Applied Ontology 7 (2012) 387–400
DOI 10.3233/AO-2012-0116
IOS Press

Terms as labels for concepts, terms as lexical units: A comparative analysis in ontologies and specialized dictionaries

Marie-Claude L'Homme * and Gabriel Bernier-Colborne
Observatoire de linguistique Sens-Texte, Université de Montréal, Montréal, QC, Canada

Abstract. Terms (or terminological units) are objects that are manipulated by different fields and that are at the core of several applications, e.g. representing knowledge, describing specialized usages, classifying documents. This inevitably leads to discrepancies (and often large ones) between resources in which terms are described and encoded (e.g., ontologies, terminologies, dictionaries, thesauri). This paper attempts to characterize these differences by analyzing two different applications: (1) the elaboration of a domain ontology; and (2) the compilation of a specialized dictionary whose aim is to account for the linguistic functioning of terms. We deliberately chose two very different resources that rely on opposing methodologies hoping that this would lead to a better understanding of some of the facets terms may have and what can be expected of resources that describe them. Examples are taken from the fields of computing and the Internet. We focus more specifically on input devices. The analysis reveals that resources manipulate terms in very different ways and questions the possibility of exchanging data from one resource to the other without jeopardizing some of the principles on which they are based.

Keywords: Term, concept, lexical unit, ontology, terminological dictionary

1. Introduction

As far as the general lexicon is concerned, the issue of the compatibility or incompatibility between lexical knowledge and ontological knowledge has raised some interest in the past years (Buitelaar et al., 2009; Hirst, 2009, among others). Researchers seem to agree that: (1) there is a true benefit in taking into account linguistic or lexical information in ontologies and connecting it somehow to ontological information; (2) linguistic data should be encoded separately from the ontology (proposals as to how this data should be represented and how it should interact with the ontology have also been made in that direction (Buitelaar et al., 2009; Cimiano et al., 2011; Montiel Ponsoda et al., 2010)).

However, to our knowledge, little has been done in order to examine the same question exclusively from the point of view of the specialized lexicon. It is often assumed that the specialized lexicon is much more organized (in the sense that it is better structured and standardized) than the general lexicon (Hirst, 2009, among others). However, as will be argued in this article, this assumption might be misleading if applied indistinctively to all reference works in which terms are encoded and described.

In this paper we examine the issue by comparing two small domain ontologies of computer input devices that we designed for the purpose of this comparison and parts of the contents of a specialized

*Corresponding author: Marie-Claude L'Homme, Observatoire de linguistique Sens-Texte (OLST), Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, QC, Canada H3C 3J7. Tel.: +1 514 343 6111 (# 32480); E-mail: mc.lhomme@umontreal.ca.

dictionary (entries related to input devices – e.g., *keyboard*, *mouse*, *trackball* – some of their properties – e.g., *wireless* – and some activities associated with them – e.g., *click*, *drag and drop*). The two types of resources were built separately.

We assume that specialized dictionaries necessarily contain terms and that ontologies need to manipulate different sorts of labels (some of them being terms) when capturing and organizing knowledge. We will analyze the similarities and differences between the objects that are called *terms* in both types of repositories hoping that this will lead to a better understanding of what terms are, what can be expected of resources that encode and describe them and, ultimately, whether one resource can serve to enrich the other. We deliberately chose two very different resources that rely on opposing methodologies hoping that this would lead to a better understanding of some of the facets terms may have.

The article is organized as follows. Section 2 reviews some of the theoretical issues debated in terminology circles. Section 3 briefly describes the principles according to which two ontologies of input devices were created and gives a few general considerations on the place of terms in the formal representations. Section 4 presents a specialized dictionary that was used in this comparative analysis and also comments on how terms appear in this repository. Section 5 analyzes more specifically how terms are taken into consideration in each resource. In Section 6, some general conclusions are drawn.

2. Different terms for different people

When asked: "What is a term?" even the most knowledgeable terminologist will probably hesitate before giving a straightforward answer. This might appear surprising but it hardly is, since terms (or terminological units) can be defined differently according to the application from which they are considered. Different applications rely on diverging theoretical backgrounds and are inevitably based on opposing methodologies. It also appears that although all refer to the objects they manipulate as *terms*, these objects can differ substantially.

This variety of viewpoints was shown in a study carried out by Estopà (2001) who asked four different categories of experts (namely, terminologists, translators, medical doctors and indexers) to identify in a text on medicine those units that appeared to be terms according to them. The study demonstrated that lists compiled by each expert category differed in terms of both content and length. Medical doctors and terminologists would select more units than the other two categories. The proportion of single-word versus multi-word terms would vary from one expert to the other as well as that of parts of speech. And these are only some of the differences that were reported.

Another study whose objective was to compare resources in which terms can be found, namely ontologies, specialized dictionaries and lexical databases (L'Homme, 2008), showed that resources differed considerably in terms of coverage, types of terms recorded and meanings taken into consideration as well as relations described. This has also been shown from a different angle by Overton et al. (2011) who used existing ontologies to annotate clinical reports and found that they covered only partly the terminology found in the reports.

In addition to these empirical studies, scholars in terminology (Cabré, 2003; Sager, 1990, among others) provided theoretical explanations for these differences, stating that terms or terminology can be viewed from a linguistic, cognitive or communicative dimension (Sager, 1990). Cabré (2003) suggested a metaphor, that of the polyhedron, to illustrate the fact that terms have a variety of facets (linguistic, cognitive, communicative, etc.) each offering a different view on the object considered, i.e. terms.

