Review

The Role of E-Vocabularies in the Description and Retrieval of Digital Educational Resources

Ana M. Fernández-Pampillón

Facultad de Filología, Universidad Complutense de Madrid, Av. Séneca, 2, 28040 Madrid, Spain;
apampi@filol.ucm.es

Abstract: Vocabularies are linguistic resources that make it possible to access knowledge through words. They can constitute a mechanism to identify, describe, explore, and access all the digital resources with informational content pertaining to a specific knowledge domain. In this regard, they play a key role as systems for the representation and organization of knowledge in environments in which content is created and used in a collaborative and free manner, as is the case of social wikis and blogs on the Internet or educational content in e-learning environments. In e-learning environments, electronic vocabularies (e-vocabularies) constitute a mechanism for conceptual representation of digital educational resources. They enable human and software agents either to locate and interpret resource content in large digital repositories, including the web, or to use them (vocabularies) as an educational resource by itself to learn a discipline terminology. This review article describes what e-vocabularies are, what they are like, how they are used, how they work, and what they contribute to the retrieval of digital educational resources. The goal is to contribute to a clearer view of the concepts which we regard as crucial to understand e-vocabularies and their use in the field of e-learning to describe and retrieve digital educational resources.

Keywords: vocabularies; e-vocabularies; digital educational resources; digital resources retrieval; information retrieval; e-learning; lexicography

1. Introduction

Vocabularies are linguistic resources that make it possible to access knowledge through words [1]. In this regard, they can constitute, in their digital format, a mechanism to identify, describe, access, and explore all the digital resources with informational content (e.g., documents, websites, and educational software) pertaining to a specific knowledge domain [2,3]. Thus, they play a key role as systems for the representation and organization of knowledge in work environments in which information is created and used in a collaborative or free manner, as is the case in digital teaching and learning environments, e-learning environments, or the web [4].

E-learning environments are spaces created on the Internet which integrate digital educational resources as a fundamental didactic component. E-learning environments currently constitute one of the basic elements in educational institutions and in professional training and updating services. By digital educational resources, also known as learning objects, we understand digital entities (files or file groups) which have at least one defined learning objective. Some examples of digital educational resources are video lessons, e-books, presentations, notes, exercises, and assessments. On the Internet, digital educational resources are usually located in educational digital repositories (also known as Learning Object Repositories). An educational digital repository is an online system to store, publish, retrieve, and reuse educational resources [5]. Two good examples are MERLOT [6] and ARIADNE [7]. The main advantage of these repositories is that they facilitate to the maximum location and selection

of the resources sought as well as their integration in e-learning environments, such as virtual courses in virtual campuses. To facilitate the location and selection of resources, mechanisms for identification and description are used—mainly metadata and e-vocabularies.

This review article focuses on one of the aforementioned mechanisms: e-vocabularies. The article describes, using a semasiological approach, what vocabularies are, how they are used, how they work, and what they contribute to retrieve digital educational resources. The final goal is to answer the main question—what is the position and purpose of e-vocabularies in the description, organization, and retrieval of digital educational resources used in e-learning environments?

To answer this question, a deductive–inductive method based on successive refinements has been applied. Starting from a theoretical core and a data sample, a double analysis is conducted going (1) from theory to the sample and (2) from the sample to the theory, comparing and trying to fit the theoretical proposals with the sample items. The theoretical core comprises the works presented in [8–10], and the sample comprises the relevant examples of open educational repositories [6,7,11,12], picture [13–16], and vocabularies [17–26]).

When a fit is found, that is, when a theoretical proposal used in any of the sample items is found or vice versa, the answer is refined by searching in the literature for information that makes it possible to (a) go deeper into the theoretical proposal; and (b) verify that it has been applied to other relevant examples.

Finally, the results are presented in the article following the same deductive–inductive format as in the analysis: the theory and the actual examples that support it. The article is structured to answer the main question incrementally, by steps: Section 2 provides a discussion of what a vocabulary is; Section 3 discusses what a vocabulary is like; Section 4 focuses on how vocabularies work in Information Retrieval; Section 5 specializes Section 4 to digital educative resources coming increasingly close to the main question: how vocabularies are used to describe and retrieve digital educational resources; Section 6 answers how to describe digital educational resources albeit refining in each possible e-vocabulary type. Finally, Section 7—the summary and conclusions—answers the main question and provides the view from the author about what the role of e-vocabularies in the description and retrieval of digital educational resources is.

2. What Is a Vocabulary?

The term *vocabulary* is ambiguous, and its meaning depends on the discipline of which it is part and the context in which it is applied. For example, in the general knowledge domain of a language such as Spanish, "vocabulario" has seven meanings (Figure 1): "(1) the set of words in a language; (2) a dictionary (book); (3) the set of words used in a region, a profession, a writer's semantic field, etc.", or simply "(4) the book in which they are included; . . . " [27]. In more specialized domains, vocabularies have a more precise meaning, form, and application. In such domains, a word (or collection of words in several languages) designates a single concept which is called a *term* and the set of terms is called a *terminology* [28,29]. Although *vocabulary* is not synonymous with *terminology*, *vocabulary* will be used in this article to cover not only words in generic language also but domain-specific words (e.g., the International Standardization Organization, ISO, uses the term *vocabulary* to denote some of their terminology standards such as ISO 472:2013-Plastics—Vocabulary). Additionally, if vocabularies are built and used in electronic format, they are referred as *e-vocabularies*. Since the majority of vocabularies are now built in electronic format, throughout the rest of the article "vocabularies" and "e-vocabularies" are used interchangeably.

vocabulario

Del lat. mediev. *vocabularium,* y este der. del lat. *vocabŭlum* 'palabra'.

1. m. Conjunto de palabras de un idioma.

2. m. **diccionario** (‖ repertorio en forma de libro).

3. m. Conjunto de palabras de un idioma pertenecientes al uso de una región, a una actividad determinada, a un campo semántico dado, etc. *Vocabulario andaluz, jurídico, técnico, de la caza, de la afectividad.*

4. m. Libro en que se contiene un **vocabulario** (‖ conjunto de palabras de una región o actividad).

5. m. Catálogo o lista de palabras, ordenadas con arreglo a un sistema, y con definiciones o explicaciones sucintas.

6. m. Conjunto de palabras que usa o conoce alguien.

7. m. coloq. Persona que dice o interpreta la mente o dicho de otro. *Hablar por vocabulario. No necesitar de vocabulario.*

Figure 1. Seven meanings of the word "vocabulary", "vocabulario" in Spanish (source: [27]).

In the domain of Library and Information Science, a vocabulary is called *term vocabulary* and it is defined as a set of terms of conventional syntactic procedures used to represent the contents of a document in order to retrieve it [30,31]. They are also known as *documentary languages* and provide a common, universal system for classification of bibliographic works and documents [31].

This concept is also shared by the vocabularies used for Information Retrieval (hereinafter, IR). In this domain, vocabularies are used as components of the IR software systems to prevent linguistic ambiguity and polysemy [8]. These are *controlled vocabularies* that are defined as lists of terms, explicitly listed, unambiguous, and non-redundant which contribute a conceptual description and a pragmatic and empirical dimension to the information domain.

In the field of Computational Linguistics, Natural Language Processing, and Linguistic Technology, vocabularies are also called *computational lexicons* or *lexical databases*. They are conceived as databases and lexical knowledge bases designed for automatic processing of natural languages [32]. In these vocabularies, lexical knowledge—at the phonetic and phonological, morphological, syntactic, semantic, pragmatic, discourse, and world levels—becomes explicit and is organized with formal data models that are suitable for automatic natural language processing [33–35]. A widely used example of a computational lexicon is WordNet [17]. Vocabularies constitute a basic component in the architecture of Natural Language Processing Systems: they are necessary for the development of linguistic applications such as spelling and style correctors, retrieval of textual information, and the indexing and description of documents and textual resources. Two sources of computational lexicon distribution are, for example, the European agency ELRA [36] and the US consortium LDC [37].

Finally, in the field of e-learning, vocabularies are a mechanism for semantic representation that allows human agents or software to locate, interpret, and retrieve educational content or to process it for didactic purposes. Vocabularies are currently mainly applied to solve two questions: (1) the representation and retrieval of digital educational resources; and (2) the interoperability between e-learning tools and content [9,10,38]. In the former case, the main type of vocabularies used are taxonomies and thesauri [10], that is, vocabularies of terms organized into categories and/or interconnected by hypernymy–hyponymy and other semantic relations. In the latter case, the type of vocabularies used are thesauri and ontologies that conceptually represent the dimensions of an e-learning system: the agents, tools, domain of knowledge, methodologies, and teaching models, etc. [39]. In the following sections, vocabulary types and examples of each of these types will be seen.

3. What a Vocabulary Is Like

A vocabulary, in any of its interpretations, comprises at least a lexical unit which can be general or specific to a domain [40]. The words or word groups that constitute it are presented in a standard form (i.e., canonical form) if they are inflected words and can be organized into classes or categories. The overall organization of the vocabulary is called its *macrostructure*. The description of each word is given in what is known as an *entry* (Figures 1 and 2). The contents of entries, known as their *microstructure*, depend on the purpose of the vocabulary [41] but can include, in addition to their meanings, grammatical, usage, phonetic, and etymological information, and relations with other vocabulary terms.

Figure 2. Lexical entry of the word "lesson" in WordNet.

Several types of vocabulary, on the one hand, can be distinguished concerning macrostructure and microstructure [9]: (i) lists of terms; (ii) glossaries; (iii) classifications and taxonomies; (iv) thesauri; (v) ontologies; (vi) dictionaries; and (vii) lexicons [10]. In Section 6, the way in which each of these vocabulary types is configured will be shown in more detail, as well as some examples used in the retrieval of digital educational resources.

On the other hand, based on the procedural criterion of the output of a lookup strategy using vocabularies, two fundamental vocabulary types can be considered [40]:

- Semasiological vocabulary: the lookup key is the word (form) and the information required is semantic. Usually different meanings of a word can be considered as well as interrelations (mainly semantic) between words.
- Onomasiological vocabulary: the lookup key is a concept and the information required is the word form or, more precisely, the term that designates the concept (and, if necessary, different terms correlated with it).

On that bases, this section and the rest of the article are mainly following a semasiological approach, although without ignoring the onomasiological angle. In this respect, the term "term" will be used to denote a word that is assigned to a concept used in special languages that occur " . . . in a domain or subject, and characterized by the use of specific linguistic means of expression . . . " [42].

A vocabulary can also be conceived as an *index* that establishes a correspondence between the orthographic form of a word and the information about that word [40]. However, it may not be a one-to-one correspondence, as in those cases in which a word has different syntactic categories, in homographs, and in polysemy, a single word form can have different entries. This would be mainly the case of dictionaries and lexicons. In other cases, a biunivocal word-information correspondence is necessary, as in *controlled vocabularies* for IR and retrieval of digital objects. Controlled vocabularies

try to ensure a biunivocal correspondence by a control process that disambiguates homograph terms using qualifiers to specify, in the case of synonymy, a single term as the preferred term, also known as the descriptor (Figure 3), and, in the case of polysemous terms, by restricting meaning through a domain note [9].

cranes (birds)
cranes (lifting equipment)

mercury (metal)
Mercury (planet)
Mercury (Roman deity)

seals (animals)
seals (law)
seals (numismatics)

socialization (economics)
socialization (social psychology)

Figure 3. Qualifiers, in parentheses, to distinguish homographs (source: [9]).

The meaning of a lexical entry is expressed in its *definition*. A definition is "a descriptive representation of a concept which serves to differentiate it from related concepts" [29]. The definition can take many forms, from a simple statement to references based on the establishment of semantic relations between the lexical units (or terms in specific languages) in the vocabulary. This latter approach—the definition of meaning based on semantic relations—is preferred in vocabularies aimed at information retrieval (IR) (Figure 4).

school
Broader Terms

BT educational institution

More specific terms

NT3 alternative school
NT3 boarding school
NT3 boys' school
NT3 coeducational school
NT3 dancing school
NT3 day school
NT3 denominational school
NT3 European school

Figure 4. Definition based on semantic relations (source: European Education Thesaurus http://vocabularyserver.com/tee/en/index.php).

Classic lexical *relations between word meanings* are equivalence, opposition, inclusion, co-hyponymy, and part–all [13]. The meaning equivalence relation is *synonymy*. Two or more words are synonymous if they can be replaced in any context without changing meaning. This condition of equality in all the meanings of a word explains why absolute synonyms are scarce and, for practical purposes, a less strict definition is preferred, so that two or more words are synonymous with respect to a meaning if they can be exchanged in that specific context. Figure 2 shows the synonym example, deterrent example, lesson, and object lesson with respect to the second meaning of a lesson.

The opposite relation to synonymy is *antonymy* (Figure 5). In the lexicographic tradition, antonyms are defined as words with opposite meanings and, as such, as opposed to synonyms. However, this definition of antonymy is too vague. In [43], the notion is refined, taking into account the fact that antonymy entails similarity between the terms. Antonymous terms are always composed by common semes, i.e., semantic features: thus, brother and sister share the semes /human being/ and /born from the same parents/, and are opposed by the seme /pertaining to sex/. Consequently, the antonymy relation can be defined as the relation that binds two words in the same grammatical category that share part of their semes and that have an opposing counterpart. Antonymy is also composed of different types of oppositions, mainly binary ones. Three types of antonyms are usually distinguished: (1) Contradictory or complementary antonyms, which express an exclusive disjunction relation, that is, the negation of either one of the words implies the affirmation of the other one and the two terms cannot be simultaneously denied, e.g., man/woman, present/past. (2) Contrary or gradable antonyms, which define the poles in an implicit gradation scale and allow the existence of intermediate degrees. Big/small, heat/cold, love/hatred. (3) Reciprocal antonyms are those that force the replacement of one by the other in a given statement to preserve the relation, e.g., doctor/patient, parent/child.

