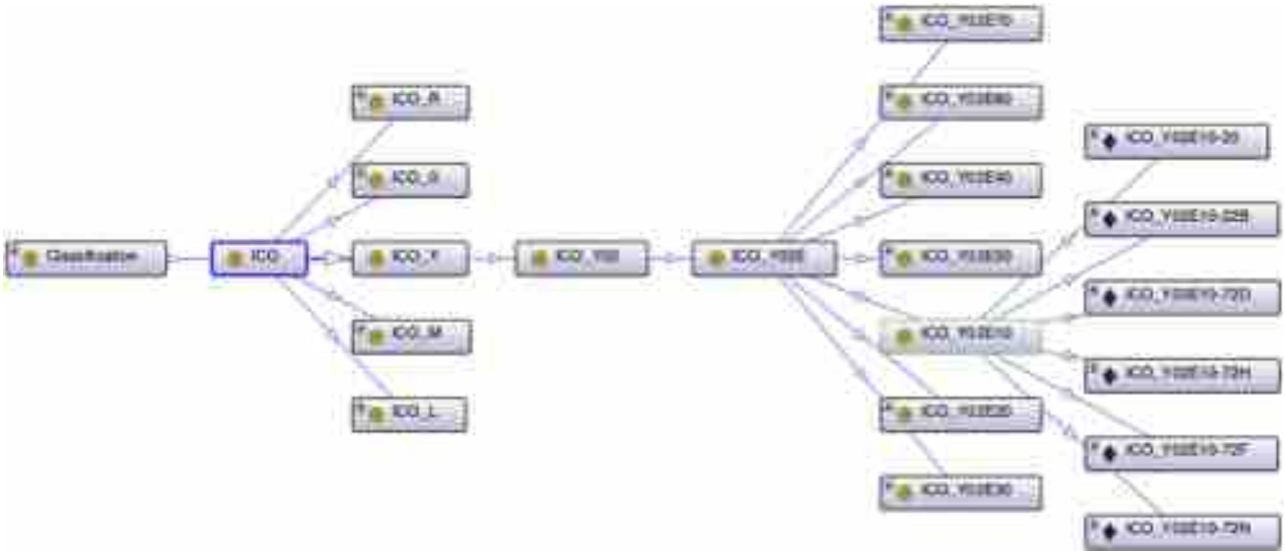

Fig. 3. Hierarchy of classes for the ICO code Y02E10/20 in Protégé.

Listing 1. Excerpt of XSD file implementing step 3.3.

```
<xsl:variable name="var8" select="concat('ICO_',substring-before(normalize-space(.),'/'))"/>

<owl:Class rdf:about="{$var8}">
<rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,4))}"/>
</owl:Class>

<owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,4))}">
<rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,3))}"/>
</owl:Class>

<owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,3))}">
<rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,1))}"/>
</owl:Class>

<owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,1))}">
<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/HCOntology.owl#ICO"/>
</owl:Class>

<owl:NamedIndividual rdf:about="{concat('ICO_', substring-before(normalize-space(.),'/'),'-',substring-after(normalize-space(.),'/'))}">
<rdf:type rdf:resource="{concat('ICO_', substring-before(normalize-space(.),'/'))}"/>
</owl:NamedIndividual>
```

technologies evolve or new technologies appear, new technological patent codes appear. The method proposed could be used not only with current technological patent codes but also with future technological patent codes, without changing the stylesheet.

4. Exploiting the ontology: reasoning through ontology hierarchy

The advantages of having the hierarchical codes represented in a hierarchy of classes in OWL lies in that the meaning of all parts of the code are explicitly available. It is therefore possible to automatically process each part of the code.

The next study cases will show the use of the meaning of parts of the code to infer new information (or new classification of patents) by using parts of the technological patent codes. In particular, this section illustrates how the proposed method enables relationships to be created between concepts using the full semantics of the technological patent codes. We draw on Bermudez-Edo et al.'s methodological proposal [17] to create the relationships between concepts.

Specifically, we will present two study cases. The first case occurs within the patent domain; we will show how to automatically reclassify patents when a new code appears (Section 4.1). The second case study will show how to automatically merge two knowledge domains: the patent domain and the financial domain (Section 4.2). In this case study, we will show how to automatically reclassify patents not only with the technological patent code of the patent domain but also with the industrial codes of the financial domain.[3]

4.1. Introducing a new technological patent code

As the technology evolves, new fields are discovered and patents must represent these new technological fields with their codes. Usually, the need for the new codes appears after a group of patents in this field have been filed and classified in other related technological fields. Our study case develops the same method used by the EPO to populate this new technological patent code, which is the same as the method used to search for patented inventions [30]. The EPO explores patents by searching for particular IPC code, or a particular keyword found in patent documents, or a combination of both (IPC codes and keywords), to reclassify them with the new code.

For example, beginning with the *Y02* term, the European Patent Office (EPO) has new ICO (In-Computer-Only) codes and sub-classifications to cover "*technologies or applications for mitigation or adaptation against climate change*", or more commonly known as "*green patent technology*". To prove the automatic reclassification of new codes, we reclassify patents as *green patent technology* using the same method as the EPO, with IPC codes and keywords, but we will use example patents that have not yet been classified as *green patent technology* by the patent authorities. With this approach, we will infer new information (ICO codes) derived from other information in the patent documents (IPC codes and/or keywords).

To this end, we use OWL [29] as the ontology language combined with the reasoner Pellet [25], and we make use of the Protégé[4] tool [24] for the implementation. This approach has been applied to PatentOntology.