Although the perspectives on terms differ, it is often assumed that the set of terms in specialized fields is somehow better structured than the general lexicon. As stated by Hirst (2009), p. 283:

> In highly technical domains, it is usual for the correspondence between the vocabulary and the ontology of the domain to be closer than in the case of everyday words and concepts. This is because it is in the nature of technical or scientific work to try to identify and organize the concepts of the domain clearly and precisely and to name them unambiguously (and preferably with minimal synonymy).

This paper aims to contribute to this debate from a slightly different point of view. We will limit ourselves to the way data is encoded in ontologies and in a specialized dictionary and focus on the role of labels (associated here with the notion of "term" in both repositories). Our view on the notion of "ontology" is based on the clarifications made by Van Rees (2003), for whom it differs from other related notions: "classification", "taxonomy" and "thesaurus". The domain ontologies to which we refer in the next sections are explicit specifications of concepts and relations between these concepts. A terminology,[1] on the other hand, is a list (that is organized alphabetically, thematically or with a set of conceptual relations) of terms that belong to the same field of knowledge (or to related fields of knowledge).

3. Two ontologies of input devices

In order to illustrate how terms may be used in knowledge representations, two ontologies of computer input devices were constructed. As is usually the case in the construction of ontologies, these ontologies were created with a specific application in mind. Both ontologies are small in size, containing about 20 concepts each.

The first ontology was designed to enable users to enter computer peripherals in a catalogue. Such an ontology could be used by a business that sells computer devices, for instance. This application determined how fine-grained the class hierarchy needed to be, as an "ontology should not contain all the possible information about the domain: you do not need to specialize (or generalize) more than you need for your application (at most one extra level each way)" (Noy & McGuiness, 2011, p. 19). It was thus decided that the instances of this ontology would be specific models of computer input devices. A different application might well change the scope of the ontology: if it was to be used for inventory, the models might have been the lower-level classes, and individual units, their instances.

The second ontology was less strictly bound to this application – it was rather oriented toward a proper definition and description of domain concepts.

The first application-driven ontology was built using the most current version of the Protégé-Frames editor available at the time (version 3.4.6), as this editor is well suited to the construction of loosely-defined, "frame-based" ontologies. The second, definition-driven ontology is based on the Web Ontology Language (OWL) and was built using the Protégé-OWL editor (version 4.0.2). Our aim was to take advantage of the formal rigour of description logics and the OWL language, as well as the inference engines that can be used with the Protégé-OWL editor.

3.1. Frame-based ontology

An initial class hierarchy for the first ontology was established using information extracted from Wikipedia (2011), which included several different classifications of input devices. This first class hierarchy, shown in Fig. 1, was then modified as classes were added to the hierarchy and their properties

[1] *Terminology* is in fact polysemous. It can refer to (1) a collection of terms related to a field of knowledge along with information about them (in the article, we use *specialized dictionary* to designate this first meaning), (2) a theory about concepts and terms or (3) the activity of collecting, describing and organizing terms in a resource, such as a term bank or a specialized dictionary (*terminography* and *terminology work* are sometimes used in the literature to refer to this third meaning).

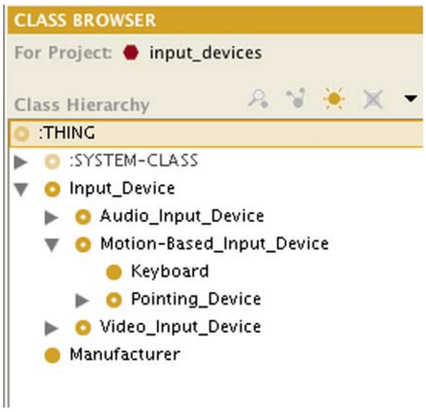


Fig. 1. Partial view of the class hierarchy in the frame-based ontology. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

(known as "slots" in Protégé-Frames) were defined – properties such as hasOpticalSensor or hasScrollWheel. For instance, if a subset of sister nodes in the hierarchy had a certain number of properties in common, this suggested the existence of an intermediate parent node which could be added to refine the hierarchy and make the classification more useful.

The classification of the various input devices made up a large part of the class hierarchy. A "manufacturer" class was added[2] in order to link each instance of an input device to its manufacturer, which could be helpful in accessing a device model by looking up its manufacturer.

The various parts that make up the input devices are not included in the class hierarchy, as the application does not require it, but they were added as slot values. For instance, there was no need to create a class for concepts such as "key" or "button", however the number of buttons was added as a slot for the class "mouse", as this information was relevant to the application. Similarly, a slot was added to indicate whether a specific model of mouse had a scroll wheel, the value of which was Boolean.

As for the function of the input devices, or other information concerning their very nature, they are not formally described in the first ontology, and the application does not require it. The essential difference between a scanner and a webcam does not need to be represented – the fact that they are sister nodes in our hierarchy suggests that some important property distinguishes them, but this is not represented formally.

3.2. *OWL ontology*

In the second, definition-driven ontology, classes are defined using property restrictions: necessary and sufficient conditions determine membership to a class. There is a formal description of the typical use of each input device and of its input mechanism; the ontology also states whether a device is an on-board component or an add-on. As seen in Fig. 2, concepts such as "graphics tablet" belong to the superclass "motion-based input device" in the *asserted* class hierarchy, yet a class may have multiple superclasses in the *inferred* class hierarchy. For example, each class has a "hasOnboardness" property, which is restricted to either the "add-on input device" class or the "onboard input device" class. These two disjoint classes are equivalent to the superclass of all classes having a specific value (onboard or

[2]It should be said that deciding whether a property should be represented as a slot or by means of a relation between classes was at times quite challenging.