Antonyms for learn

ignore	neglect	overlook
lose	pass	teach
miss	release	
misunderstand	spend	

Roget's 21st Century Thesaurus, Third Edition Copyright © 2013 by the Philip Lief Group.
Cite This Source

Figure 5. Antonyms of the word "learn" (Source: Thesaurus.com website http://www.thesaurus.com/browse/learn/).

The main inclusion relation is generalization or *hypernymy*, and its inverse relation is specialization or *hyponymy*. An example of hypernymy–hyponymy relations is shown in Figure 4. The hypernym appears marked as the Broader Term (BT), and the hyponyms are marked as Narrow Terms (NTs). Hypernymy–hyponymy relations are asymmetric and transitive so they "arrange" words into simple or multiple hierarchies from more general to more specific meanings. Structures are usually lattices, but the term hierarchy is used to emphasize the conceptual dependence of hyponyms on hypernyms.

Co-hyponymy within the same hierarchy is established between the words related to the same hyperonymy, e.g., in Figure 4, "alternative school" and "boarding school" are co-hyponyms of "school". Co-hyponyms are distinguished by one or more delimiting characteristics. A delimiting characteristic is a "necessary characteristic that distinguishes a concept from related concepts" [29]. Unlike in antonymy based on a binary opposition, the negation of one of the co-hyponyms does not necessarily imply the affirmation of the other co-hyponym, but the choice remains open: if x is not an "alternative school", it can be a "boarding school" or a "dancing school". Co-hyponyms are mutually exclusive: a "school" is an "alternative school" or a "boarding school" or a "dancing school", etc. Moreover, co-hyponyms can have synonymy or antonymy relations between each other. For example, "college" and "university" can be regarded as co-hyponyms of "school" and can pass for synonyms; contrariwise, "virtual learning" and "face-to-face learning", co-hyponyms of "learning", are antonymous [43].

The part-all or *holonymy/meronymy* relation is also a hierarchical, antisymmetric, and transitive relation in which one of the terms denotes a part and the other the whole pertaining to that part (Figure 6). The difference with respect to the hyponymy–hypernymy relation is that it is a relation of belonging rather than inclusion, so meronyms do not inherit the homonyms' attributes. Meronymic dependence relations are varied and complex: member/set (teacher/faculty), component/assembly

(leg/chair), portion/mass (slice/cake), material/object (steel/bicycle) [43]. In the WordNet vocabulary, three types of holonymic relations are used: member, substance, and part [14].

- S: (n) **school**, schoolhouse (a building where young people receive education) *"the school was built in 1932"; "he walked to school every morning"*
  - *direct hyponym* / *full hyponym*
  - *part meronym*
    - S: (n) classroom, schoolroom (a room in a school where lessons take place)
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - ***part holonym***
    - S: (n) school system (establishment including the plant and equipment for providing education from kindergarten through high school)

Figure 6. Meronyms and holonyms with respect to the second meaning of "school" (source: WordNet [17]).

In addition to "classic" lexical relations, vocabularies can include many other semantic *associative relations*, such as family (fire/fireman), agent (student/learning agent), instrument (exercise/practice instrument), or location (school/student location).

As will be seen in the following sections, semantic relations, particularly those of synonymy and hyponymy–hypernymy, are basic for building vocabularies for the classification and indexing of information and of digital educational resources.

4. How Vocabularies Work in Information Retrieval

Information retrieval (IR) is a field of Artificial Intelligence whose object is the search for information in documents, document search, the search for metadata that describe documents, and the search for data in databases, be it via the Internet, an intranet, for texts, images, sound, or other data, in a pertinent and relevant manner [15]. Retrieval of digital educational resources is thus a particular case of IR. In this section, the basic IR mechanisms that are also shared by *retrieval of digital educational resources* are reviewed.

In the context of IR, vocabularies are *information organization systems* whose purpose is to increase the effectiveness of the processes of indexing, search, or navigation of collections of digital resources with informational content (hereinafter, simply resources). They constitute a "bridge" between information representation at a high level of abstraction and representation at the level of computer information system, as well as between the natural language in which people make a search query and the formal query languages used by machines [8,10,16]. The idea is to use the terms to represent the contents of resources and relations between the terms in the vocabularies to improve the exhaustiveness and precision of the indexing results and the search for resources, as well as the usability of the user interface.

In this regard, *exhaustiveness* in search processes improves if the terms employed by the user in their request (i.e., "dog") are expanded by using all the possible terms related either orthographically (i.e., "dog", "doggie"), in what is known as the form control, or else by their meaning, through synonym control (e.g., "pooch", "canine", "cur"), quasi-synonyms, hyponyms/hypernyms (e.g., "quadruped", holonyms/meronyms (e.g., "pack") and other associated terms. Regarding the improvement in *precision*, vocabularies make it possible to combine terms to fine-tune the user's query, what is known as coordination. They make it possible to distinguish between homographs, use definitions to disambiguate, restrict, or clarify the meanings of the query, and the frequency of use of the terms to statistically differentiate the most likely terms on the basis of use, context, or even user profile. Finally, as regards interface *usability*, vocabularies offer users an alternative system for access to resources as a terminological-conceptual map of the resource collections (Figure 7).

Figure 7. Vocabulary for navigation in the archeological digital repository Coclé in Panama (source: http://oda-fec.org/cocle/).

The functioning of vocabularies in the indexing, search, and navigation of resources are discussed in more detail in the next two subsections.

4.1. Vocabulary in Resource Indexing

The first process for IR is *resource indexing*. In indexing, resources are represented by means of terms and the indices that link terms and resources are created (Figure 8). A small, limited set of terms that represent the contents of the resource is selected and associated with this resource. The terms may come from a pre-existing controlled vocabulary, and it is then termed assignment indexing by assignation; otherwise, it can be extracted from within the resources themselves, what is known as extraction indexing [8]. The index generated could be used, if so wished, as a terminological map for users to explore and retrieve the collection resources (Figure 7).

Indexing is basically carried out in *three stages*: (1) analysis of the domain (collection of resources), which entails the extraction of the key terms in each of the resources; (2) the selection of the most representative terms of a controlled vocabulary to describe the resources; and (3) the creation of data structures (called indices) to store the associations between resources and terms (Figure 8).
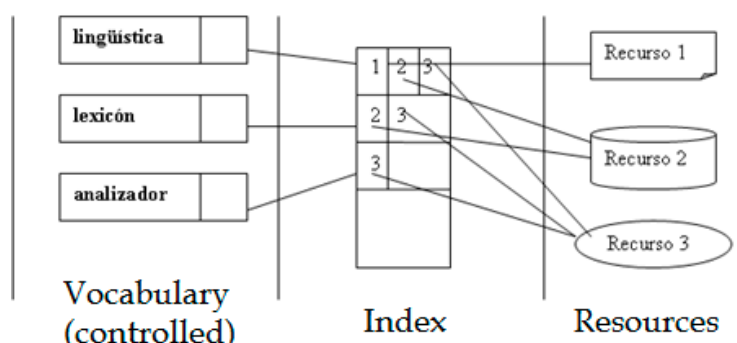
Figure 8. Indexing of digital resources with vocabularies.

The indexing process can be *manual, automatic,* or *semiautomatic* [2]. In the first case, an expert (or expert committee) analyzes the contents, context, and structure of each resource and assigns to it the set of terms from a controlled vocabulary that best describes it (categorization by assignment).

Another option in this first case is the extraction of the terms that are most representative of the resource or its metadata (categorization by extraction). Manual indexing has the advantage of having a high degree of precision in the description of resources, particularly in non-textual resources such as images or software applications. However, it is quite costly and in some cases lacks consistency in descriptions due to differences in criteria among indexers. The data provided in [44] can serve as a sample: " ... Yahoo uses 200 employees to index web pages in accordance with its 500,000-term taxonomy; MEDLINE (the national medical library) [18] spends $2 million per year on the indexing of articles with the MeSH thesaurus ... " [19].

Automatic indexing, in the second case, uses algorithms that statistically analyze the words in the contents or metadata of the resources [45]. Metadata are a set of properties and values that describe a resource (Figure 9). In indexing by automatic extraction, patterns of word behavior are identified on the basis of such variables as frequency of use, placement, order, and proximity. In this way, the words and relations that best represent the contents of the resource are identified [46]. In automatic indexing by assignment, the words extracted are compared to the terms of a controlled vocabulary to select those that are most similar as descriptors of the resource. The result are clusters of resources that show similar content patterns, tagged by means of the sequence of terms extracted from the controlled vocabularies or within the resources themselves , which best represent their contents [8,47]. As a sample of automatic indexing systems (see [48–50]) and American Society for Indexing list of software [51]. Automatic indexing is faster, less costly, and more consistent than manual indexing, but the level of accuracy of descriptions is lower and usually requires a certain degree of human involvement to correct the results obtained.

```
<html>¶
<head><title>Introduction to Prolog programming</title>¶
<meta name="DC.Type" content="text">¶
<meta ¶
    →     name="DC.Subject" ¶
    →     scheme="UCD" ¶
          content="Language.Linguistics.Literature/Linguistics and 
          Language/Languages/Artificial Languages">¶
<meta name="DC.Description" content="This document is an introduction 
          to the Prolog logical programming language. It is a teaching 
          element for the Computational Linguistics course in the 
          Linguistics degree at the UCM School of Philology">¶
<meta name="DC.Creator" content="Ana Fernández-Pampillón">¶
<meta name="DC.Title" content=" Introduction to Prolog programming">¶
<meta name="DC.Identifier" scheme="URI" 
          content="https://campusvirtual.ucm.es/linguistica-informati
          ca/intro.html">¶
................¶
```

Figure 9. Use of the UDC in the Dublin Core metadata. The subject attribute (DC.Subject) takes as its values a set of related vocabulary terms (hypernymy).

In the third case, *semiautomatic* or *hybrid indexing* systems combine human involvement to identify the potential meanings of the resources and the efficiency of automatic indexing [47]. They are able to

learn with experience, which makes them increasingly efficient. One example is the Verity documentary indexing systems used by Oracle [52].

To *create indexing vocabularies*, three points should be taken into account: firstly, in the design stage, it should be verified whether there are *standard reference vocabularies* for the domain of knowledge of the collection of resources that can be used totally or partially before creating a new one ad hoc. The reason for this is twofold: not only does it reduce costs, but also facilitates interoperability between repositories and IR systems, for example, to transfer the vocabulary to potential new versions of the repository or to other information systems. If there are no such vocabularies and one must be created from scratch, the standard guidelines for the creation of vocabularies should be followed, to guarantee the interoperability of the vocabulary in every case [9,10,53–56].

Secondly, it should be taken into account whether the vocabulary will be used in *combination with metadata* to describe the resources. This is a frequent option, which is described in Section 5. Basically, it consists in using the terms of one or several vocabularies to assign values to the properties of the resource metadata. Figure 9 shows use of the Universal Decimal Classification (UDC) to describe a resource in html format. The metadata model is Dublin Core [57], which is probably the one most frequently used on the Internet.

Thirdly and finally, when creating new vocabularies, it should be borne in mind whether they will be also used in the IR system *search*, *navigation*, and *personalization* systems. This option has the advantages that not only are the costs of creation capitalized, but also the maintenance of the conceptual and designatory consistency of the collection resources is facilitated, which helps users to have a mental image of what the collection of resources is like, and thus locate more easily the resources sought.

4.2. Search and Navigation Vocabulary

To understand the *role of vocabularies* in search and navigation, the following types of IR and online resource retrieval systems should be distinguished: browsing, searching, and filtering systems [47].

1. *Browsing systems* offer users an organized term structure, which is the vocabulary, which includes the information resources and a mechanism that allows users to browse the vocabulary to locate the resources sought. The way in which the vocabulary is presented ranges from simple navigation in an alphabetical list restricted to the vocabulary terms (Figure 10) to hybrid systems that make it possible to select terms while browsing through the vocabulary to combine them in an expression that describes what is being sought [58].

2. *Search systems* offer users the possibility of defining the resource(s) sought by entering the terms that describe said resource in a text box (Figure 11). These systems are suitable if users are able to express in sufficient detail what they are looking for, that is, they know beforehand which information or resources they seek (and consequently do not need to "browse" as in the previous case). The vocabulary in this kind of system basically serves as a terminological reference system for users to use the same terms employed by the IR system to index the information or resources. In this way, better search results are guaranteed (recall and accuracy). In browsing and retrieval systems, users interact in real time with the IR search module.

3. *Filtering systems* allow users to define their content preferences in their profile. These preferences (e.g., Spanish Literature) are used by the IR system to filter the resources and offer users only those resources in which they are interested. In this kind of system, the vocabulary allows users to select the terms that define their preferences, which are the same terms used by the IR system to index the resources. One example of use of this system is the "My MERLOT" option in the MERLOT educational repository (Figure 12).