The ICO codes were not implemented in PatentOntology. Therefore, before implementing the concepts, relationships and restrictions, a new class (*ICO*), and a new property (*hasICOCode*) with a domain (*PatentDocument*) and range (*ICO*), have been created, in accordance with description logics notation [31].

$$ICO \sqsubseteq Classi\ fication$$
$$Domain(hasICOCode, PatentDocument)$$
$$Range(hasICOCode, ICO)$$

Certain terms for the concepts and relationships in PatentOntology are misleading. For example, there is a property called "*hasIPCClass*" that links the patent with its IPC codes (individuals) and not the IPC Class (classes) as one might think. Therefore, for reasons of clarity, an equivalent object property called "*hasIPC-Code*" is introduced.

Next, a hierarchical classification following the Y02 classification of ICO in the EPO databases has been defined, as well as a new class (*Y02_code_pending*) that represents the *green patent technology* without an official classification from the EPO.

The three scenarios of the method used by the EPO to reclassify new codes have been implemented. The main objective behind each scenario is the summarized as follows:

1. In the first scenario, we found an IPC code that can be directly classified as a green patent technology code. For example, the IPC class *H04N7/26* addresses television systems using bandwidth reduction. We can assume that all patents classified in this field are green patent technology because they attempt to reduce energy use when transmitting a television signal [32,33].
2. A second scenario is intended to find certain environmentally friendly keywords in patents that have the same meaning in all of the technological fields and always lead to an environmental patent. For example, the keywords "*energy saving*", "*using less energy*", or "*reduce the energy*", all refer to attempts to use less energy and therefore to be environmentally friendly, regardless of the field of the invention.
3. The third scenario is to have an IPC class that is a priori non-environmentally friendly (such as *H04 "transmission in electric communications*") but when combined with certain keywords related only to this field (such as "*bit reduction*" or "*less overhead*") will lead to *green patent technology*.

The scenarios are described in detail in the next subsections.

4.1.1. Reclassification using IPC codes

In this first scenario, we will show how to infer a new code based on an existing code. To illustrate this, we will reclassify all of the patents belonging to the IPC class *H04N7/26* "*transmission of television signal using bandwidth reduction*" into the new class *Y02_H04N7_26*, which we have created as a subclass of *Y02_code_pending*. This process includes the following steps:

1. Create the classes: *Y02_H04N7_26* (a subclass of *Y02_code-pending*) and *PatentH04N7_26* (a subclass of *Patent*)
2. Populate the ontology with individuals with the IPC code *H04N7/26*
3. Define the equivalent classes for the class *PatentH04N7_26* (see Fig. 4) with Axioms 1 and 2. The first axiom will search for the patents with the IPC code *H04N7_26*. The second axiom will add the value *Y02_H04N7_26* to the patents found

$$PatentH04N7\_26 \equiv PatentDocument \sqcap \exists hasIPCCode.H04N7\_26 \tag{Axiom 1}$$

$$PatentH04N7\_26 \equiv PatentDocument \sqcap \exists hasICOCode.\{Y02\_H04N7\_26Individual\} \tag{Axiom 2}$$

With these axioms, the reasoner will classify all of the patents with the IPC class *H04N7_26* into the class *PatentH04N7_26*. Next, it will link these patents with the individual *Y02-H04N7_26* of the class *ICO*.

Therefore, when searching for *Y02* patents, because the re-classified patents are from one subclass of *Y02*, the re-classified patents will show-up.

For this first scenario, we have inserted individual patents with the *H04N7/26* IPC code, such as *US794843* (see Fig. 5), with the title "*Method of Implementing Improved Rate Control for a Multimedia Compression and Encoding System*" as well as the IPC codes *H04N7/24*, *H04N7/26* and *H04N7/30*. The patent addresses a method for rate control and bandwidth reduction. The code *H04N7/26* means '*bandwidth reduction of TV signals in communication transmission*', and therefore (as explained in Section 4.1), we can classify the patent as environmentally friendly.

The reasoner has inferred the subclass of the class *Y02_code_-pending*, *Y02-H04N7_26* for the patent *US794843* (see Fig. 5). Therefore, the reasoner has reclassified patents as *green patent technology* by making use of part of the codes, i.e. using the semantic meaning of the hierarchy found in HCOntology.

Thus, this approach enables the automatic reclassification of patents based on the IPC codes of the patents and can be easily implemented by an IT non-expert with a couple of lines. Furthermore, it is not necessary to change the application on top of the ontology. We have also shown how to create transversal relationships between different concepts.

[3] http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27.
[4] http://protege.stanford.edu.

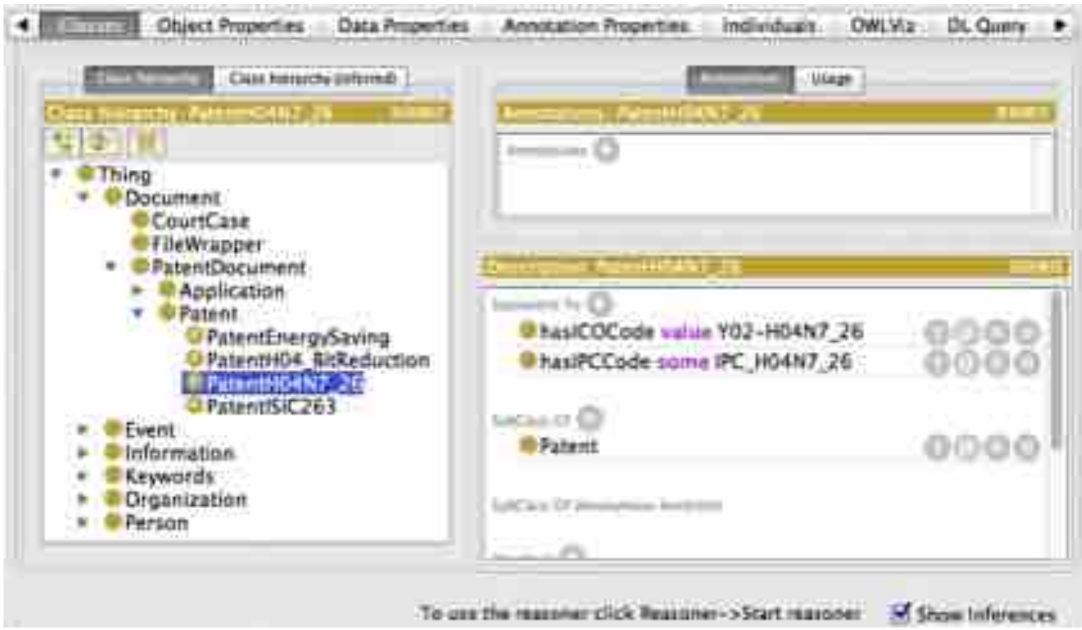

Fig. 4. Equivalent classes that classified H04N7/26 patents as *green patent technology*.