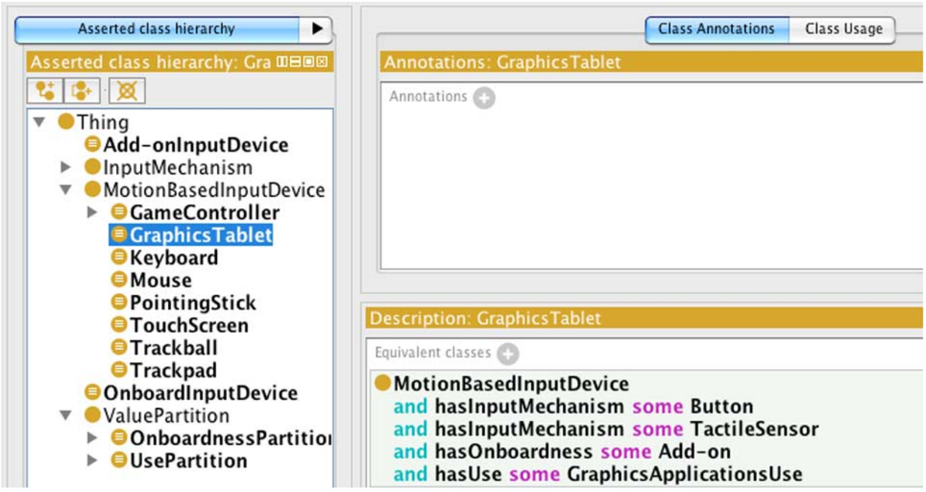


Fig. 2. Class hierarchy and definition of the class "graphics tablet". (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

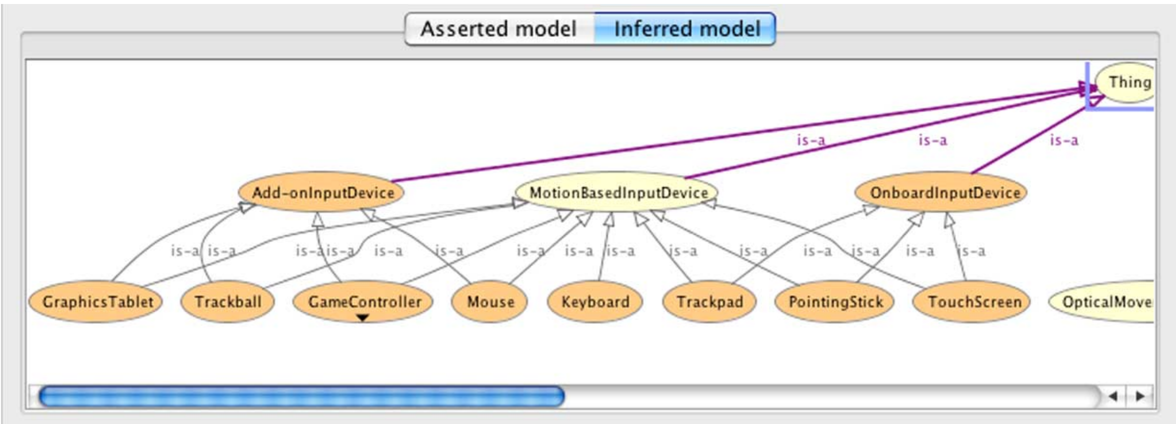


Fig. 3. Graph representing the asserted and inferred superclasses of input devices. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

add-on) for the property "hasOnboardness". The inference engine can then populate them automatically. We thus obtain a system of classes that can have multiple superclasses, as shown in Fig. 3 (a graphic representation of the class hierarchy made using the OWLViz plug-in).

As mentioned previously, OWL enables ontologists to define classes by means of property restrictions, and one must ensure that these class definitions are consistent. In the input device ontology, before the "hasUse" property was added to link each concept to its typical use, the inference engine classified "mouse" and "keyboard" as subclasses of "game controller", which highlighted a discrepancy in the class definitions. This illustrates how the inference engine can be used to test the consistency of an ontology.

3.3. *Terms in ontologies*

In the ontologies described in the two subsections above, terms can be associated with the labels used for class names (e.g., *mouse*, *keyboard*, *trackball*).[3] The labels chosen for the classes in these ontologies are not necessarily representative of actual language use: certain class labels might not actually appear in dictionaries or texts related to the domain, yet the classes themselves play an important role in creating a coherent class hierarchy. At some levels of the ontology, class names correspond to terms that can be found in terminological dictionaries or running text (e.g., *input device*, *mouse*, *trackball*, etc.); in others, class names are created for purposes of classification (e.g., *add-on input device*). This seems to be common practice among ontologists, as evidenced by many of the ontologies available on Bioportal, which contain class names such as "Radlex entity" or "NCBO Bioportal concept". Thus, if we view terms as class names, we face a degree of ambiguity, as certain class names correspond to terms used in natural language, whereas others are created solely for the purpose of creating consistent classifications and coherent class definitions. Adding to this ambiguity is the use of coordinated terms (e.g., "cellular component organization or biogenesis" in the Gene Ontology) as class names.