Figure 10. Presentation of the Thesaurus for Graphical Material (http://www.loc.gov/rr/print/tgm1/toc.html) for browsing through the images in "US Steel Gary Works Photograph Collection, 1906–1971" (http://www.dlib.indiana.edu/collections/steel/index.html).

Figure 11. Vocabulary (ETB thesaurus) of the Spanish AGREGA educational repository (http://agrega2.red.es/buscador2/BuscarAvanzadoCU/MostrarFormularioAvanzadoAnalizaPulsacion.do).

Figure 12. "My MERLOT" in the MERLOT repository allows personalized filtering of content (www. merlot.org/merlot/index.htm).

To conclude, vocabularies for resource searches play *two basic roles in IR*: (i) serving as the *language of reference* to make queries and index the resources; and (ii) *guiding* users, or the system, *in the search* by offering them all the potential options for information and resource retrieval through a map of related terms.

In this respect, the key issue solved by vocabularies is that of making the *users' query language correspond to the indexing language in the IR system*. Thus, two operations are carried out in the search process: analysis of the user's query, and its translation into the resource indexing language. If users make their query the same language that is employed for indexing, the search process is simplified and results are optimized, as the analysis of the query only involves checking that the terms used by the user belong to the IR system vocabulary, and the translation is limited to an adjustment between the terms of the query and the terms of resource indexing.

Despite the effectiveness of this solution, there are, however, two drawbacks that restrict it: users are obligated to express what they are searching for using a limited vocabulary that is not always capable of expressing what they really need [8], and users also need to be familiar with that vocabulary. Two approaches have been put forward to solve these drawbacks: (i) expressing the query in natural language that is as close as possible to users, and analyzing the query using natural language processing techniques to translate it into the IR system indexing language. The second approach (ii) consists in dynamically adapting the vocabulary, automatically or manually, so that it approaches the user's vocabulary. This approach entails applying an update methodology that includes the criteria to accept new terms and relations without losing the vocabulary's consistency [2].

Finally, to understand the role of vocabularies in IR, it is also necessary to know *when and why vocabularies fail* as "intermediaries" between users and IR systems. The two main sources of failures that can be attributed to vocabularies are the specificity of the vocabulary and spurious relations [8].

Specificity is probably the factor that has the greatest impact on search effectiveness. When a vocabulary is very specific, it makes it possible to describe a resource by means of many terms or small categories; this means greater indexing accuracy but also complicates resource location, as users must have in-depth knowledge of the vocabulary to express the query with sufficient accuracy. By contrast, if the vocabulary is general, users are more likely to find the resource(s) sought, using concepts with a broad meaning, but the results obtain will be irrelevant in a high percentage (Figure 13). The goal is thus to achieve a compromise between generality and specificity in such a way that the vocabulary terms are sufficiently specific to make it possible to retrieve the information or resources in a sufficiently fast and accurate manner (with the minimum number of undesired results), but the vocabulary should also contain sufficiently general terms to respond to the queries of users that are not experts in the domain of knowledge—e.g., learning students—or in the IR system vocabulary.
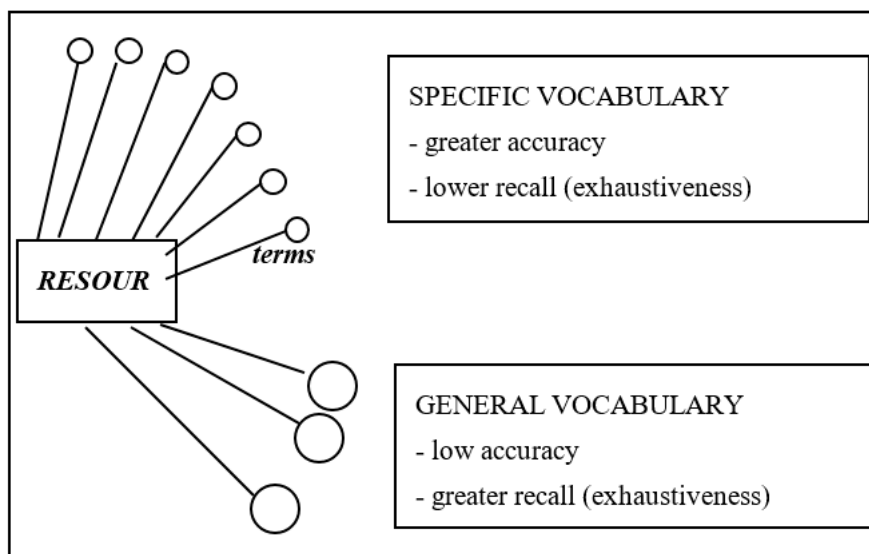
Figure 13. Ratio between the specificity of a vocabulary and the accuracy and exhaustiveness of the IR [8].

The second source of failure in IR vocabularies is *uncontrolled synonymy relations* and *ambiguous relations* between terms. Synonymy control is necessary to make the IR system indexing language coincide with the user's query language. For example, if a user searches for information about "car" engines, they should also receive information that is indexed as "automobile" engines. The ambiguity of relations appears in systems that index resources by using multiple terms but do not coordinate, that is, do not take syntax into account. They give rise to such failures as false term coordination or incorrect relations. False coordination takes place, for example, when a document on "engines for clean nuclear fusion" is ambiguously indexed under the terms "engine", "fusion", "nuclear", and "cleaning" when it should be indexed under the coordination "nuclear fusion engine". If the query searches for documents about "engine cleaning", the false coordination between "engine" and "cleaning" in the indexed document will include it as an erroneous search result.

To summarize, vocabularies play the role of the language of reference in IR systems to ensure effective communication between users and the system. However, this role is not free of problems which should be taken into account to make it effective. The first problem is that the vocabulary used by the IR system and the vocabulary used by users to make queries are often not the same. The second problem arises if coordination (combination of terms) is not properly performed when retrieving information or resources. Consequently, a vocabulary will be effective in an IR system if it is close to the users' vocabulary and if coordination of the indexing and query terms is properly performed.

5. How Vocabularies Are Used to Describe and Retrieve Digital Educational Resources

Retrieval of digitalized resources from the Internet and from large repositories is a specific case of IR, so use of vocabularies is basically the same but with distinct characteristics which should be taken into account to improve its effectiveness in the specific context of educational resources.

As an IR system, vocabularies for the retrieval of digital educational resources are used as a reference system *to describe and retrieve* those resources. The resource repository or storage system uses these descriptions to index and then find them. Three approaches can be taken to describe and retrieve digital resources: metadata, vocabularies, or metadata with vocabularies. Metadata are a set of properties and values that describe a resource from multiple points of view—content, authorship, format, intellectual property, purpose, etc. (Figure 14).

Figure 14. Example of Dublin Core metadata (source: Christine Apicul http://slideplayer.com/slide/ 1515416/).

In the first approach, resources are described and indexed using a common *metadata scheme*. The metadata schemes used are generally standard to facilitate interoperability between repositories [59]. Interoperability is the ability to take resources from a repository or e-learning space and use them in another location with another set of tools or platforms. The standard schemes most frequently used for the description of educational resources are Dublin Core [60] (Figure 14) and Learning Object Metadata (LOM) [61] (Figure 15).
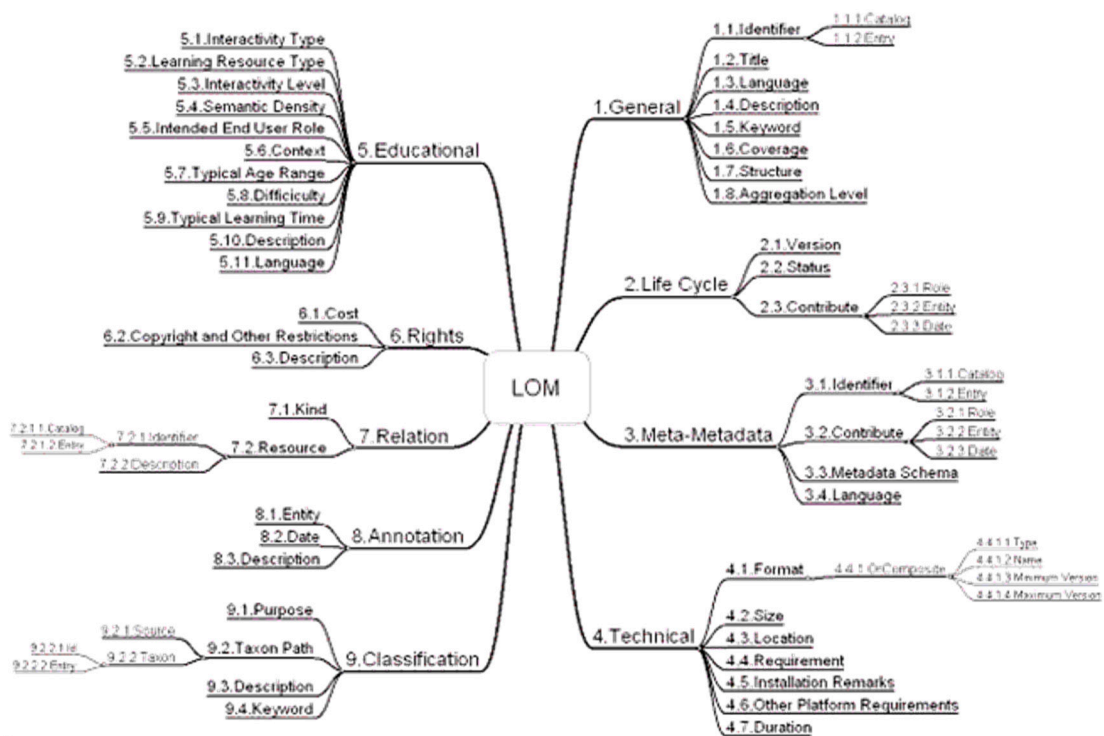
Figure 15. Scheme of the IEEE Learning Object Metadata (LOM) metadata standard (source: IMS Global Learning Consortium https://www.imsglobal.org/metadata/mdv1p3/imsmd_bestv1p3.html).

Retrieval of the educational resources in a repository is carried out on the basis of the properties and values of its metadata. To do so, search forms (Figure 16) or navigation menus (Figure 17) are used which show the properties (also known as attributes) of the metadata for users to select the values sought.

Figure 16. Resource search form based on the LOM properties of the MERLOT educational repository (source: https://www.merlot.org/merlot/advSearchMaterials.htm).

Figure 17. Navigation menu and metadata search in the MERLOT international repository.

Metadata-based retrieval is very effective but not flexible. Indexers and users must follow the repository's metadata scheme or storage system, which can be a problem for two reasons: (i) in highly specialized disciplines (e.g., pre-Columbian archeology in Figure 7), the properties and values of metadata schemes are too general and cannot describe the resources; and (ii) the values used for properties (e.g., the subject property in the Dublin Core scheme) may be not fixed, so users can employ values that are not those used in indexing. This has a negative impact on the effectiveness of retrieval.

The second option is based on use of *indexing languages* [62]. As described in the previous section, an indexing language is a vocabulary, and in some cases a set of syntactic rules that standardize the combination of terms in resource indexing. Two types of indexing languages can be established: (a) those that use terms from controlled vocabularies; and (b) those that use the symbols of bibliographic classifications and formal knowledge representation systems, such as ontologies [63]. In both cases, indexing languages provide a shared vocabulary to express the indexing and search terms and to create navigation menus.

The limitation of indexing languages as opposed to metadata is that they only make it possible to describe one property or aspect of the resource collection, usually the topic, and no other properties or attributes are used that could help refine the search, such as key words, language, type of material, technical format, etc. (Figure 16). Moreover, they pose the problems, mentioned in the previous section, of differences between the indexing terminology and the user's terminology, and a lack of accuracy in coordination, which have an impact on effectiveness of retrieval.

A third option is *joint use of metadata and indexing languages*. The idea is to describe educational resources using metadata schemes in which certain properties take as their values the terms of one or several vocabularies set or recommended in the metadata scheme. If the metadata and vocabularies are standard, they also solve the problem of the interoperability and compatibility between repositories and e-learning platforms, allowing the exchange, shared use, and integration of resource collections. In this regard, it is worth pointing out the role currently played by vocabularies as "mediators" (software modules that employ the encoded knowledge to create information at higher level than the applications that they integrate and connect) in the sharing and exchange of resources between people and computer systems (e.g., repositories and e-learning platforms) [64]. MERLOT [6], EUROPEANA [11], and ARIADNE [7] are examples of repositories that use standard metadata and controlled vocabularies.