4.1.2. Reclassification using keywords

In the second scenario, the reasoner will infer a new reclassification of patents based on keywords found in the patent documents. All of the patents (regardless of the technological fields they are classified in) that have keywords related to "*energy-saving*" (e.g. energy saving, reduce energy, power reduction, etc.) will be classified as *green patent technolo*gy. This process includes the following steps:

1. Create the classes: *Y02_EnergySaving* (a subclass of *Y02_code-pending*), *Keywords* (and its subclass *EnvironmentalKeywords*) and *PatentEnergySaving* (a subclass of *Patent*).
2. Add the property *hasKeyword*, with range *Patent* and domain *Keyword*.
3. Populate the ontology with individuals with the keyword "*energy-saving*". These are patents that we found with keywords related to energy saving, and we add to them the property *ObjectHasValue (hasKeyword energy-saving)*.
4. Define the equivalent classes for the class *PatentEnergySaving* (see Fig. 6). The first equivalent class, Axiom 3, will search for the patents with the keyword "*energy-saving*". The second equivalent class, Axiom 4, will add the value *Y02_EnergySavingIndividual* to the patents found.

$$PatentEnergySaving \equiv PatentDocument \sqcap \exists hasKeyword.\{energy - saving\} \quad \text{(Axiom 3)}$$

Fig. 5. Patent US794843 reclassified.

$$PatentEnergySaving \equiv PatentDocument \sqcap \exists hasICOCode.\{Y02\_EnergySavingIndividual\} \quad \text{(Axiom 4)}$$

With these axioms, the reasoner will classify all of the patents with the value "*energy-saving*" in the class *PatentEnergySaving*. Next, it will link all of the patents that are in the class *PatentEnergySaving* with the individual *Y02-energy-saving* of the class *ICO*. Hence, when someone searches for *Y02* patents, because these patents are in a subclass of *Y02*, the re-classified patents will show-up.

For this second scenario, we have inserted individual patents dealing with energy saving, such as *WO2012010194* (see Fig. 7) with the title "*Energy Saving in a Mobile Communications Network*", that deals with energy savings in a mobile communication network and with the IPC code *H04L5/00* "*affording multiple use of the transmission path in the transmission of digital information*". In said patent, the IPC code is, a priori, non-environmental in nature, but the patent has in the text selected vocabulary related to *energy-saving*, and therefore, we have inserted into the patent the property "*hasKeyword*" with the individual item "*energy-saving*".

The reasoner has inferred the subclass of the class *Y02_code_-pending*, *Y02-energy-saving* for the patent *WO2012010194* (see Fig. 7). Therefore, the reasoner has reclassified patents as *green patent technology* making use of the transversal relationships between concepts, in this case, the keywords and patent classification.

4.1.3. Reclassification using IPC codes and keywords

This scenario will show how to reclassify patents based on two different concepts (keywords and IPC codes) that will be a combination of the previous scenarios. Specifically, the reasoner will infer a new re-classification of *green patent technology* based on the IPC code *H04* (and all the subcodes of *H04*) and keywords related to "*bit reduction*" because the energy consumed in the transmission of data in computer networks depends on the number of bits transmitted, among other things. We will link these patents to the *ICO* code *Y02_H04_BitReduction*, created as an individual of the class *Y02_tobeClassified*. In this scenario, the hierarchical classification of the IPC codes aids in reasoning regarding the *Y02* codes. In this case, by only selecting the class

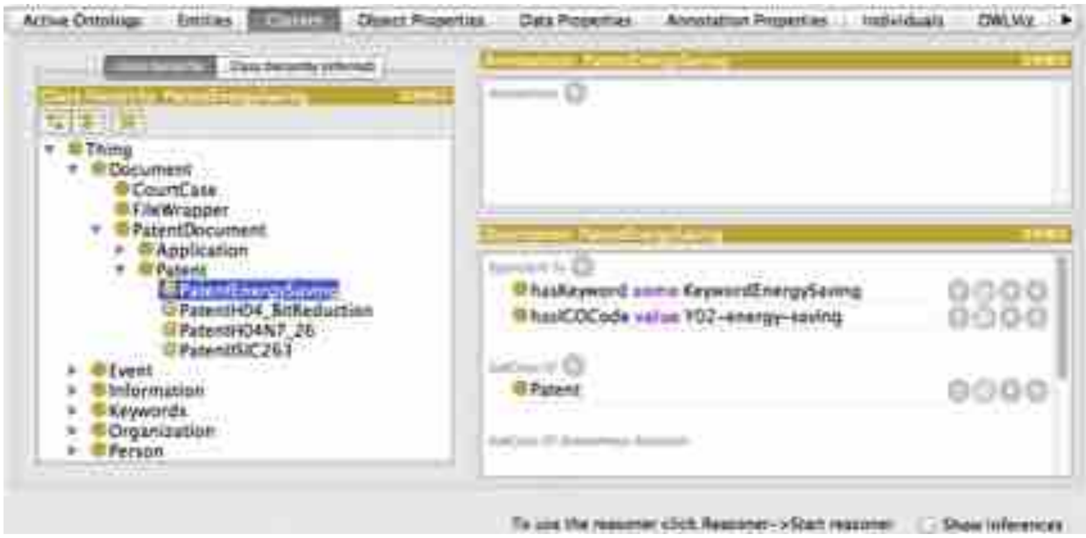

Fig. 6. Equivalent classes that classify patents related to energy saving as green patent technology.