For applications that require it (and many do), terms can also be defined as the list of possible linguistic forms that can be used to express a given concept (e.g., *mouse*, *computer mouse*), as class labels can also have synonyms (exact synonyms, "narrow" synonyms, etc.) or variations. Other linguistic information may be linked to a class as well, typically a textual definition of the concept. Such linguistic information is generally placed in the "documentation" field of the classes in the ontology (in Protégé-Frames) or in annotations such as "rdfs:label" or "rdfs:comment" (in Protégé-OWL) or in a separate resource (a parallel lexicon, as proposed in Buitelaar et al., 2009; Montiel Ponsoda et al., 2010) that is linked to the ontology. Furthermore, someone interested in the terminology of the domain would not necessarily find useful information in the class hierarchy (where classes might be represented by their unique identifier), and might turn their attention to other components of the ontology (documentation, annotations, etc.) or to the properties that bind concepts to other concepts or to values (numbers, strings, Boolean values).

Coming back to terms viewed as class names, in order to be used as such, they must be "standardized" in the sense that if many possible linguistic expressions exist in a given language to express a concept, one needs to be chosen to represent the class; this is referred to as the "class name" or "preferred name". In addition, names should not be repeated as each serves to identify a unique concept. Ideally, a class name (a term or a label for the purpose of classification) should be transparent in the sense that its form should be somewhat indicative of the concept for which it stands.

In ontologies, terms may be viewed as devices to access individual concepts that appear in a general knowledge structure. The focus of ontologies is to describe, characterize and organize concepts (identify their place within a knowledge structure, chiefly but not exclusively organized in terms of ISA relationships), and distinguish concepts from one another in terms of necessary and sufficient conditions. In ontologies, the objects that are represented (that may or may not be labelled by means of existing terms) may be of different natures (entities, parts of these entities, manufacturer, use, etc.) and this difference must be acknowledged and taken into account (although there are different options as to how to take them into account, i.e. create new classes, properties of a specific class, relations).

The focus is thus not on the description of the labels used to identify concepts. Hence, concepts are viewed as independent objects and can be analyzed without considering the name that was selected to

[3]Concepts are also given language-independent unique identifiers. Human-readable class names or annotations are linked formally to the class identifiers. Thus, there is an important distinction between unique identifiers (non-linguistic) and human-readable class names, which can (but do not necessarily) correspond to actual terms. "Non-linguistic" identifiers for classes have the advantage of being language-independent (provided that slight differences that may exist between languages are taken into account elsewhere in the ontology). Hence, a domain ontology can be used for applications related to different languages. This was not done in the two ontologies of input devices described in Sections 3.1 and 3.2, but is often done in large scale ontologies.The mapping of class names to unique identifiers is helpful to help interpret the classes (or give them a more intuitive flavour), but it is unnecessary from the point of view of knowledge modelling.

represent them. Terms are thus important but secondary from this perspective: in fact, from a strictly formal point of view, they are not even necessary, since other non-linguistic devices can have the same function.

Hence, when considering matters from a concrete point of view, terms have an ambiguous status in ontologies. They can be viewed as means to express concepts (in such a case, they need to be standardized) or considered independent from concepts. In both cases, linguistic expressions still need to be distinguished from the classes, and lists of terms, variants and synonyms will be linked to certain (not all) classes in the hierarchy. In addition, ontologists will make decisions as to what terms should be and not necessarily use them as they appear and behave in natural language.

4. Input devices in a specialized dictionary

The specialized dictionary used in this study is a multilingual database called the *DiCoInfo, Dictionnaire fondamental de l'informatique* (http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi) and is compiled under the responsibility of one of the co-authors. It contains terms in English and in French and a Spanish version is under development. It is based both theoretically and methodologically on the framework of Explanatory Combinatorial Lexicology, ECL (Mel'čuk et al., 1995).

Terms are selected as relevant headwords if they correspond to lexical units in the sense of Cruse (1986)[4] and if they convey a meaning related to the fields of computing and the Internet. Hence, multi-word units are selected if their meaning is non-compositional from the point of view of the field (e.g., *touch pad*); units with multiple meanings related to computing or the Internet are described in separate entries (e.g., *address*$_1$ "location in the memory", *address*$_2$ "identifier for a computer", etc.). A specific meaning is delimited by observing the set of relationships (similarities and oppositions) a lexical unit shares with other lexical units.

Finally, non-specialized meanings, even if they appear in texts on computing, are not taken into consideration (*memory* as computer storage is described but not *memory* when it refers to the human faculty).

4.1. Description of the entries

The DicoInfo, which is still under construction,[5] contains terms that refer to input devices (*keyboard*, *mouse*, *touch pad*, *trackball*). Its word list also includes semantically related terms such as hypernyms (*device*, *peripheral*) and collocates that express properties (*ergonomic*, *wireless*) or activities (*click*, *drag and drop*, *type*) associated with input devices.

The DicoInfo aims to give a description of the lexico-semantic properties of terms. As shown in Figs 4 and 5, it contains the following data categories: (1) actantial (argumental) structure along with linguistic realizations of actants (arguments); (2) synonyms and variants; (3) contexts extracted from the corpus used to compile the dictionary; (4) equivalents in French and Spanish; (5) a list of lexical relationships.[6] In this list, lexical units that are related semantically to the headword (either paradigmatically or syntagmatically) are provided along with a short explanation of the relationship. When the related unit has its own entry, a hyperlink allows users to access it directly.

[4] A lexical unit is a lexical form (that can behave as an autonomous unit in a sentence) that conveys a specific meaning.