The vocabularies used with standard metadata meet a number of requirements: authority (i.e., they are evaluated by a committee or recognized authority), stability (i.e., they do not disappear or change frequently), maintenance (i.e., there is someone responsible for their being operative on a permanent basis), dissemination, coverage, multilingualism, adaptation to users' needs, degree of compliance with standards, and specifications such as those of [10] and [61]. In this respect, it is important to point out that no ad hoc vocabularies are allowed, nor is using folksonomies, which are categorization systems that describe a resource domain by means of tags created collaboratively: for example, the categorizations of *del.icio.us* websites, which categorize favourite links and pictures on *Flickr*. The vocabularies recommended for metadata include the DDC [20] and LCSH [58] general classification systems, the vocabularies of specific scientific area such as the Medical Subjects Heading [19], the NASA Thesaurus [22], the ACM taxonomy [23], or general vocabularies such as Cyc [24], WordNet [17], the UNESCO thesaurus [25], and the ARIADNE thesaurus [65] (Figure 18).
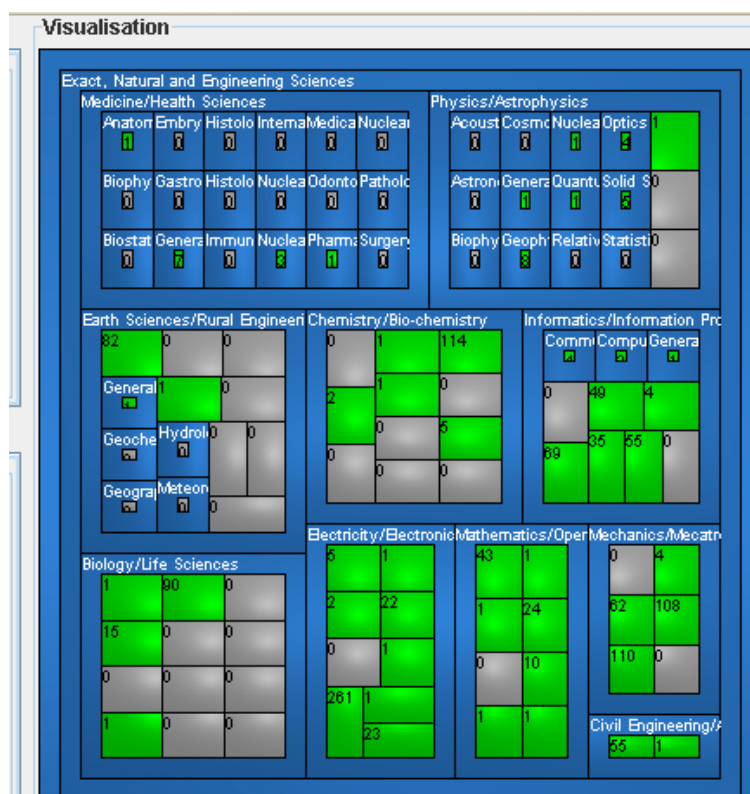
Figure 18. Presentation of the ARIADNE thesaurus for the Knowledge Pool repository. The terms are structured by means of the generalization–specialization relation. The numbers indicate the number of resources indexed by each term.

These recommendations are meant to ensure the interoperability and reusability of the metadata schemes and vocabularies. However, as for IR vocabularies and standard metadata, the vocabularies recommended for metadata have two drawbacks: (i) they require an effort on the part of the authors, the indexers, and the users of educational resources repositories to learn the vocabulary; and (ii) it is not guaranteed that the vocabulary will be able to accurately describe all the educational resources [66–68], particularly in highly specialized academic contexts [69]. These drawbacks can be solved, as described in the previous section, by adapting the standard vocabulary to the knowledge domain of the educational resources, but this has a significant cost, as it requires (i) being fully familiar with the vocabulary (its terms, expansion rules, and technology); (ii) knowledge of the specialized vocabulary, and (iii) having sufficient technical resources and staff to make the adaptation. This cost frequently accounts for the fact that new vocabularies are created, which specifically target the knowledge domain of the collection of educational resources and the educational community that uses them [10,66,69].

Finally, it is important to highlight *three characteristics of use of vocabularies* to retrieve digital educational resources. The first one is that, despite the recommended standards, the vocabularies used are frequently *non-standard* and *highly specialized* in the knowledge domain of the resources described, targeting the educational communities that use them. This, however, is not the case with other contexts for retrieval of digital resources, such as digital libraries [10,67,69]. One of the reasons that may account for this non-standard use is, as was previously stated, that the authors, indexers, and users of the educational resources cannot spend the time and effort required to learn the recommended standard vocabularies, or do not wish to spend it because they are not certain that they adequately represent their educational resources.

Secondly, retrieval vocabularies are frequently used as a *conceptual representation system* for the knowledge domain to which the collections of educational resources belong. These representations

or conceptual maps are useful not only to find what is sought but also to support and reinforce the learning of the domain terms and concepts [70,71].

Thirdly, it is important to point out that the vocabularies for the retrieval of digital educational resources are often created in an *inductive* and *collaborative* manner by the groups of teachers and students that use them, by means of virtual learning environments [72,73].

To summarize, vocabularies are used in the retrieval of educational resources as in IR, but embedded in the standard educational metadata that document resources. The vocabularies pertaining to the contents of the resources or their teaching use are often created as needed and collaboratively by the educational community that uses them. They are non-standard, highly specialized vocabularies, adapted to the educational community that creates them. Thus, the vocabulary, in addition to a communication language with the repository retrieval system, is in itself an educational resource that helps students to create a conceptual map of the knowledge domain and learn its terminology. Despite the specific nature of these vocabularies, their joint use with standard metadata provides sufficient interoperability to ensure that the educational resources will be enduring from the technological point of view.

In the following section, the last aspect of vocabularies is described, namely their type. There are five types of vocabularies: lists, classifications and taxonomies, thesauri, ontologies, and dictionaries and glossaries. Each type plays a specific role in the retrieval of digital educational resources.

6. Types of Vocabularies for Retrieval of Digital Educational Resources

The types of vocabularies distinguished in this section are based on the ANSI/NISO Z39.19 [9] standard for the creation of monolingual vocabularies and on the recommendations for use of vocabularies for the description of "learning objects" made by the European Commission for Standards [10]. There are five types:

1) Simple vocabularies or lists of values;
2) Classifications and taxonomies;
3) Thesauri;
4) Ontologies;
5) Dictionaries and glossaries.

The selection of a specific type for retrieval of educational resources depends on (i) the specific role to be played; (ii) the users targeted; (iii) compatibility with the information system where it is integrated; (iv) the resources (financial, personal, and time) for their creation and maintenance; (v) the coverage; (vi) the size; (vii) the nature of the educational resources; and (viii) the type of queries, searches, and user profiles [2]. The types of vocabulary and their use experience complete the review made in this paper of the role of vocabularies in the retrieval of digital educational resources.

6.1. Simple Vocabulary or List of Values

The simplest form of vocabulary is a list of terms (Figure 19). They can be accompanied by unique alphanumerical identifier and some other types of very simple associated information, but they do not contain definitions or semantic relations with other terms. Their structure is flat: completely linear and in alphabetical order when viewed.

Simple vocabularies are suitable for providing values for certain properties (also known as attributes) of the digital resource metadata. For example, Figure 20 shows an excerpt of the metadata of an educational resource in the former ARIADNE European repository [7]. The properties Educational, Media (MIME), Type, and Operating System (OS) take their values from simple vocabularies. Figure 21 shows use of the ISO 3166 vocabulary [74] in Figure 19 to describe the property "language" in an XML document.

English short name	▲ French short name	Alpha-2 code	Alpha-3 code	Numeric
Afghanistan	Afghanistan (l')	AF	AFG	004
Åland Islands	Åland(les Îles)	AX	ALA	248
Albania	Albanie (l')	AL	ALB	008
Algeria	Algérie (l')	DZ	DZA	012
American Samoa	Samoa américaines (les)	AS	ASM	016
Andorra	Andorre (l')	AD	AND	020
Angola	Angola (l')	AO	AGO	024
Anguilla	Anguilla	AI	AIA	660
Antarctica	Antarctique (l')	AQ	ATA	010

Figure 19. A standard Vocabulary: ISO 3166 Codes for the representation of names of countries and their subdivisions—Part 1: Country codes (source: https://www.iso.org/obp/ui/#search).

Figure 20. Metadata of an educational resource in the former ARIADNE European repository.

Figure 21. Use of the ISO 3166 standard vocabulary to give values to the property "language" of the XML element <string>.

6.2. Classifications and Taxonomies

A *classification* is a vocabulary comprising categories. Categories are usually presented in alphabetical order (Figure 22) or hierarchically if they are linked by hypernymy–hyponymy or generalization–specialization relations. For example, Figure 23 shows the "Language. Linguistics. Literature" category, the "Linguistics and language" subcategory, and the subcategories within this subcategory in the Decimal Universal Classification.

Figure 22. Alphabetical classification of papers, specifying the frequency of the number of links to each term by means of their greater or smaller size (Source: http://cent.uji.es/octeto/taxonomy/).

Figure 23. Root category number 8 "Language, Linguistic, Literature" in the Decimal Universal Classification with some of the terms included (Source: http://www.udcc.org).

A *taxonomy* is a classification for a speciality, discipline, or specific thematic area [10]. It constitute specialized representation of a specific knowledge domain. The structure of a taxonomy is hierarchy. Categories or general descriptors are located at the top level. The successive hierarchy levels refine the top-level categories or descriptors (Figure 24). Descriptors can also have an alphanumerical ID. They are frequently integrated in more general vocabularies. For example, the Cyc ontology includes the Mesh medical taxonomy and the Open Directory topic taxonomy [75].

This difference between classification and taxonomy is very subtle and in practice both terms are frequently used as synonyms. For example, Figure 24 shows the ACM taxonomy for the Computer Science domain, and yet it is called as "classification system".

In any case, use of classifications and taxonomies to conceptually describe digital educational resources consists in associating each resource with the suitable category or categories. The semantic description of the resource is a set of terms, called taxonomic paths, from the terminal categories, called "leaf", which include them, to the top or "root" categories in the classification or taxonomy

(Figure 25). The indexing tools and the tools for search by concept and topic in the educational resource repositories are based, for the most part, in categories and taxonomies (Figure 26).
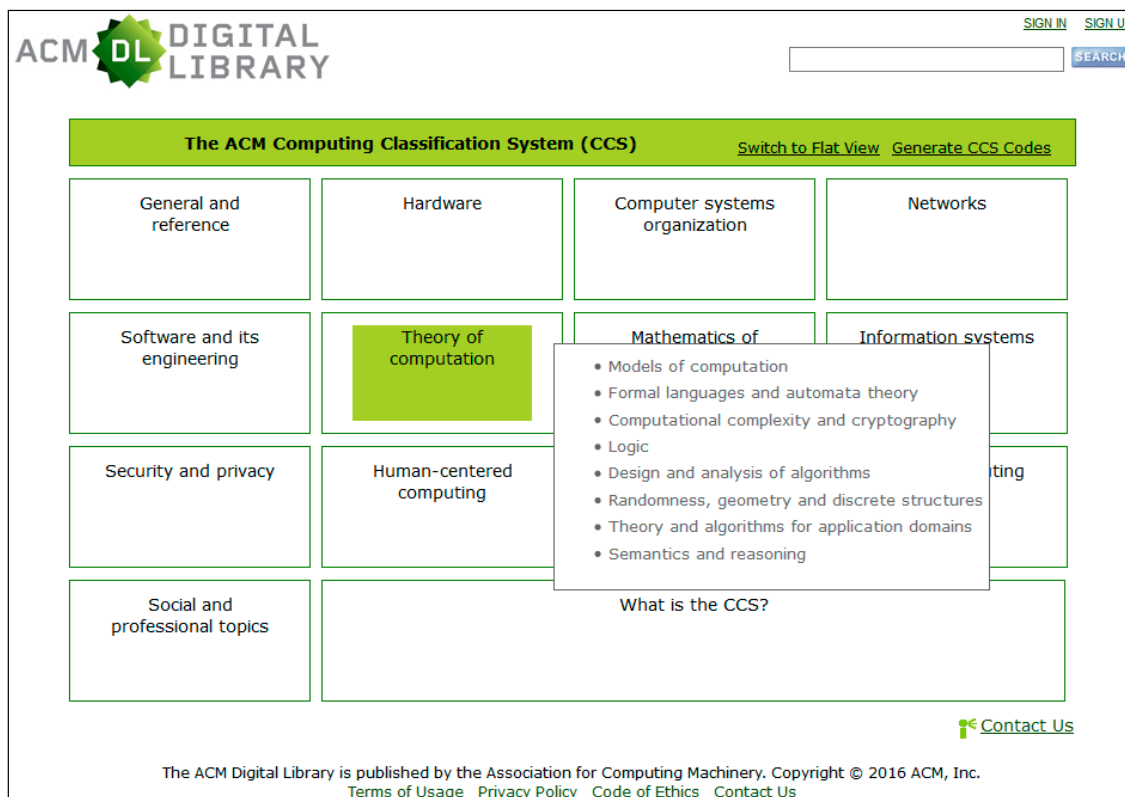