H04, all of the patents with IPC codes that begin with *H04* will be retrieved.

This process involves the following steps:

1. Create the classes: *Y02_H04BitReduction* (a subclass of *Y02_code-pending*) and *PatentH04BitReduction* (a subclass of *Patent*).
2. Populate the ontology with the individuals with IPC codes beginning with *H04* and keyword '*bit reduction*'. Add to them the property *ObjectHasValue (hasKeyword bit-reduction).*
3. Define the equivalent classes for the class *PatentH04BitReduction* (see Fig. 8). The first equivalent class, Axiom 5, will search for the patents with the IPC codes *H04* and keyword "*bit-reduction*". The second equivalent class, Axiom 6, will add the value *Y02_H04BitReduction* to the patents found.

$$PatentY02\_H04\_BitReduction \equiv PatentDocument \sqcap \exists hasIPCCode.H04 \sqcap \exists hasKeyword.\{bit-reduction\} \qquad \text{(Axiom 5)}$$

$$PatentY02\_H04\_BitReduction \equiv PatentDocument \sqcap \exists hasICOCode.\{Y02\_H04BitReductionIndividual\} \qquad \text{(Axiom 6)}$$

The reasoner will classify all of the patents with the IPC code *H04* and the value "*bit-reduction*" in the class *Keyword* into this new class of *PatentH04_BitReduction*. Hence, when searching for *Y02* patents, because these individuals are in a subclass of *Y02*, the re-classified patents will show-up. Next, all of the patents that are in the class *PatentH04_BitReduction* will be linked with the individual *Y02_H04-bit-reduction* of the class *ICO*.

The ontology has been populated with more patents according to the designed process. A number of these additional patents meet

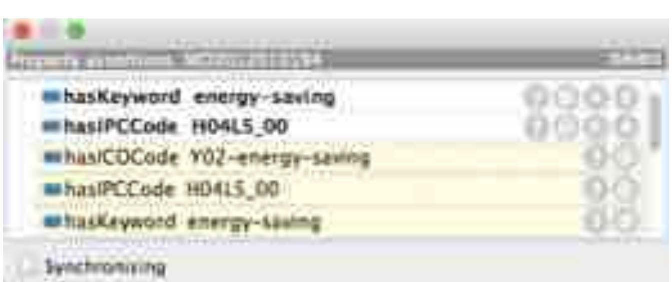

Fig. 7. Patent WO2012010194 reclassified.

the criteria of this scenario, such as the patent *US20080107132* (see Fig. 9) with the title "*Method and apparatus for transmitting overhead information*", and classified with the IPC code *H04J3/24*. The patent addresses a method that attempts to reduce the overhead in the transmission. This IPC code, *H04J3/24*, is defined as "*multiplex communications in which the allocation is indicated by an address*", and therefore an a priori non *green patent technology* concept is appreciated. However, the patent belongs to the more general code *H04* and includes certain keywords such as "*bit-reduction*". Therefore, the patent meets the criteria shown in Fig. 9.

The reasoner has inferred the subclass of the class *Y02_code_-pending*, *Y02-H04-bit-reduction* for the patent *US20080107132* (see Fig. 9). Therefore, the reasoner has reclassified patents as *green patent technology* making use of part of the codes, i.e. using the semantic meaning of the hierarchy found in HCOntology combined with transversal relationships such as keywords and patent classification.

4.2. Reasoning with HCOntology in heterogeneous domains

In this section, we will present a case study that automatically merges two knowledge domains, namely, the patent domain and the financial domain.

Multiple works have attempted to show the relationship between patent metadata and the financial situation of the firm [34,35]. The joint analysis of the patent and financial databases for specific fields of activities will lead to more robust results that avoids the perturbing effect of different contextual factors.

Financial databases include industrial codes of the companies, used to classify the activities of the firms. There are several classification schemes of industrial codes (i.e. *Standard Industrial Classification – SIC*, *International Standard Industrial Classification – ISIC*, *Nomenclature statistique des activités économiques dans la Communauté Européenne – NACE*) that allow business analysts to compare firms' economic activities on a statistical basis. However, as we have already mentioned, patent databases are classified by technological fields, and they are different from the industrial codes.

For this study case, we will focus on the ISIC[5] codes (International Standard Industrial Classification) because they are known worldwide, and they are usually included in the

[5] http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27.