[5] As of March 2012, the English version contained approximately 800 entries that include nearly 5000 lexical relationships.

[6] In the French version (which is more complete that the English one), definitions are also provided. In the English and French versions, some entries also include tables with annotated contexts.

mouse $_1$, n

Status: 2

Actantial structure: a mouse: ~ used by Agent{user $_1$} to act on Patient1{icon $_1$} or Patient2{pointer}

Linguistic realizations of actants

Synonym(s): computer mouse

Contexts

The movement of the mouse over a flat surface is mirrored by a pointer on the monitor screen. (Source: DEVICE02)

Extending this concept, "hypermedia" lets you click on something with your mouse and bring up not only text, but also graphics, sound, and animation. (Source: WEB)

This minitower is accompanied by a wireless mouse (which operates flawlessly at a distance of up to 15 feet), a USB color scanner, a programmable keyboard, and an antenna that receives wireless Internet feeds. (Source: PCMAG2)

Lexical relations

Spanish: ratón $_1$

French: souris $_1$

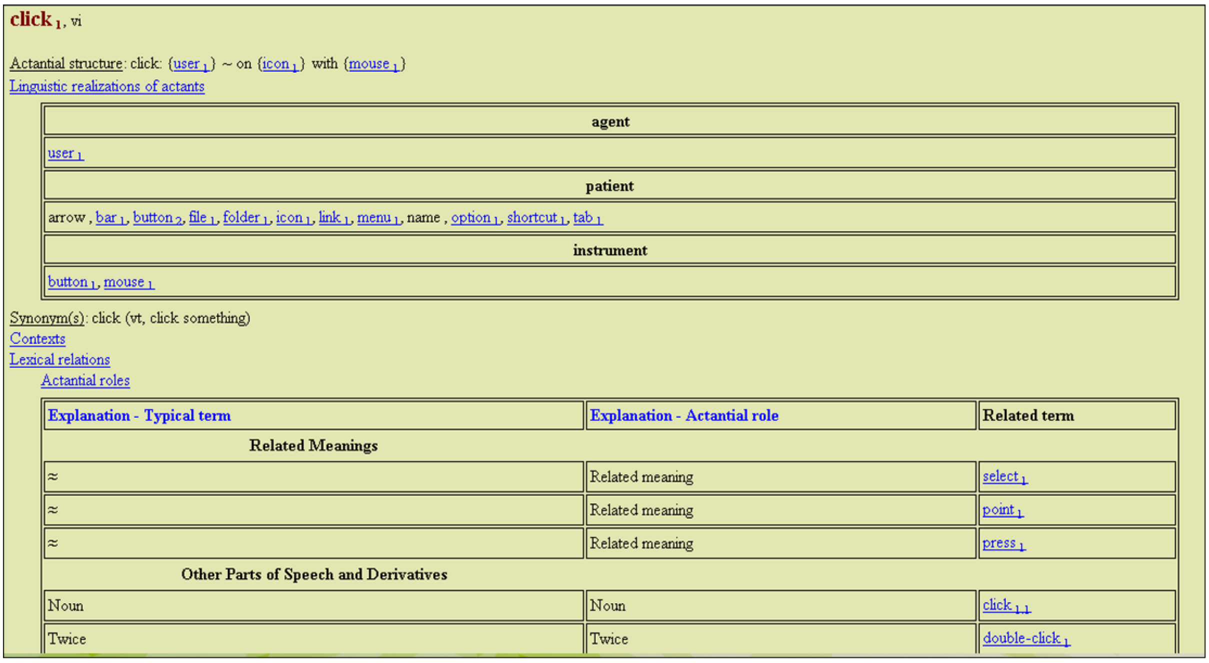
Fig. 4. Term record for *mouse*. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

Explanation - Typical term	Explanation - Actantial role	Related term
Related Meanings		
≈	Generic	device $_1$
≈	Generic	peripheral $_1$
≈	Related meaning	keyboard $_1$
≈	Related meaning	touch pad $_1$
≈	Related meaning	trackball $_1$
Types of		
The movements of which are captured with an optical system	The movements of which are captured with an optical system	optical ~
That is easy and comfortable to use	That is easy and comfortable to use	ergonomic $_1$ ~
That does not need to be linked to a computer to function	That does not need to be linked to a computer to function	wireless $_{1}$, ~
Combinations		
The user uses a m.	The agent uses a « Key word »	move a ~
The user uses a m. to act on the icon	The agent uses a « Key word » to act on the patient1	click $_1$ on … with a ~
The user uses a m. to act on the icon	The agent uses a « Key word » to act on the patient1	drag and drop $_1$ … with a ~
The user uses a m. to act on the pointer	The agent uses a « Key word » to act on the patient2	move … with a ~
Others		
Part	Part	ball
Part	Part	button
Part	Part	scroll wheel

Fig. 5. Lexical relationships listed in the entry *mouse*. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

The compilation of the list of relationships is an ongoing process in the sense that part of the relationships are discovered when first writing the entry (e.g., collocates found in the corpus examples used to build the entry) while others are captured when writing other entries (e.g., writing the entry for the verb *click* will help complete the entries for all the terms referring to devices associated with this activity).