Figure 24. Categories in the Association for Computing Machinery (ACM) Classification System and terms in the "Theory of computation" category (Source: http://dl.acm.org/ccs/ccs.cfm).
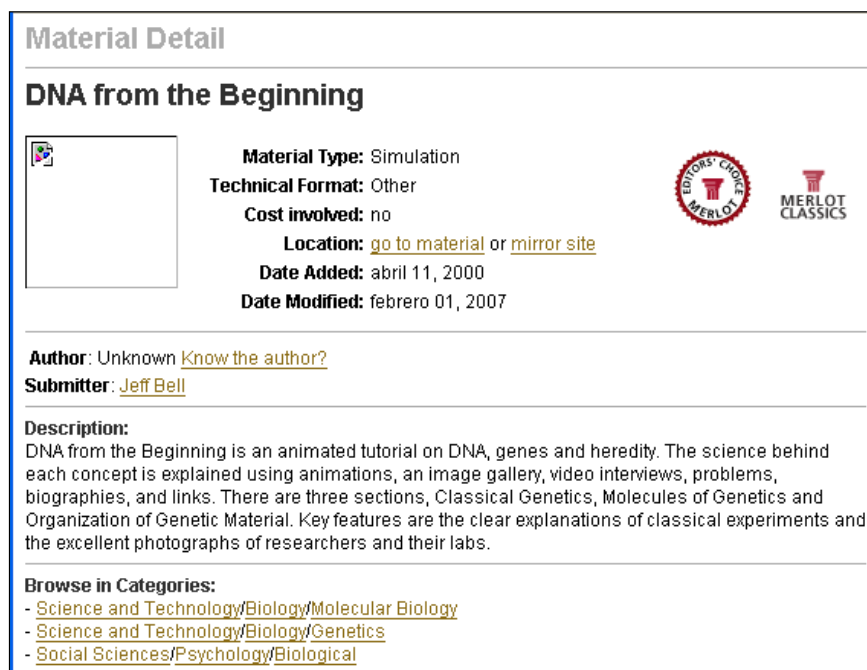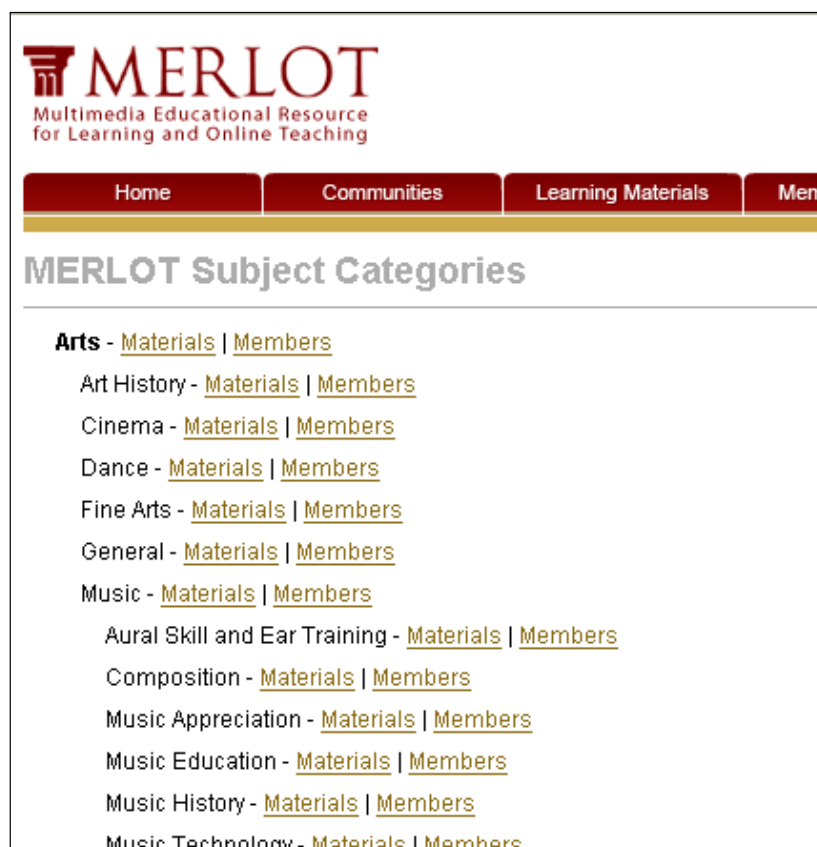

Figure 25. Classification of a resource with several taxonomic paths specified in the field: Browse in Categories (Source: [6]).

Figure 26. Categories in the MERLOT educational resource repository (Source: MERLOT [6]).

6.3. Thesauri

A *thesaurus* is a limited vocabulary, usually of specialized terms, including their semantic correspondence, selected to represent the notions that appear in a given context for use in computing and in the establishment of indices [76]. The concept of thesaurus has been adapted to new Information and Communications Technologies uses. Thus, later definitions are oriented towards IR: Aitchison defines a thesaurus as "the vocabulary of a controlled indexing language, formally organized so as to make relations between concepts explicit beforehand" [2], and the ANSI/NISO Z39.19 standard as "a controlled vocabulary that is organized and structured in a known form, where relations between equivalent terms, homographs, hierarchical and associative relations are clearly viewed by means of standard and reciprocal markers" [9].

The difference between a thesaurus and other types of vocabularies lies in the fact that its priority is formally representing semantic relations between concepts using natural language terms. Simple vocabularies do not contain semantic relations: classifications and taxonomies implicitly represent the generalization/specialization relation only. Dictionaries and glossaries, as will later be seen, contain natural language descriptions of works in which relations between words are not always explicit (Figure 27). Finally, ontologies represent concepts (not words) and their relations at a language-independent level.

Thesauri explicitly define the standard semantic relations of associativity, equivalence, and hierarchy and other potential relations pertaining to the speciality domain, for each of the terms that constitute it. They can also organize terms by aspect (i.e., separate categories) or by category (i.e., categories that are not necessarily separate), including information to specify the definition of the concept designated by the term. Figures 28–30 show three examples of thesauri used to describe digital educational resources.

Figure 27. Content and structure of the "learning" entry in *The American Heritage* e-dictionary (**top**) and *Roget's Thesaurus* (**bottom**).

(LINGUISTIC)·ACCEPTABILITY¶

T.E.:¶

     AGRAMMATICALITY¶

     GRAMMATICALITY·(LINGUISTICS)¶

T.G.:¶

     SOCIOLINGUISTICS¶

SPEECH·ACTS·(LINGUISTICS)¶

T.R.:¶

     LANGUAGE·AND·LANGUAGES-·PHILOSOPHY¶
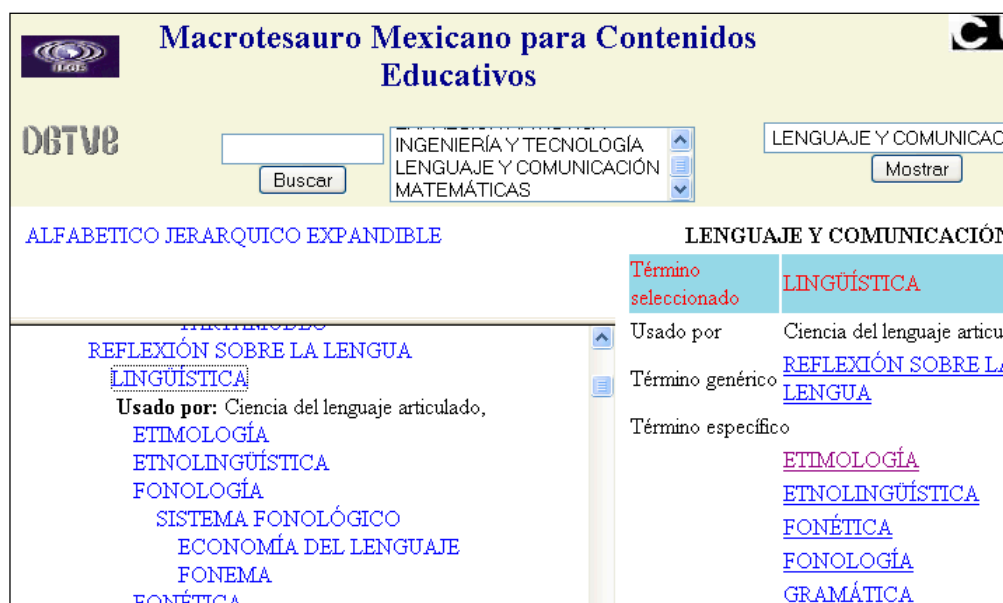
ACTION·AND·COMPETENCE·(LINGUISTICS)¶

U.S.:¶
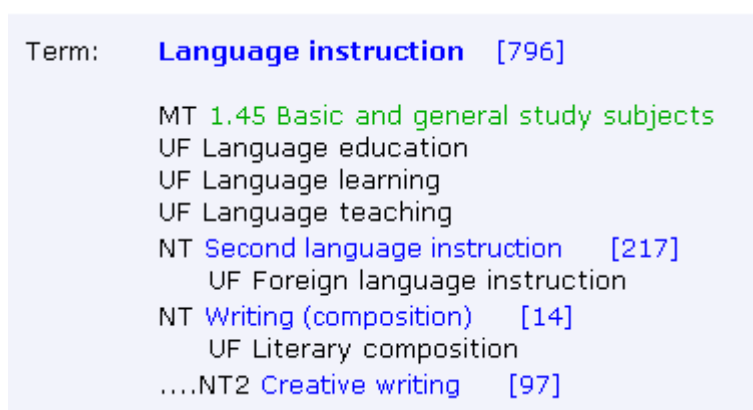
     COMPETENCE·AND·ACTION·(LINGUISTICS)¶

ANAPHORA·(LINGUISTICS)¶

Figure 28. Excerpt from the UCM Library thesaurus for documentary management. The standard alphabetical presentation is used (Source: http://alfama.sim.ucm.es/tesauro/).

Figure 29. Hierarchical presentation (left) and by detailed term (right) of the MMCE thesaurus for description of educational content (Source: http://iibi.unam.mx/~tesauro/).

Figure 30. Presentation by selected term from the UNESCO documentary resource repository thesaurus. The number of resources associated with the term is specified and linked to the right of the terms (Source: http://vocabularies.unesco.org/browser/thesaurus/en/).

One of the main applications of thesauri is currently the retrieval of digital educational resources, but it is not the only one. Other uses include support in general understanding of a knowledge area, providing conceptual maps and conceptual schemes that show the interrelations between concepts and entities (e.g., resources), the search for alternative terms in text writing or reading, the learning of the terms in a discipline, and the generation of keyword lists [2].

Regarding the retrieval of educational resources, thesauri are used for exploration and search in digital repositories. Resources are associated with one or several terms in the thesaurus (Figure 31), and during the search all the combinations of the query terms that are related in the thesaurus are calculated to specify and complete the query, thus reducing the number of unwanted results or the likelihood of obtaining no results (Figure 32) [10].

UNESDOC

3/796

Title:	Directory of ICT resources for teaching and learning of science, mathematics and language
Publ Year:	2006
Corporate author:	UNESCO Office Bangkok and Regional Bureau for Education in Asia and the Pacific; ASEAN Foundation
Imprint:	Bangkok, UNESCO Bangkok, 2006
Collation:	50 p.
Original Language:	English
Main descriptors:	educational technology; teaching materials; science education; mathematics education; language instruction; secondary education; teaching guides
Identifiers:	ICT in Education Programme; Strengthening ICT in Schools and SchoolNet Project in ASEAN Setting
Document Type:	Unesco publication
Catalog Number:	145633

Figure 31. The "Main descriptors" field contains a description of the content of the educational resource using the UNESCO thesaurus terms.

Figure 32. Search for educational resources by means of the ERIC Thesaurus (Source: http://eric.ed.gov/).

For exploration, the thesaurus can be visually presented as a conceptual map through which users can browse. While browsing, users can interactively discard those thesaurus terms that do not meet their request so as to specify the results. Results can also be expanded by finding new resources from the terms related to the search term(s). Figures 33 and 34 show two examples.