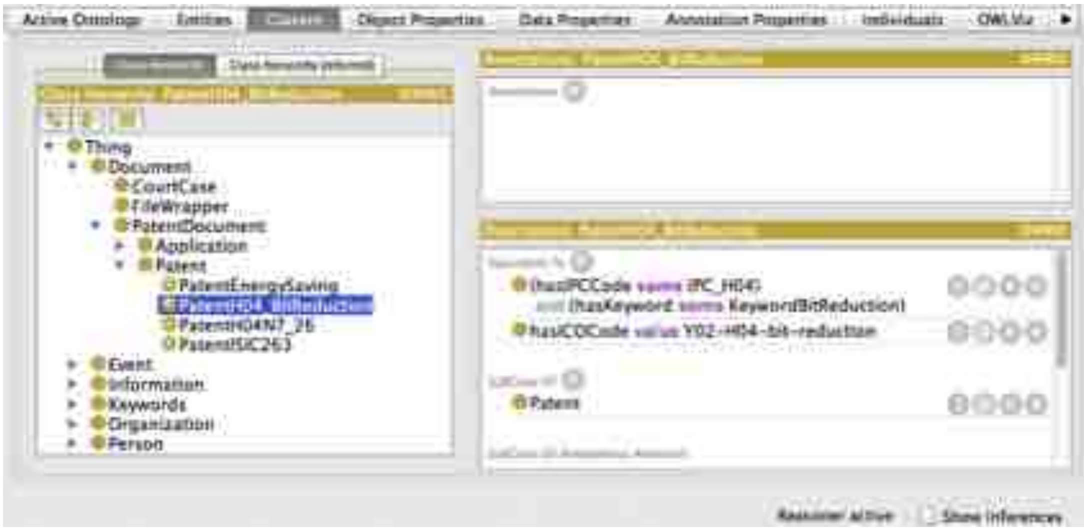

Fig. 8. Equivalent classes that classify patents H04 related to bit reduction as green patent technology.

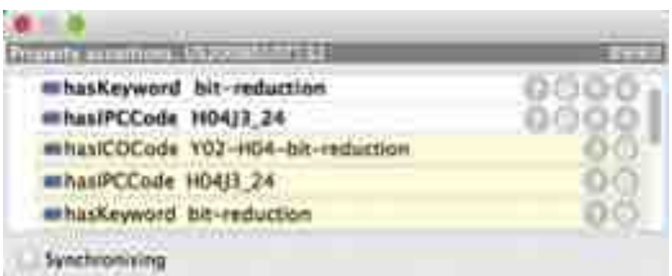

Fig. 9. Patent US20080107132 reclassified.

databases regarding firm financial information. However, our findings can be applied to any industrial classification codes.

Section 4.2.1 will show the internal structure of the ISIC codes, and Section 4.2.2 will present a case study of reasoning in heterogeneous domains using technological patent codes and industrial codes.

4.2.1. Industrial codes and mapping with technological fields

To develop the reasoning to automatically reclassify patents into industrial codes, this subsection will introduce the ISIC codes and previous efforts to match the industrial codes with the technological patent codes.

ISIC codes are organized hierarchically, which allows the use of the advantages of the hierarchy concepts; however, a specific code cannot be understood automatically by computers to provide semantic meaning to each part of the code.

ISIC codes[6] are divided into 4 parts:

- The first part is the section, which is represented by a letter from *A* to *U*
- The second part is the division and is represented by 2 digits (from *01* to *99*). The *01* to *03* digits belong to *A*, the *04* to *09* digits belong to *B*, etc.
- The third part is the group, represented by 1 digit
- The last part is the class, represented also by 1 digit

For example, the code *2630* belonging to section *C (Manufacturing)*, has the division (*26*), the group (*3*) and the class (*0*). The entire code *2630* means:

26: *manufacture of computer, electronic and optical products*
263: *manufacture of communication equipment*
2630: *manufacture of communication equipment*
2630 is the same as *263* because this group (*263*) has only one class (*2630*).

The first attempt made to match industrial codes with IPC codes was performed using the *Yale Technology Concordance* (YTC) [36]. However, the industrial sectors and the IPC codes have evolved since then. More recently, the European Commission has published a report mapping industrial codes with corresponding IPC codes [37]. This report focuses on section *C (manufacturing)* of the industrial codes and found 44 matching technological fields with industrial sectors.

The European commission report contains the most recent mapping; therefore, we will use this mapping for reasoning in the ontology with the IPC classes and to infer the industrial ISIC sectors of the patents.

4.2.2. Reasoning with HCOntology merging information domains

In this subsection different knowledge domains will be linked through various classification schemes, using the mapping between the different hierarchies of classifications. For that purpose, we will focus on the matched sector *35, telecommunication manufacturing*, of the European Commission report. This sector corresponds with code *263* of ISIC rev.4,[7] and with 20 IPC codes, among them *G06F, G06G, G06K, H03K* and *H03L*. For clarity reasons, only these 5 IPC codes are shown in this paper.

The HCOntology, developed in Section 3 of this paper, will allow the implementation of this reasoning and will demonstrate the advantage of the semantic meaning of parts of the IPC codes.

The implementation of this matching in the ontology includes the following steps:

1. Create the classes: *ISIC* (a subclass of *Classification*), a subclass *ISIC263* and *PatentISIC263* (a subclass of *Patent*)
2. Create the hierarchy of classes with the ISIC codes in a similar way as we have done with the IPC codes. The superclass is ISIC. Next, create the subclasses sections (*ISIC_A* to *ISIC_U*), then the

[6] http://unstats.un.org/unsd/cr/registry/regcs.asp?Cl=27&Lg=1&Co=0111.

[7] The European commission report matches with the old ISIC rev3.1 code 332, which correspond to the new ISIC v.4 code 263 http://unstats.un.org/unsd/cr/registry/regso.asp?Ci=60.