The dictionary takes into account terms that belong to different parts of speech, namely nouns (*byte*, *computer*, *mouse*, *program*, *user*), verbs (*click*, *configure*, *download*), adjectives (*virtual*, *wireless*) and adverbs (*dynamically*). Entries for all these terms have the same structure, but their contents (especially, lexical relationships) can vary as shown in Fig. 6, which illustrates the entry for the verb *click* (compare to the previous entry for *mouse*).



click 1, vi

Actantial structure: click: {user 1} ~ on {icon 1} with {mouse 1}
Linguistic realizations of actants

agent
user 1
patient
arrow , bar 1, button 2, file 1, folder 1, icon 1, link 1, menu 1, name , option 1, shortcut 1, tab 1
instrument
button 1, mouse 1

Synonym(s): click (vt, click something)
Contexts
Lexical relations
Actantial roles

Explanation - Typical term	Explanation - Actantial role	Related term
Related Meanings		
≈	Related meaning	select 1
≈	Related meaning	point 1
≈	Related meaning	press 1
Other Parts of Speech and Derivatives		
Noun	Noun	click 1.1
Twice	Twice	double-click 1

Fig. 6. Term record for *click (v.)*. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/AO-2012-0116.)

4.2. *Terms in dictionaries*

In specialized dictionaries, terms represent the focus of the description and are in fact the main reason for which they are compiled. They form the list of headwords enabling access to information and they are the objects that are described in entries that are created for them. Terms appearing as headwords necessarily correspond to linguistic expressions that occur in specialized texts or that are used by experts.

As was mentioned above in Section 4.1, in the specific dictionary described here, an entry corresponds to a lexical unit, i.e. a lexical form considered from one of the meanings it can convey. Hence, the information provided in entries can focus on form (part of speech, syntactic behaviour) or on meaning (definition, semantic relationships), but is necessarily coherent with the meaning and form under consideration (i.e., a lexical relationship is only valid when a specific meaning is considered). A lexical form that conveys more than one meaning in the field of computing appears in separate entries.

The same meaning can be expressed with more that one form (e.g., *mouse*, *computer mouse*; *touchpad*, *touch pad*). In most terminological dictionaries, one of these forms is chosen as the headword (in some cases, this headword is preferred or standardized) and the others are listed as synonyms or variants. A cross-reference system can be created to allow users to access the entries from any of the existing forms.

Terms can belong to different parts of speech (however, they all belong to the open class category) and this must be acknowledged in the dictionary. Although all entries have the same structure, differences can be seen it their contents (e.g., a verb does not have the same kinds of relationships with other words as a noun). Terms can also have various meanings (refer to concrete objects, units of measurement, animates, activities, properties, etc.). Again, this appears in the contents of the entries since a term that refers to an activity does not participate in the same kinds of relationships as a term that refers to a concrete object, but this does not impact on the form of entries and all are represented using the same basic structure.

5. Terms in ontologies and terms in dictionaries: A comparison

Considering what was stated above, it can easily be seen that there is no straightforward relationship between terms as they appear in domain ontologies and terms that are listed in dictionaries, at least when considering the ontologies we built and the dictionary we used in this comparison. The next subsections will examine this issue more closely.

5.1. A comparative look at the labels used in ontologies and dictionaries

In ontologies, terms can appear as class names and as linguistic realizations for them in class annotations (e.g., *input device*, *mouse/computer mouse*, *trackball*, etc.). In addition, class names can be created for the purpose of classification and take forms that cannot be found in dictionaries (e.g., *add-on input device*). Lower-level classes in the hierarchy are designed to capture the objects for which the ontology is built (e.g., *mouse*, *trackball*) and these normally have names that are used in natural language (hence that appear in dictionaries), but many parent classes are defined for the sole purpose of capturing generalizations and names are often created to label them in a transparent way. On the other hand, dictionaries should contain (in the list of headwords) units that correspond to linguistic expressions that are found in specialized texts or are used by experts. In the specific dictionary to which we refer in this paper, an additional constraint is added, since terms must act as lexical units.[7] This explains why the generic term for *mouse* is *device* or *peripheral* whereas in a well-formed ontology, several steps must be taken before reaching the class Input Device (mouse → pointing device → motion-based input device → input device). In dictionaries, linguistic expressions cannot be created and considered as relevant headwords for the purpose of facilitating the compilation process.

Hence, in ontologies, terms act as labels (one of them is the preferred label; others are variants or synonyms), whereas in dictionaries, terms are considered as linguistic forms attached to specific meanings. In other words, ontologies and dictionaries are dealing with different objects: ontologies are dealing with terms = class labels or with terms = linguistic expressions listed in annotations; dictionaries are dealing with terms = form + meaning.

Looking at the labels themselves, many discrepancies can be noticed between the way ontologies and dictionaries manipulate them. First, some "labels" that are listed as class names in ontologies do not appear in dictionaries. Conversely, some units listed as headwords in dictionaries do not appear in ontologies or appear in different places (types of input devices as class names; parts listed in the properties; potentially typical uses as properties as well). Figure 7 gives a few examples of expressions found in both resources or exclusively in either the ontology or the dictionary.

5.2. Some explanations for these discrepancies

The previous subsection discussed some differences that can be observed when examining the labels in the two ontologies and the dictionary used in this study.[8] Both types of repositories are attempts to

[7] This is not the case in all specialized or terminological dictionaries, which can contain entries for multi-word terms that are compositional (e.g., *configuration file*, *input device*, *impact printer*).