Thesauri are systems for the representation and retrieval of educational resources that are easy to use by people as they contain natural language terms or speciality terms to represent the contents of the educational resources, avoiding the ambiguity intrinsic to natural language. However, strictly speaking thesauri cannot be regarded as a conceptual representation of the knowledge domain or as a natural-language description. Rather, they constitute an intermediate system between conceptual representations and natural language.
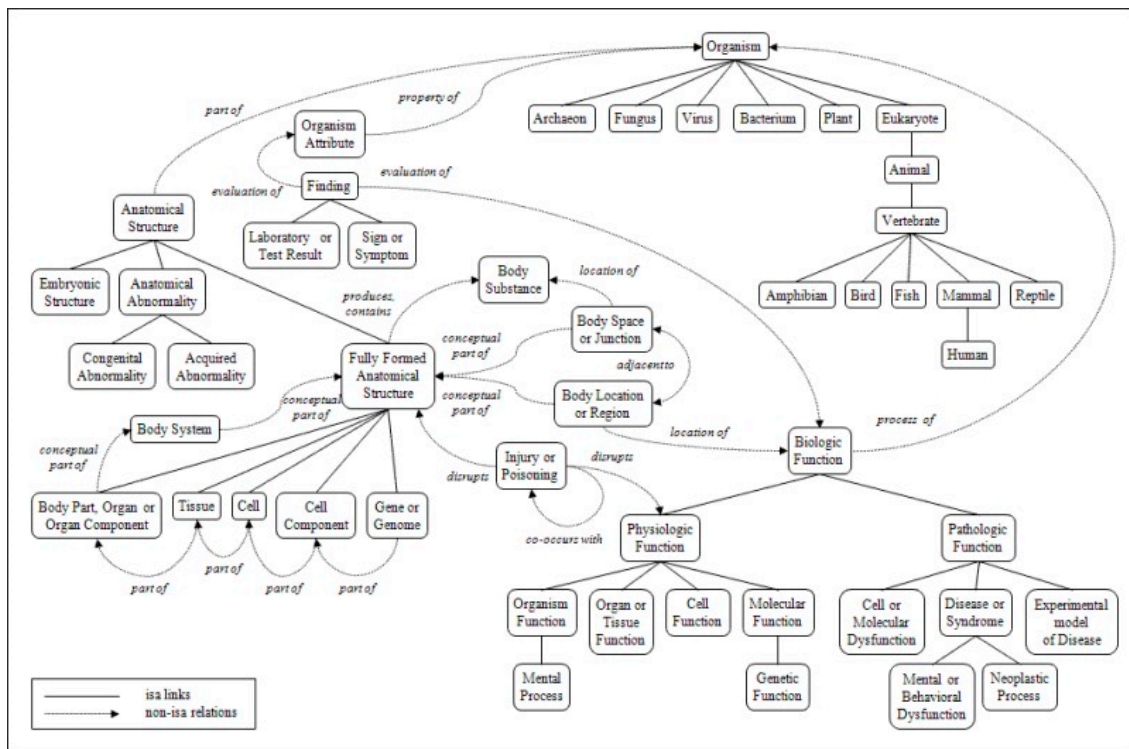
Figure 33. Graphic presentation of part of the Semantic Network of the Unified Medical Language System (UMLS) thesaurus [77] (source: https://www.ncbi.nlm.nih.gov/books/NBK9679/, Figure 3).
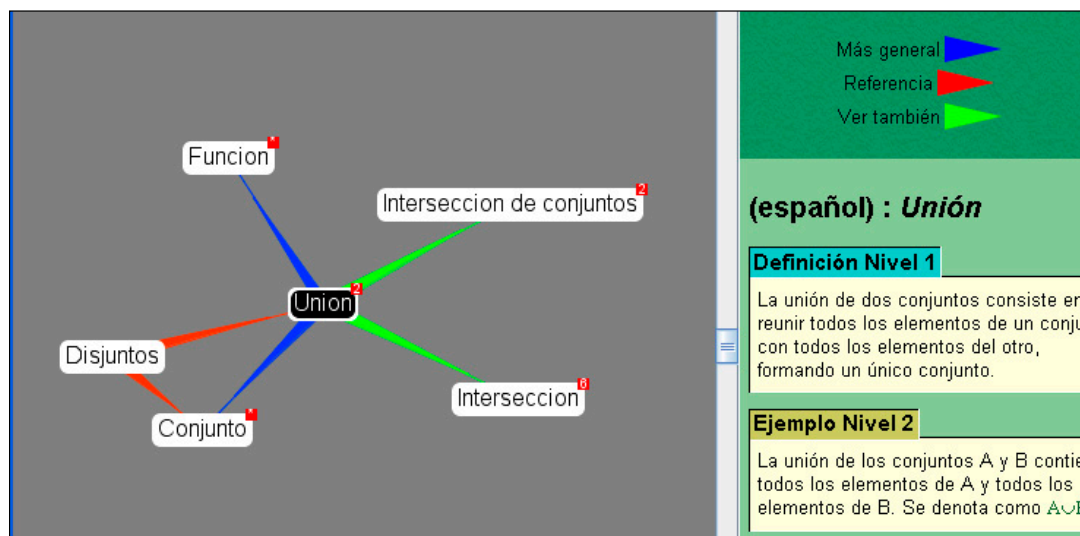
Figure 34. Graphic presentation of the thesaurus.maths.org thesaurus, with the terminological and conceptual organization of the subject "mathematics" (currently unavailable on the web).

6.4. Ontologies

The term *ontology* is taken from Philosophy and has been adapted to Computer Science and to Library and Information Science to handle the representation, management, and exchange of knowledge. From a computer point of view, ontology is defined as the specification of a conceptualization [78], or, more specifically, the *formal specification of a shared conceptualization* [79]. A conceptualization is an abstract, simplified view of the world to be represented (known as the domain model); it is shared because it is agreed, and it is formal because it is precise and non-ambiguous, which

makes it automatic processing possible (Figure 35). Ontologies provide a shared, agreed understanding of a knowledge domain which can be used by people and by computer systems [80]. In this regard, they are a models for knowledge representation, exchanging knowledge by reusing and incorporating it from different sources [35].
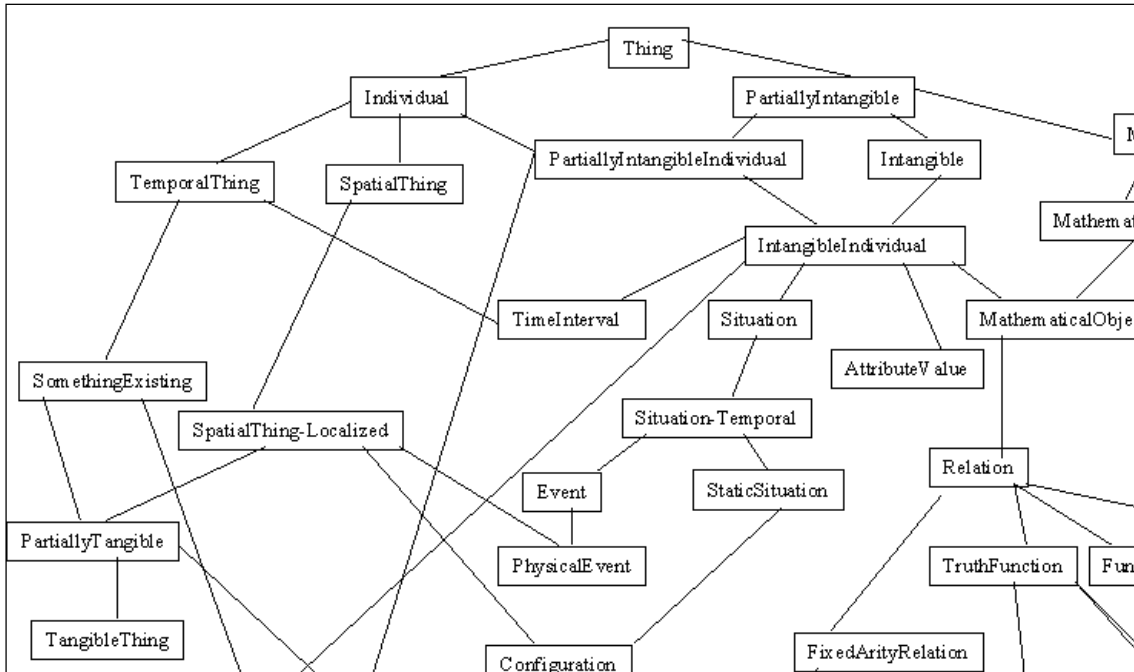
Figure 35. Top level of the Cyc ontology (Source: http://www.cyc.com).

Ontologies comprise a set of entities or concepts, relations and instances of entities and relations. They define the structure of a domain and use a shared vocabulary and semantics (Figures 35 and 36). They are usually structured by generality levels: a top level, which represents knowledge of the world in general and provides basic notions and abstract concepts (Figure 35), and specific level, corresponding to specific domains or knowledge areas.
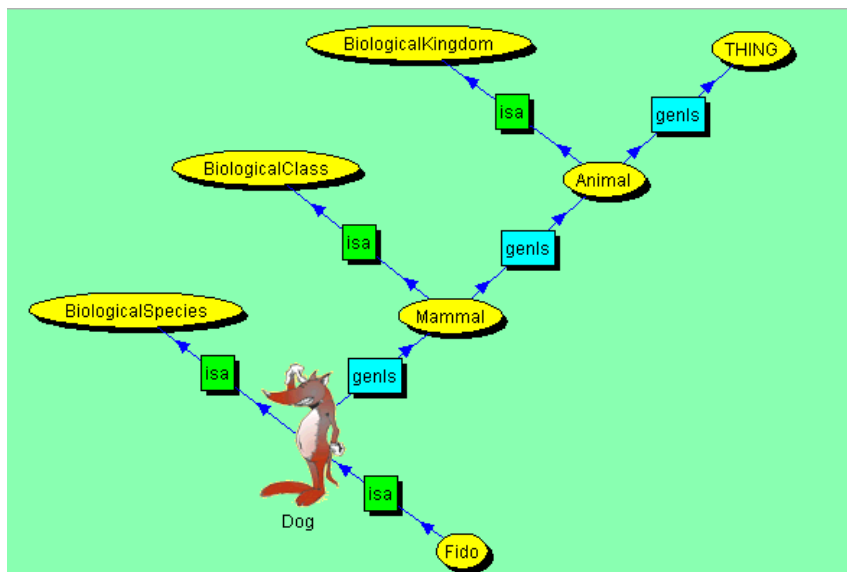
Figure 36. Example of an instance of "dog" in the Cyc ontology: "Fido".

Multiple models and types have been put forward to build ontology, depending on the purpose or the domain to be modeled [81–84]. More specifically, in [82], four types of ontology are put forward: formal, terminological, prototypes, and mixed.

In *formal ontologies*, concepts are defined by means of axioms and definitions in logic or in some programming language that can be automatically translated into logic (Figures 37 and 38). These ontologies have the greatest expressive power, as they make it possible to make queries not only regarding the knowledge explicitly stored but also regarding implicit knowledge by means of the application of inference rules and automatic reasoning. Their drawback is that they are costly to build and hard to maintain.
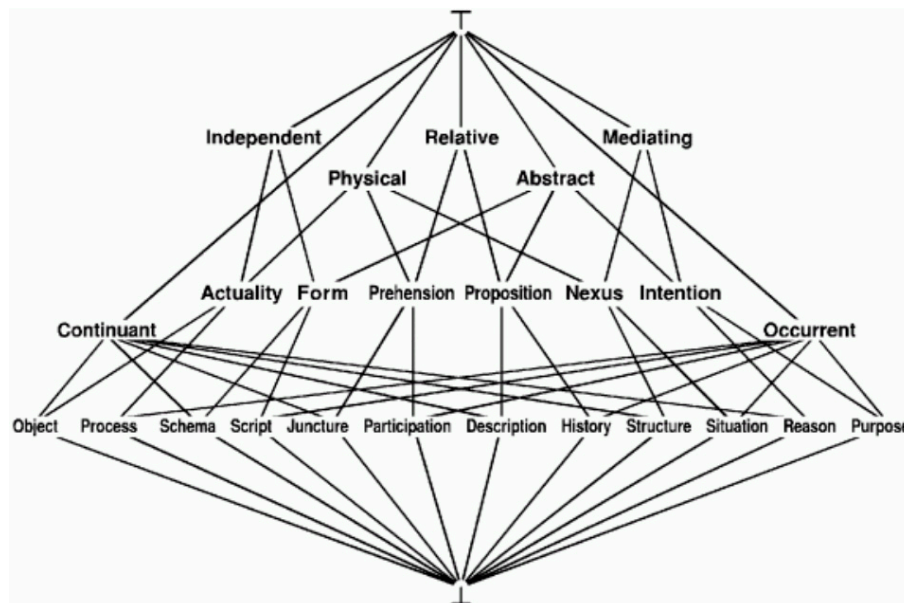
Figure 37. Categories of the top level of the KR Ontology (Source: http://www.jfsowa.com/ontology/toplevel.htm).

Abstract (A).

Pure information as distinguished from any particular encoding of the information in a physical medium. Formally, Abstract is a primitive that satisfies the following axioms:

- No abstraction has a location in space: ~(x:Abstract)(y:Place)loc(x,y).

- No abstraction occurs at a point in time: ~(x:Abstract)(t:Time)pTim(x,t).

As an example, the information you are now reading is encoded on a physical object in front of your eyes, but it is also encoded on paper, magnetic spots, and electrical currents at several other locations. Each physical encoding is said to *represent* the *same* abstract information.

Figure 38. Axioms to define the concept "Abstract".

Terminological ontologies express concepts by means of natural-language terms and semantic relations [82]. A standard example is WordNet [17], which calls concepts "Synsets", represents them by means of one or several language terms, and links them by means of semantic relations that depend on the grammatical category of each term. For example, the basis semantic relations for nouns are hyponymy–hypernymy and part–whole (Figure 39).

Prototype ontologies define concepts by means of categories that are defined by the instances that contain them. Thus, for every category (or concept) "c" of the ontology, there must exist a prototype or instance "p", and a semantic distance function "d(p,x,c)" which measures the similarity or difference between "p" and the other "x" instances in the ontologies. The new instances in the ontology are "placed" in categories that contain similar instances, that is, with small semantic differences with respect to it. The prototype theory has its philosophical roots in Wittgenstein's famous "family resemblance theory", which states that objects (or instances) covered by a term often share a family resemblance [85].

Thus, prototype categories are constructed on the basis of experientially perceived similarities among members, and these similarities may involve one or more dimensions, or characteristics.

Display Options: (Select option to change) ∨ | Change |
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
    - *direct hyponym* / *full hyponym*
    - *part meronym*
    - *domain term category*
    - *direct hypernym* / *inherited hypernym* / *sister term*
    - ***derivationally related form***
- S: (n) **car**, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad) *"three cars had jumped the rails"*
- S: (n) **car**, gondola (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)
- S: (n) **car**, elevator car (where passengers ride up and down) *"the car was on the top floor"*
- S: (n) cable car, **car** (a conveyance for passengers or freight on a cable railway) *"they took a cable car to the top of the mountain"*

Figure 39. Concepts (Synsets) of "car" and semantic relations with other concepts (source: WordNet Search 3.1).

Finally, *mixed ontologies* contain concepts defined as in formal ontologies in the top, more abstract levels, and concepts defined by means of prototypes or instances at the more specialized levels. They are a "mix" of formal and prototype ontologies.

Regardless of their type, ontologies have a complex structure, as they are networks of organized concepts with partial order relations, such as subtype–type (i.e., hyponymy–hypernymy) or part–whole (meronymy–holonymy). The resulting structure is called a lattice in mathematics, and one or several types of hierarchies can be distinguished within it. The operations and properties defined in lattices can be programmed to automatically manage the knowledge that they contain.