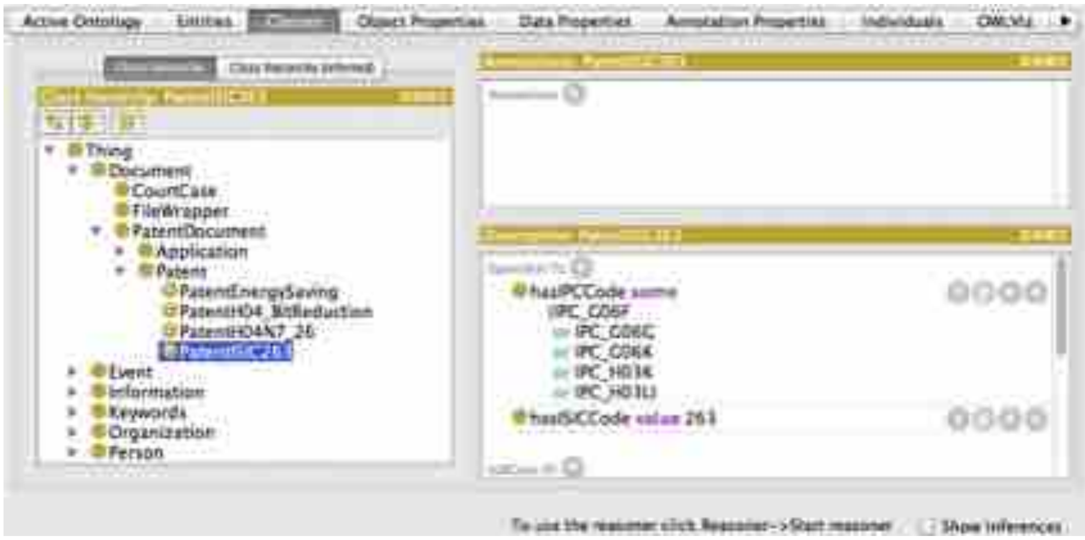

Fig. 10. Definition of equivalent classes matching IPC codes with the ISIC code.

Fig. 11. Reclassification of a patent with industrial code.

subclasses (*ISIC_01* to *ISIC_99*) and populate them with individuals (e.g. *263*)

3. Add property *hasISICCode* with domain *PatentDocument* and Range *ISIC*

$$Domain(hasISICCode, PatentDocument)$$
$$Range(hasISICCode, ISIC)$$

4. Populate the ontology with individuals with the mentioned 5 IPC codes
5. Define the equivalent classes for the class *PatentISIC263* (see Fig. 10). The first equivalent class will search for the patents containing at least one of the corresponding 5 IPC codes (see Axiom 7). The second equivalent class will add the value *ISIC_263* to the patents found (see Axiom 8)

$$PatentISIC263 \equiv PatentDocument \sqcap \exists hasIPCCode.$$
$$(G06F \sqcup G06G \sqcup G06K) \sqcup H03K \sqcup H03L) \qquad \text{(Axiom 7)}$$

$$PatentISIC263 \equiv PatentDocument \sqcap \exists hasISICCode.\{263\} \qquad \text{(Axiom 8)}$$

The reasoner, with the first equivalent class, will classify all of the patents with the IPC codes beginning with *G06F* or *G06G* or *G06K* or *H03K* or *H03L* or *H03K* into the class *ISIC263*. Next, with the second equivalent class, all of the patents that are in the class *PatentISIC263* will be linked with the individual *263*of the class *ISIC*.

For example, we have populated the ontology with the patent *US2012059512* that has the title "*Apparatus for dispensing solid pharmaceutical articles*" and with the IPC codes *G06F17/00* and *B65D83/00*. The first IPC code meets the criteria of axiom 7 because it begins with *G06F*, and the second one does not meet the criteria. As shown in Fig. 11, the reasoner has introduced the ISIC code *263* into the patent. The equivalent classes (defined above) and HCOntology allow inferences such as the one in this example.

Therefore, the semantic meaning of the hierarchy found in the different classification codes can be used to link two different knowledge domains resulting in better patent analyses. In this particular case, the financial studies could combine the information from financial databases with that of the patent databases. Therefore, it is possible to automatically use patents as a measure of innovation and relate them to data from the firms gathered in the financial databases.

5. Discussion

Information about patents is stored in large patent document databases with different data structures that make it difficult to interoperate or automatically and efficiently process the information contained therein. Additionally, in these databases there are no semantic annotations that allow further relationships and links between patents and other knowledge domains. In this context, ontologies have proved to be useful for sharing information by providing formal and sharable representation about patent domain. One of the most used information of patent metadata is technological field that delimits the categories a patent may pertain to or be associated with [4]. Patent databases use codes for technological fields that assume a hierarchical classification. However, previous patent ontologies (*Patexpert* and *PatentOntology*) do not automatically retrieve the semantics of the hierarchy of technological patent codes and do not merge information from different patent offices.

This work introduces a method to transform the representation of the hierarchical technological patent codes into a hierarchy of concepts, HCOntology filling the gaps in the previous patent metadata ontologies. The creation of HCOntology draws on the Noy and Guinness's methodology [15], and automates steps 4 and 7 of said methodology.

The proposed method enables the current and future emerging codes to be translated automatically from the patent databases to our ontology, without the need for any further implementation. The resulting hierarchy of concepts in OWL (which we called HCOntology) allows the exploitation of the information gathered in each part of the hierarchy, enriching the information retrieval process with new relationships, properties and inferred information. Thus, it is possible to automatically process each part of a code. This exploitation requires the creation of transversal relationships between concepts that reclassify existing and future patent documents. Our analysis provides opportunities to automatically update the previous classification of the technological fields in an efficient way. The approach has been applied to PatentOntology, but it could be used in any other patent ontology.

This proposal enables the automatic inference of implicit knowledge based on reclassification techniques and relationships between different application domains without changing the applications that make use of patent information. Additionally, HCOntology presents a technical solution that automates the process of populating the ontology, parsing the query results from different databases in XML format.

As any research development, our proposal presents a number of limitations. Although ontologies have strengths they also have some weaknesses. We have spot three limitations in our approach to create, reason and manage change with patent information. The first limitation is that the axioms, or relationships between concepts cannot be automatically created. An expert on the area needs to think and manually create the set of explicit assumptions regarding the mapping between the different hierarchies of classifications. In our examples in Section 4 we have manually written the axioms that reclassify patents as "*green patent technology*", and the axioms that map patents into industrial classifications.

The second limitation is that the current reasoning engines need processing time to perform the subsumption relationships. However, the proposed methodology is meant to be applied in enterprise information systems, which utilized the reasoning engines in patent searches without critical real time constraints. In this context, OWL-DL ensures decidability, and some previous works have stated the feasibility of similar applications in the patent domain (e.g. [17]).