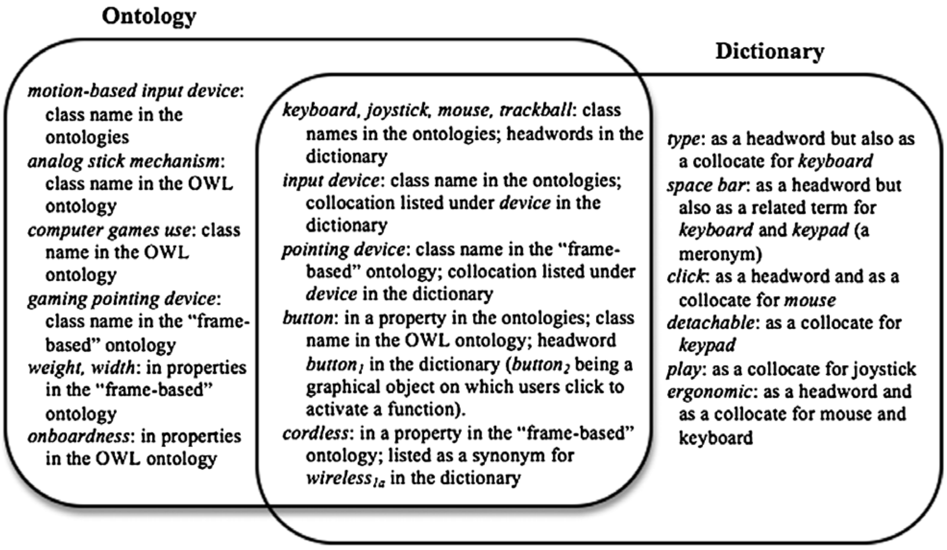


Fig. 7. Names and terms in the ontologies and the dictionary.

organize domain-related information (concepts or meanings) and describing relations between them. As stated by Hirst (2009), p. 474: "Most of the issues in the relationship between lexicons and ontologies pertain to the nature of the word senses in the lexicon and to relationships between those senses – that is, to the semantic structure of the lexicon".

Hirst is referring to the general lexicon, but we believe that the statement also applies to the specialized lexicon. This belief about the similarities between ontologies and the specialized lexicon is shared by many terminologists – and perhaps even more strongly – based on the assumption that terms reflect the organization of concepts in a field. We hope we showed that this is not completely true in the previous subsection. In this one, we will examine some of the causes of this state of affairs.

In ontologies, the focus of the description is placed on the classes (or concepts) and ontologists go to great lengths to capture the properties of the objects they represent in such a way that each class becomes distinct from others and refers to a closed set of objects. This explains why labels are secondary. In dictionaries, parts of the description focus on the form (part of speech and information derived from the fact that a term belongs to a specific part of speech: syntactic behaviour, for instance) whereas others deal with meaning (definitions, semantic relations). However, the entry itself is linked to a term considered as a linguistic form attached to a specific meaning and these two properties cannot be separated.

Ontologies, because it is essential that they capture the specificities of classes, are highly structured entities, and the relationships between classes must be clearly, logically and consistently defined, especially if inferences are made from them.

An increasing number of dictionaries formally describe relations between units (the most prototypical example for general language being WordNet) based on the assumption that the lexicon is also structured. However, some relationships can appear in dictionaries that are clearly not central in ontologies (syntagmatic relations such as collocations). In addition, although some relationships are similar (ISA vs.

[8]We have not examined the problem of multilinguality where even more differences can be observed. This issue is discussed in Hirst (2009) and Montiel Ponsoda et al. (2010). Different languages make semantic distinctions that are not always made by others.

hypernymy, for instance), the units they link can be different (e.g., the *input device* example mentioned previously).

The nature of classes (entity, activity, etc.) also has an impact on the way knowledge is represented and taken into account in ontologies. Ontologists must also make a number of decisions as to how and where to better represent the different items of knowledge. In dictionaries, relations are not always described formally, but when they are, they can be encoded based on more flexible criteria. For instance, dictionaries take into account relationships that are not defined based exclusively on logical criteria (near synonymy or collocations). The nature of meanings (entity versus activity) impacts indirectly on the contents of entries (different lists of relationships, different parts of speech), but does not affect the organization of the information per se.

Finally, the elaboration of ontologies and dictionaries – since they are dealing with very different objects – rely on diverging methodologies. The construction of an ontology is mostly top-down and requires that a clear delimitation of concepts be made (at least for the most important concepts) before starting to populate it. Adding a new class, especially at a high level of the hierarchy, often leads to a redefinition of parts of the modelling. On the other hand, the compilation of a dictionary is a bottom-up process and, since relationships between lexical units are defined between pairs of words and not from the point of view of an entire field of knowledge, entries can be added without having to reconsider the structure of the existing word list.

6. Concluding remarks

In this paper we set out to analyze the use of objects called *terms* in domain ontologies and in specialized dictionaries. Although it was carried out on small ontologies and a small set of terms in a dictionary, we believe it reveals some fundamental differences between both types of repositories and the way they address the notion of "term".

Among the major differences that were noted, are the following: (1) the focus of repositories (concepts in ontologies and terms in dictionaries); (2) the types of labels that are dealt with, the place they are assigned and their purpose; (3) the respective methodologies.