In the context of e-learning, ontologies are applied to describe the components of Learning Management Systems (applications to build virtual e-learning spaces), such as user types, academic organization and administration, courses, activities, and resources. These semantic descriptions for each component make it possible to create applications to personalize and adapt teaching and learning to each student's needs, facilitating collaborative learning in a distributed environment, exchange and reuse of information and knowledge, and management of and access to digital educational resource repositories [10,39,72,86,87].

Regarding the role of ontologies in management of and access to digital educational resource repositories, these are basically used as indexes to index resources with respect to their contents. One example is the ALOCoM ontology, which defines a *content model* to describe the Learning Objects in the ARIADNE European repository (ARIADNE Learning Objects Content Model) [88]. Ontologies also serve to interpret users' queries by using concept-based search strategies [89]. Finally, they are also used as conceptual maps and schemes of the domain, like thesauri, through which teachers and students can browse to locate resources that have related content, topics, or concepts (for example, [90]) (Figure 40).
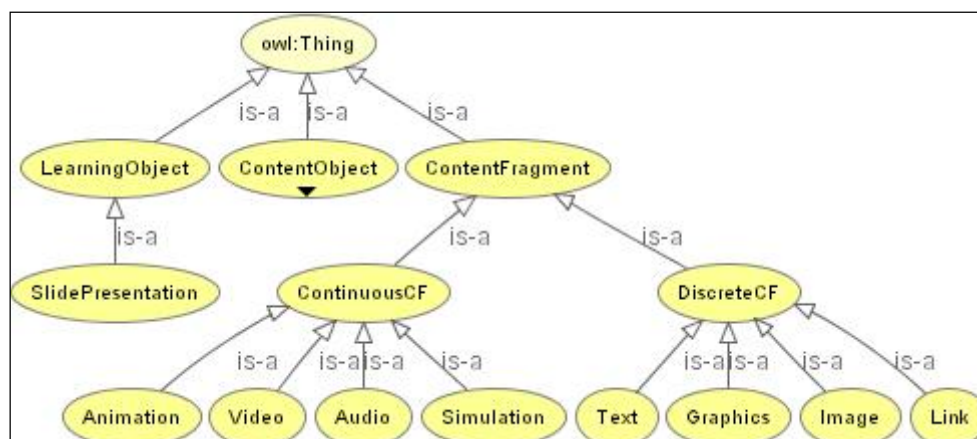
Figure 40. Concepts of the top level of the ALOCOM ontology (Source: [88]).

In practice, the distinction between ontologies and thesauri is not always clear. In the bibliography cited in this paper we find many examples of thesauri called ontologies, particularly in the case of terminological ontologies that model specific knowledge domains. For example, in the ontology classification given in [84], vocabularies—lists, glossaries, and thesauri—are regarded as a type of ontology.

However, other authors such as Sowa [26,40] clearly establish the differences between vocabularies understood as lexicons—lists, taxonomies, thesauri, dictionaries, and glossaries—and ontologies. They believe that ontologies have three clear characteristics with respect to other vocabularies: (1) they constitute the conceptual representations of the knowledge domain at an abstract semantic level; (2) they must be representations that are independent from natural languages even though their notation must be close so as to be legible (for which reason they have been used as interlanguage modules in multilingual systems [91]); and (3) they can include inferential knowledge (rules) that make it possible to reason on the basis of the knowledge explicitly stored. Sowa [92] rightly defined the role of the lexicon with respect to the role of ontologies in the knowledge representation as follows: "the lexicon is the bridge between language and the knowledge expressed with that language".

In any case, ontologies are preferred as systems for the description of computer applications because they facilitate interoperability between said applications, while thesauri are preferred as mechanisms for the indexing, search, and browsing through information and digitalized resource collections as they facilitate operability for people.

6.5. Glossaries and Dictionaries

Glossaries and dictionaries are vocabularies organized and written for human use even if they are presented in an electronic format, among other reasons because the information which they contain is not always explicit, it is semi-structured, and its interpretation depends on the user's linguistic competence and knowledge of the world [93].

Generally speaking, when information is semi-structured, the structure of the content elements (such as the lemma, the definition, and the examples) is embedded in the text and is recognized because a convention or simple syntax is used [94]. Thus, the preamble of dictionaries describes the conventions to recognize each information item, such as the convention to present the structural element "lemma" shown in Figure 41 [27].

The lemma is the term that heads the entry and can be presented in different ways:

Usually in a regular font.

In italics if the term in question is a foreign term or a Latin term not adapted to Spanish (kilobyte, ab initio).

In capital letters if it is an acronym.

Preceded or followed by a dash in the case of a compositional element (pre-, -aico, ca).

With a superindex, if there are two or more entries with the same form (champán1, champán2).

In the masculine form, followed by the feminine termination, when there is a gender variation

Figure 41. Description of the structural element "lemma" and the presentation system in the Dictionary of the Royal Spanish Academic (Source: http://dle.rae.es).

A *dictionary* is a "repertoire in the form of a book on paper or on an electronic support which brings together, in a given order, the words or idioms of one or more languages or of a specific subject, together with their definition, equivalence, or explanation" (Figure 42) [27]. As for a *glossary*, it is a "catalogue of words of one subject, field of study, work, etc., defined or discussed" [27]. The difference between them lies in their coverage, complexity, and purpose. Glossaries have a limited number of entries and more specific content than dictionaries. Their purpose is to accurately define the specific terms of a given text, work (Figure 43), or subject (Figure 44) to help readers to understand the material [95].

diccionario

Del b. lat. *dictionarium.*

1. m. Repertorio en forma de libro o en soporte electrónico en el que se recogen, según un orden determinado, las palabras o expresiones de una o más lenguas, o de una materia concreta, acompañadas de su definición, equivalencia o explicación.

2. m. Catálogo de noticias o datos de un mismo género, ordenado alfabéticamente. *Diccionario bibliográfico, biográfico, geográfico.*

Real Academia Española © Todos los derechos reservados

Figure 42. Query for the word "dictionary" in the online version of DRAE [27].

Glossary

Canadian Index of Well-being: An initiative to develop a new comprehensive measure of well-being that reflects economic, social and environmental conditions that matter most to Canadians.

Capital: Wealth, in whatever form, used or capable of being used to produce more wealth. Also an accumulated stock of such wealth (money or property) or its value which is owned or used in business by a person, corporation, etc.

Chrematistics: An ancient Greek term which literally means the art of money-making or the science of wealth (e.g. riches).

CIW: Canadian Index of Well-being.

Common Wealth: This phrase (or "the common weal") incorporates the old meaning of wealth which is well-being. Thus the term literally means "common well-being."

Competition: Originally from the Latin *competere* meaning to strive together for some common interest or purpose.

The Economics of Happiness: Building Genuine Wealth

Escrito por Mark Anielski

Buscar en este libro   Ir

Figure 43. Glossary for a work (Source: Google Books).

SOFTWARE ENGINEERING GLOSSARY

Software configuration management domain

baseline: 1. A description of a system and its components (configuration items) at a particular period of time, and any approved updates to the baseline.
2. A work product that has been placed under formal configuration management. See also *work product*.
baseline document: A system/software document that defines a work product that has been placed under configuration management. Examples are system specifications, requirements specifications, and design specifications.
build: An operational version of a software product incorporating a specified subset of the capabilities that the final product will include. Sometimes synonymous with *version*.
configuration: 1. The arrangement of a system or network as defined by the nature, number, and chief

port change processing and implementation status, and verify compliance with specified requirements. [IEEE Std. 610.12-1990]
configuration management system: The discipline of identifying the components of a continually evolving system for the purpose of controlling changes to those components and maintaining integrity and traceability throughout the life cycle.
control point (project control point): An agreed-on point in time or times when specified project agreements or controls are applied to the software configuration items being developed. Examples are an approved baseline or release of a specified document or code. [IEEE Std. 828-1998]
document control: The application of configuration management to the control of documents.
engineering change: An alteration in the configuration of a hardware/software configuration item or items, delivered, to be delivered, or under development, after formal establishment of their configuration identification. See also *engineering change proposal*.
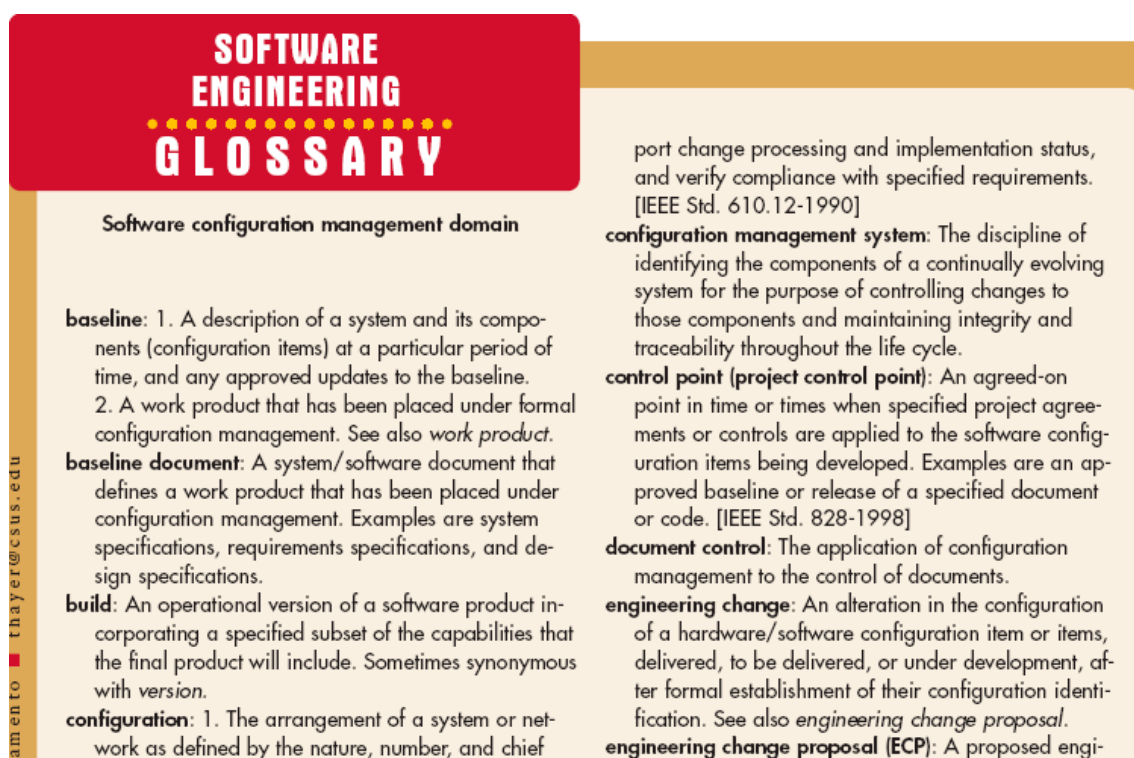**engineering change proposal (ECP)**: A proposed engi-

Figure 44. Glossary for "Software Engineering" (source [96]).

Use of glossaries and dictionaries in the *retrieval of digital resources* is very limited. In the case of dictionaries, they are used as a source of lexical knowledge to create other types of vocabularies that are suitable for retrieval (classifications and taxonomies, thesauri, ontologies, and lexical databases) [97]. Glossaries, by contrast, have been directly used as a tool to access digital resources, both to expand search mechanisms [98] and to organize and explore digital information, content, and resources [99].

Glossaries are used in the retrieval of educational digital resources, not only to delimit the set of natural-language terms used to index and describe the stored resources, as in the case of lists or classifications and taxonomies, but also to provide the definition of the sense in which said terms are used. In this way, users can compare whether what is sought for corresponds to what was indexed and stored in the system, which also helps them to understand and grasp the concepts in the knowledge domain that encompasses the educational resource repository.

7. Summary and Conclusions

To understand the role of vocabularies in the retrieval of digitalized educational resources, we have reviewed (i) their nature, which depends on the vocabulary type; (ii) their role in IR systems; and in particular (iii) their role in the retrieval of educational resources, taking the vocabulary type also into account. The conclusions of this review are the following:

Firstly, in the context of retrieval of educational resources in electronic environments, the vocabulary constitutes a linguistic, computer, and educational resource that brings together and formalizes the lexical knowledge of a knowledge domain in order to describe and index the resources of said domain so that they can be easily located by people or computer systems.

Secondly, regarding the nature of the vocabularies for retrieval, a vocabulary usually comprises a set of terms that can be associated and described. When terms have a single meaning, the vocabulary is regarded as being controlled and suitable for IR, and in particular for the retrieval of educational resources. Associations between terms and descriptions depend on the vocabulary type. Thus, (i) lists do not contain associations or descriptions; (ii) taxonomies, thesauri,

and ontologies only contain semantic associations between terms, basically hyponymy–hypernymy in classifications and taxonomies, synonymy, hyponymy–hypernymy and related terms in thesauri, and hyponymy–hypernymy and varied typology in ontologies; (iii) glossaries and dictionaries contain natural-language descriptions that implicitly and explicitly contain relations with other words. From the point of view of the retrieval of digital educational resources, the most effective systems are classifications, taxonomies, and thesauri.