The third limitation is the lack of generalization of the proposed method. Our method is only applicable to codes that follow a strict hierarchical structure. In any case the field of structured hierarchical codes is wide, including codes such as postal codes, bar codes, or telephone numbers, among others.

6. Conclusions and future research

Patent documents provide information particularly useful to analyze dimensions of innovations. One of the most used information of patent metadata is the technological field. This metadata is widely used in searches in databases to determine the field in which a firm may infringe upon another company's industrial rights or where there exists a gap in which a company could innovate. Our proposal in this paper provides benefits from two different perspectives. First the enriched knowledge can be obtained in a more automatic way for a better analysis of the patent information. Second the new classifications (or changes in exiting codes) can be developed without the need of changes in the rest of applications that make use of the ontology. This opportunity is especially appealing in the context of steadily increasing technological changes, in which multiple technologies (e.g. *green patent technology*) are developing quickly.

Furthermore, the Organization for Economic Co-operation and Development [2] has already highlighted the need for matching industrial codes with the technological fields of patents. Thus, the proposed method merges two knowledge domains and allows automatically reclassify patents not only with the technological patent code (patent domain) but also with the industrial code (financial domain). This link between industrial codes and technological patent codes opens excellent opportunities for multiple useful analyses, in particular for the business and economics experts.

Future research could extend our proposal in multiple directions. One direction is the combination of this proposal with artificial intelligence algorithms, such as text mining. This combination would be particularly useful to explicitly show the relationships of any data gathered in patents, not only metadata. Additionally, future research moves forward to apply our proposed methods to other domains that use hierarchical codes. The hierarchical codes would therefore be enriched with semantics, enabling the definition of new relationships, properties and inferred information. For example, large engineering companies use document part numbering conventions. The bar codes also follow a strict hierarchy in their structures. All these codes are good illustrations of the potential to use our methodology not only in the patent field and automatically translate the codes into OWL following our proposal.

References

[1] M. Giereth, A. St, M. Rotard, T. Ertl, Application of semantic technologies for representing patent metadata, 1st International Workshop on Applications of Semantic Technologies, AST06 94 (2006) 297–305.

[2] OECD, OECD Patent Statistics Manual, Statistics (2009), Available at http://www.oecd.org/science/innovationinsciencetechnologyandindustry/oecdpatentstatisticsmanual.htm.

[3] M. Bermudez-Edo, N.E. Hurtado-Torres, N. Ortiz-de-Mandojana, The influence of international scope on the relationship between patented environmental innovations and firm performance, Business & Society (2015), http://dx.doi.org/10.1177/0007650315576133.

[4] P. Foglia, Patentability search strategies and the reformed IPC: a patent office perspective, World Patent Information 29 (1) (2007) 33–53.

[5] S. Taduri, G.T. Lau, K.H. Law, H. Yu, J.P. Kesan, Developing an ontology for the U.S. Patent System, in: Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, 2011, 157–166.

[6] N. Anand, M. Yang, J.H.R. Van Duin, L. Tavasszy, GenCLOn: an ontology for city logistics, Expert Systems with Applications 39 (15) (2012) 11944–11960.

[7] T.-Y. Chen, Value ontology-based multi-aspect intellectual asset valuation method for decision-making support in K-commerce, Expert Systems with Applications 38 (5) (2011) 5471–5485.

[8] L. Ramos, R. Gil, D. Anastasiou, M.J. Martin-Bautista, Towards a Machine of a Process (MOP) ontology to facilitate e-commerce of industrial machinery, Computers in Industry 65 (1) (2014) 108–115.

[9] H. Wi, S. Oh, M. Jung, Virtual organization for open innovation: semantic web based inter-organizational team formation, Expert Systems with Applications 38 (7) (2011) 8466–8476.

[10] N. Chungoora, R.I. Young, G. Gunendran, C. Palmer, Z. Usman, N.A. Anjum, K. Case, A model-driven ontology approach for manufacturing system interoperability and knowledge sharing, Computers in Industry 64 (4) (2013) 392–401.

[11] D. Nadoveza, D. Kiritsis, Ontology-based approach for context modeling in enterprise applications, Computers in Industry 65 (9) (2014) 1218–1231, http://dx.doi.org/10.1016/j.compind.2014.07.007.

[12] M. Noguera, M.V. Hurtado, M.L. Rodríguez, L. Chung, J.L. Garrido, Ontology-driven analysis of UML-based collaborative processes using OWL-DL and CPN, Science of Computer Programming 75 (8) (2010) 726–760.

[13] A.B. Pelegrina, K. Benghazi, M.V. Hurtado, M. Noguera, J. Barjis, W. Aalst, J. Mylopoulos, et al., A framework for the semantic representation of business processes within business organizational models, Enterprise and Organizational Modeling and Simulation (63) (2010) 79–94.

[14] H. Seol, S. Lee, C. Kim, Identifying new business areas using patent information: a DEA and text mining approach, Expert Systems with Applications 38 (4) (2011) 2933–2941.

[15] N.F. Noy, D.L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, 2001 Available at http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf.

[16] W3C, XSL Transformations (XSLT), 1999 Available at http://www.w3.org/TR/xslt.

[17] M. Bermudez-Edo, M. Noguera, N. Hurtado-Torres, M.V. Hurtado, J.L. Garrido, Analyzing a firm's international portfolio of technological knowledge: a declarative ontology-based OWL approach for patent documents, Advanced Engineering Informatics 27 (3) (2013) 358–365.

[18] L. Wanner, S. Brügmann, B. Diallo, M. Giereth, Y. Kompatsiaris, E. Pianta, G. Rao, et al., PATExpert: Semantic Processing of Patent Documentation, 2009 Available at http://recerca.upf.edu/patexpert/.