We have shown that ontologies and dictionaries are dealing with different objects and manipulate them very differently. Of course, there is a degree of overlap between both repositories, but looking at this specific analysis, this overlap is rather narrow. On an abstract level, we can consider that a correspondence between the objects considered in both repositories can be viewed as merely accidental. In our opinion, trying to exchange contents from one resource to the other will inevitably result in making several adjustments (in the ontology or in the dictionary),[9] unless one compromises on the very nature of the objects that are represented in them. When importing terms in ontologies, they must be emptied of most of their linguistic properties and – from the view of knowledge representation – their "flaws" (variation, polysemy, ambiguity). As mentioned by Hirst (2009), p. 276: "a lexicon, especially one that is not specific to a technical domain ((...)), is not a very good ontology". We believe that the statement also applies to dictionaries in specialized domains, unless these dictionaries comply with some of the requirements of ontologies (aim to organize concepts, standardize definitions and labels, etc.; in other words, mimic ontology work). We also add that a domain ontology would not be the most interesting specialized dictionary.

[9]This also applies to taxonomies versus ontologies as shown in Soergel et al. (2004).

What should we conclude from all this? Why do resources that aim at representing concepts (in ontologies) or meanings (in dictionaries) hardly contain the same labels? Part of the answer can be found in the studies mentioned in the introduction, i.e. different perspectives can be taken on terms or on the meanings or concepts they represent. We could say that ontologies focus on the cognitive dimension of terms; in contrast, dictionaries take into account both their cognitive and linguistic dimension in varying proportions. Here is another possible answer: ontologies aim to organize knowledge in an ideal way and this entails the introduction of abstract categories in order to better organize things or organize them in a way such that they can be manipulated by tools that are available; dictionaries aim to describe terms that (unfortunately, some would say) are not that well organized, exactly like the natural language to which they belong.

In this article, we limited ourselves to examining the notion of "term" and the way it is handled in specialized dictionaries and ontologies. Another topic of utmost importance is that of the types of relations that are taken into account in both repositories. A close examination of sets of relations that appear in them would probably shed even more light on the debate about terminologies versus ontologies. It would also be interesting to include thesauri (Soergel et al., 2004) in this kind of study and analyze the sets of relations that are taken into consideration in this third form of resource. This issue should be debated in another article.

Acknowledgement

The work presented in this article is supported by the Quebec funding agency *Fonds de recherche société et culture*.

References

Bioportal (2011). Available at: http://bioportal.bioontology.org/ (accessed 15 September 2011).

Buitelaar, P., Cimiano, P., Haase, P. & Sintek, M. (2009). Towards linguistically grounded ontologies. In *The Semantic Web: Research and Applications*. Lecture Notes in Computer Science (Vol. 5554, pp. 111–125), New York: Springer.

Cabré, M.T. (2003). Theories of terminology: their description, prescription and explanation. *Terminology*, *9*(2), 151–161.

Cimiano, P., Buitelaar, P., McCrae, J. & Sintek, M. (2011). LexInfo: a declarative model for the lexicon-ontology interface. *Journal of Web Semantics*, *9*(1), 29–51.

Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge Univ. Press.

DiCoInfo (2011). Dictionnaire fondamental de l'informatique et de l'Internet. Available at: http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi (accessed 15 September 2011).

Estopà, R. (2001). Les unités de signification spécialisées élargissant l'objet du travail en terminologie. *Terminology*, *7*(2), 217–237.

Gómez-Pérez, A., Fernández-López, M. & Corcho, O. (2004). *Ontological Engineering, with Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web*. London/Berlin/Heidelberg: Springer.

Hirst, G. (2009). Ontology and the lexicon. In S. Staab and R. Studer (Eds.), *Handbook on Ontologies* (pp. 269–292). Heidelberg: Springer.

Horridge, M. (2011). *A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools* (Edition 1.3). University of Manchester. Available at: http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/ (accessed 15 September 2011).

L'Homme, M.C. (2004). A lexico-semantic approach to the structuring of terminology. In *Computerm'2004* (pp. 7–14). Geneva: University of Geneva.

L'Homme, M.C. (2008). Ressources lexicales, terminologiques et ontologiques: une analyse comparative dans le domaine de l'informatique. *Revue Française de Linguistique Appliquée*, *13*(1), 97–118.

Mel'čuk, I., Clas, A. & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Belgium: Duculot/Aupelf-UREF.

Montiel Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A. & Peters, W. (2010). Enriching ontologies with multilingual information. *Journal of Natural Language Engineering*, *17*(3), 283–309.
Noy, N.F. & McGuiness, D.L. (2001). Ontology development 101: a guide to creating your fist ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory and Technical Report SMI-2001-0880, Stanford Medical Informatics. http://protege.stanford.edu/publications/ontology_development/ontology101.html (accessed 15 September 2011).
Overton, J.A., Romagnoli, C. & Chhem, R. (2011). Open biomedical ontologies applied to prostate cancer. *Applied Ontology* *6*, 35–51.
OWL 2 Web Ontology Language Primer (2011). Available at: http://www.w3.org/TR/owl2-primer/ (accessed 15 September 2011).
Protege-Frames User's Guide (2011). Available at: http://protegewiki.stanford.edu/wiki/PrF_UG (accessed 15 September 2011).
Sager, J.C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia, PA: John Benjamins Publishing.
Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. & Katz, S. (2004). Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information*, *4*(4), 1–23.
Van Rees, R. (2003). Clarity in the usages of the terms ontology, taxonomy and classification. In *Construction Informatics Digital Library* (paper w78-2003-432). Ljubljana University.
Wikipedia contributors (2011). Input device. Wikipedia, The Free Encyclopedia. Available at: http://en.wikipedia.org/wiki/Input_device (accessed 15 September 2011).
WordNet (2011). Available at: http://wordnet.princeton.edu/ (accessed 15 September 2011).