[19] S. Taduri, G.T. Lau, K.H. Law, H. Yu, J.P. Kesan, An ontology to integrate multiple information domains in the patent system, in: IEEE International Symposium on Technology and Society ISTAS, 2011, 23–25.

[20] LexisNexis Website, 2012, Available at www.lexisnexis.com/.

[21] S. Choi, D. Kang, J. Lim, K. Kim, A fact-oriented ontological approach to SAO-based function modeling of patents for implementing function-based technology database, Expert Systems with Applications 39 (10) (2012) 9129–9140.
[22] V.-W. Soo, S.-Y. Lin, S.-Y. Yang, S.-N. Lin, S.-L. Cheng, A cooperative multi-agent platform for invention based on patent document analysis and ontology, Expert Systems with Applications 31 (4) (2006) 766–775.
[23] The Protégé Ontology Editor and Knowledge Acquisition System, 2012, Available at http://protege.stanford.edu/.
[24] N.F. Noy, R.W. Fergerson, M.A. Musen, The knowledge model of Protégé-2000: combining interoperability and flexibility, Knowledge Engineering and Knowledge Management Methods Models and Tools 1937 (1) (2000) 1–20.
[25] E. Sirin, B. Parsia, B. Grau, A. Kalyanpur, Y. Katz, Pellet: a practical OWL-DL reasoner, Web Semantics Science Services and Agents on the World Wide Web 5 (2) (2007) 51–53.
[26] A. Malhotra, J. Melton, N. Walsh, XQuery 1.0 and XPath 2.0: Functions and Operators, Working draft W3C, 2007 Available at http://www.w3.org/TR/xpath-functions.
[27] H. Bohring, S. Auer, Mapping XML to OWL ontologies, Leipziger InformatikTage 72 (2005) 147–156.
[28] T. Rodrigues, P. Rosa, J. Cardoso, Moving from syntactic to semantic organizations using JXML2OWL, Computers in Industry 59 (8) (2008) 808–819.
[29] M.K. Smith, C. Welty, D.L. McGuinness, OWL Web Ontology Language Guide. W3C Recommendation, The World Wide Web Consortium (W3C), 2004 Available at http://www.w3.org/TR/owl-guide.
[30] M.W.E. Goebel, F.S. Hintermaier, Searcher's little helper – ICO index terms, World Patent Information 33 (3) (2011) 260–268.
[31] F. Baader, The Description Logic Handbook: Theory, Implementation and Applications, Cambridge University Press, 2003.
[32] J. Lorchat, T. Noel, Energy saving in IEEE 802.11 communications using frame aggregation, GLOBECOM'03, in: IEEE Global Telecommunications Conference, 2003, 1296–1300.
[33] W. Peng, X.-C. Lu, On the reduction of broadcast redundancy in mobile ad hoc networks, in: Proceedings of the 1st ACM International Symposium on Mobile Ad Hoc Networking & Computing, 2000, pp. 129–130.
[34] V. Oltra, R. Kemp, F. de Vries, Patents as a measure of eco-innovation, International Journal of Environmental Technology and Management 13 (2) (2010) 130–148.
[35] D. Vries, Leveraging Patents Financially. A Company Perspective, Springer, 2012.
[36] S. Kortum, J. Putnam, Assigning patents to industries: tests of the Yale technology concordance, Economic Systems Research 9 (2) (1997) 161–176.
[37] U. Schmoch, F. Laville, P. Patel, R. Frietsch, Linking Technology Areas to Industrial Sectors Final Report to the European Commission, 2003 Available at ftp://ftp.cordis.europa.eu/pub/indicators/docs/ind_report_isi_ost_spru.pdf.

Maria Bermudez-Edo is a Research Fellow in the Institute for Communication Systems (ICS) at the University of Surrey. She received her PhD in Information and Communication Technologies from the University of Granada, and her Engineering degree in Telecommunications from the Polytechnique University of Valencia. Prior to join the academia she worked in the industry (Siemens, Telefonica) and international organizations (e.g. European Patent Office) in the field of network communications. She is currently involved in several European projects in the Internet of the Things field. Her research interests include semantic technologies, Internet of Things and machine learning. Her research has appeared in the proceedings of various peer-reviewed international conferences and journals

María V. Hurtado is an associate professor of computer science at the University of Granada (Spain), where she received a PhD in computing. At present, she is in charge of the coordination of the programme master degree "*Master of Business Processes and Technology*". Her research interests include knowledge and information systems modelling, ontology-driven approaches to systems and software engineering, assistive technology and ICTs applied to education. She has published around hundred research papers in international conferences and journal specialized on these topics. She has served on programme committees and referee of different international conferences, and she is also part of the editorial board of Information & Management journal. She is involved in various (European, national and regional) R&D&I projects.

Manuel Noguera is a permanent Assistant Professor of Computer Science at the University of Granada, Spain, where he obtained his Ph.D. in 2009. Most of his work focuses on ontology-driven approaches to Systems and Software Engineering. The results of his Ph.D. work have been mainly applied in the groupware, semantic Web and business process modelling domains. Other research interests include systems and software architectures, mobile computing, m-health, sports engineering, remote activity monitoring, collaborative systems, business process management and service-oriented architectures. He is also co-founder of Everyware Technologies Ltd, a spin-off company specialized in the development of mobile apps.

Prof. Dr. Nuria E. Hurtado-Torres PhD in Business Administration and Economics, University of Granada. She is a full professor of Management at University of Granada (Spain) and visiting professor at Queen Mary University (UK). Her research interests include the connections between internationalization, environmental strategy, innovation and patents. She has published in top tier journals such as Journal of Business Ethics, International Business Review, Journal of Business Research, Decision Support Systems, Social Indicators Research, Journal of Environmental Management, among others. She has lead different research projects.