Thirdly, regarding the role of vocabularies in IR systems, they provide a precise language to describe and guide the search for the information that is the object of a query. They are used in indexing and queries. Vocabularies are IR-effective if they are able to represent the resources in indexing in the same way as users make searches in their queries. For this reason, a vocabulary that is effective in an IR context is not always effective in a different IR context within the same knowledge domain. Consider, for example, the educational context of a university, where teachers, researchers, and students are usually the authors, indexers, and users of the information generated. In this context, vocabularies of reference such as the Universal Decimal Classification, which is widely used for IR in libraries and documentation centers, seem not to be effective in educational repositories, and, if possible, new vocabularies closer to the speciality language shared by the creators and users of the knowledge domain are created.

Fourthly and finally, regarding the role of vocabularies in the retrieval of digital educational resources, they provide a precise, shared language to index, describe, and locate the resources. They are usually employed with metadata as they benefit from the features of metadata, which are simpler to process and, if standard, provide a minimum level of interoperability, as well as from the benefits of vocabularies, which provide a language for metadata values that is suited to users' knowledge and needs (teachers and students).

To conclude, experience in use of vocabularies for the retrieval of the digitalized educational resources reviewed indicates that vocabularies constitute the language for the retrieval of educational resources and are usually employed to add value to metadata properties. Vocabularies could be more effective if: (i) they are adapted to the knowledge domain of the collection resources and the teaching and speciality language of the teachers who are the main users and in many cases the authors of the resources; (ii) they provide mechanisms that facilitate the collaborative and inductive creation of these vocabularies in a framework that is as standardized as possible; and (iii) they can be used as an educational resource in themselves to learn the speciality language for the subject(s) which they encompass.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bougarev, B.K. Building a Lexicon: The Contribution of Computers. *Int. J. Lexicogr.* **1991**, *4*, 227–260. [CrossRef]
2. Aitchison, J.; Gilchrist, A.; Bawden, D. *Thesaurus Construction and Use: A Practical Manual*, 4th ed.; Aslib IMI: London, UK, 2000.
3. Rodríguez, K.; Ronda, R. Web Semántica: Un Nuevo Enfoque Para la Organización y Recuperación de Información. *ACIMED* **2005**, *13*. Available online: http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm (accessed on 17 February 2017).
4. Huynh, D.; Mazzocchi, S.; Karger, D. Piggi Bank: Experience the Semantic Web inside Your Web Browser. In Proceedings of the 6th International Semantic Web Conference, 2nd Assian Semantic Web Conference, Busan, Korea, 11–15 November 2007; Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Springer: Berlin, Germany, 2007. Available online: http://www.websemanticsjournal.org/index.php/ps/article/download/107/105 (accessed on 30 October 2016).

5. Sotiriou, S.; Manouselis, N.; Tzikopoulos, A.; Costopoulou, C.; Sideridis, A.B. Towards teacher competence on metadata and online resources: The case of agricultural learning resources. In METASCHOOL Challenges in Training Teacher about Metadata and Online Resources, Proceedings of the European Workshop, Alcalá de Henares, Spain, 20–22 October 2010; European Commission, Project No 141942-LLP-1-2008-1-GR-COMENIUS-CMP, Life Long Learning, Tranversal Programmes; Ellinogermaniki, A., Sotiriou, S., Tzikopoulos, A., Eds. Available online: http://www.ea.gr/ep/Metaschool/deliverables/Metaschool_WP5_Dissemination_V1.0_30Oct2010_EA.pdf#page=9 (accessed on 30 October 2016).

6. MERLOT. Multimedia Educational Resources for Learning and Online Teaching. Available online: https://www.merlot.org/ (accessed on 30 October 2016).

7. Ariadne. Available online: http://www.ariadne-eu.org (accessed on 30 October 2016).

8. Lancaster, F.W. *Vocabulary Control for Information Retrieval*, 2nd ed.; Information Resource Press: Arlington, VA, USA, 1992.

9. ANSI/NISO Z39.19. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*; NISO Press: Bethesda, MD, USA, 2010. Available online: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf (accessed on 9 February 2017).

10. European Committee for Standarization. Controlled Vocabularies for Learning Objects Metadata: Typology, Impact Analysis, Guidelines and a Web Based Vocabulary Registry, CEN CWA 14871, 2003. Available online: http://www.cen-isss-wslt.din.de/sixcms_upload/media/3050/CWA14871.pdf (accessed on 30 October 2016).

11. Europeana. Available online: http://www.europeana.eu/portal/es (accessed on 30 October 2016).

12. Sierra, J.L.; Fernández-Valmayor, A.; Guinea Bueno, M.; Hernanz, H.; Navarro, A. Building Repositories of Learning Objects in Specialized Domains: The Chasqui Approach. In Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies, Kaohsiung, Taiwan, 5–8 July 2005; IEEE Computer Society: Washington, DC, USA, 2005. Available online: https://www.computer.org/csdl/proceedings/icalt/2005/2338/00/23380225-abs.html (accessed on 30 October 2016).

13. Lyons, J. *Semantic*; Cambridge University Press: Cambridge, UK, 1977.

14. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]

15. Baeza-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval*; Addison-Wesley: Wokingham, UK, 1999.

16. Aitchison, J.; Clarke, S.D. The thesaurus: A historical viewpoint, with a look to the future. *Cat. Classif. Q.* **2004**, *37*, 5–21. [CrossRef]

17. WordNet. A Lexical Database for English. Available online: https://wordnet.princeton.edu/ (accessed on 30 October 2016).

18. MEDLINE PubMed Resources Guide. Available online: https://www.nlm.nih.gov/bsd/pmresources.html (accessed on 30 October 2016).

19. Medical Subject Headings MeSH. Available online: https://www.nlm.nih.gov/mesh/ (accessed on 30 October 2016).

20. DDC: Dewey Decimal Classification. Available online: https://www.oclc.org/dewey (accessed on 30 October 2016).

21. LCSH: Library of Congress Subject Headings. Available online: http://id.loc.gov/authorities/subjects.html (accessed on 30 October 2016).

22. NASA Thesaurus. Available online: https://www.sti.nasa.gov/sti-tools/#.WBe6wCTGqBo (accessed on 30 October 2016).

23. ACM Computing Classification System. Available online: http://www.acm.org/about/class (accessed on 30 October 2016).

24. Cyc Knowledge Base. Available online: http://www.cyc.com/ (accessed on 30 October 2016).

25. UNESCO Thesaurus. Available online: http://vocabularies.unesco.org/browser/thesaurus/en/ (accessed on 30 October 2016).

26. Sowa, J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*; Course Technology: Boston, MA, USA, 2000; p. 594. Available online: http://www.jfsowa.com/krbook/ (accessed on 30 October 2016).

27. Diccionario de la Lengua Española. Real Academia Española. Available online: http://www.rae.es/ (accessed on 30 October 2016).

28. Arnold, D.J.; Balkan, L.; Meijer, S.; Lee Humphreys, R.; Sadler, L. *Machine Translation: An Introductory Guide*; Blackwells-NCC: London, UK, 1994; p. 101. Available online: http://www.essex.ac.uk/linguistics/external/clmt/mtbook/PostScript/ (accessed on 4 January 2017).

29. ISO-1087:2000 Terminology Work—Vocabulary—Part 1: Theory and Application. Available online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=20057 (accessed on 5 January 2017).

30. Van Slype, G. *Los Lenguajes de Indización: Concepto, Construcción y Utilización en Los Sistemas Documentales*; Hipola, P.; Félix de, M., Translators; Fundación Germán Sánchez Ruipérez: Salmanca, Spain, 1991.

31. Lewis, D.D.; Jones, S.K. Natural Language Processing for Information Retrieval. *Commun. ACM* **1996**, *39*, 92–101. [CrossRef]

32. Allen, J. *Natural Language Understanding*, 2nd ed.; Benjamin-Cummings Publishing: Redwood City, CA, USA, 1995.

33. Brachman, R.J.; Levesque, H.J. *Readings in Knowledge Representation*; Morgan Kauffmann: Los Altos, CA, USA, 1985.

34. Bertino, E.; Catania, B.; Zarri, G.P. *Intelligent Database System*; Addison Wesley Longman Publishing: Boston, MA, USA, 2001.

35. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities. Scientific American. 2001. Available online: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (accessed on 30 October 2016).

36. ELRA. Available online: http://www.elra.info/en/ (accessed on 30 October 2016).

37. LDC. Linguistic Data Consortium. Available online: https://www.ldc.upenn.edu/ (accessed on 30 October 2016).

38. MS VDX: Vocabulary Definition Exchange Version 1 Final Specification. IMS Global Learning Consortium, 2004. Available online: http://www.imsglobal.org/vdex/index.html (accessed on 30 October 2016).

39. Sampson, D.G.; Lytras, M.D.; Wagner, G.; Díaz, P. Ontologies and the Semantic Web for E-learning. *Educ. Technol. Soc.* **2004**, *7*, 26–28.

40. Hirst, G. Ontology and the Lexicon. In *Handbook on Ontologies*; Staab, S., Studer, R., Eds.; Springer: Berlin, Germany, 2004; pp. 209–230.

41. Gibbon, D. Computational Lexicography. In *Lexicon Development for Speech and Language Processing*; Eynde, F.V., Gibbon, D., Eds.; ELSNET, Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000; pp. 1–42.

42. International Organization for Standardization. ISO 10241-1:2011 Terminological Entries in Standards—Part 1: General Requirements and Examples of Presentation. Available online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=40362 (accessed on 5 January 2017).

43. Lehmann, A.; Martin-Berthet, F. *Introduction à la Lexicologie*; Dunod: Paris, France, 1998.

44. Bennett, P. Introduction to Text Categorization. Web-Based Information Architectures. 23 July 2002. Available online: http://users.softlab.ntua.gr/facilities/public/AD/Text Categorization/Introduction to Text Categorization.ppt (accessed on 9 February 2017).

45. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: New York, NY, USA, 2008.

46. Salton, G.; McGill, M.J. *Introduction to Modern Information Retrieval*; McGraw-Hill Book Company: New York, NY, USA, 1986.

47. Centelles, M. Taxonomías Para la Categorización y la Organización de la Información en los Sitios Web. *Hypertext.net* **2005**, *3*. Available online: https://www.upf.edu/hipertextnet/numero-3/taxonomias.html (accessed on 30 October 2016).

48. Montejo, A. Proyecto de Indexado Automático para Documentos en el Campo de la Física de Altas Energías. *Procesamiento del Lenguaje Natural* **2001**, *27*. Available online: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/articl/view/3373/1861 (accessed on 30 October 2016).

49. Gómez-Hidalgo, J.M.; Cortizo-Pérez, J.C.; Puertas-Sanz, E.; Buenaga-Rodríguez, M. Experimentos en Indexación Conceptual Para la Categorización de Texto. In Proceedings of the Conferencia Iberoamericana WWW/Internet, Madrid, Spain, 7–8 October 2004; Gutiérrez, J.M., Martínez, J.J., Isaías, P., Eds.; 2004; pp. 251–258.

50. Mao, W.; Chu, W.W. The phrase-based vector space model for automatic retrieval of free-text medical documents. *Data Knowl. Eng.* **2007**, *61*, 76–92. [CrossRef]
51. Software. American Society for indexing. 2017. Available online: http://www.asindexing.org/reference-shelf/software/ (accessed on 9 February 2017).
52. Understanding Verity Search Indexes. Available online: https://docs.oracle.com/cd/E41633_01/pt853pbh1/eng/pt/tsvt/concept_UnderstandingVeritySearchIndexes-07101e.html (accessed on 30 October 2016).
53. Nilsson, M.; Baker, T.; Johnston, P. Interoperability Levels for Dublin Core Metadata. Dublin Core Metadata Initiative. 2009. Available online: http://dublincore.org/documents/2009/05/01/interoperability-levels/ (accessed on 30 October 2016).
54. IMS Digital Repositories. Digital Repositories Specification v. 1.0 Final Specification. Available online: http://www.imsglobal.org/digitalrepositories/ (accessed on 30 October 2016).
55. ISO 24613:2008 Language Resource Management. Lexical Markup Framework (LMF). Available online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=37327 (accessed on 30 October 2016).
56. LEMON. The Lexicon Model for Ontologies. Available online: http://lemon-model.net/ (accessed on 30 October 2016).
57. Powell, A.; Nillson, M.; Naeve, A.; Jonhston, P.; Baker, T. DCMI Abstract Model. Dublin Core Metadata Initiative. 2005. Available online: http://dublincore.org/documents/abstract-model/ (accessed on 30 October 2016).
58. Dalmau, M.; Floyd, R.; Jiao, D.; Riley, J. Integrating thesaurus relationships into search and browse in an online photograph collection. *Libr. Hi Tech.* **2005**, *23*, 425–452. [CrossRef]
59. Koutsomitropoulos, D.A.; Solomou, G.D.; Papatheodorou, T.S.P.; Alexopoulos, A.D. The Use of Metadata for Educational Resources in Digital Repositories. Practices and Perspectives. *D-Lib Mag.* **2010**, *16*, 1–2.
60. Dublin Core Specifications. Available online: http://dublincore.org/specifications/ (accessed on 30 October 2016).
61. IEEE 1484.12.1:2002. IEEE Standard for Learning Object Metadata. Available online: https://standards.ieee.org/findstds/standard/1484.12.1-2002.html (accessed on 30 October 2016).
62. Slavic, A. A Definition of Thesauri and Classification as Indexing Tools. 2000. Available online: http://dublincore.org/documents/2000/11/28/thesauri-definition/ (accessed on 30 October 2016).
63. Woods, W.A.; Bookman, L.A.; Houston, A.; Kuhns, R.J.; Martin, P.; Green, S. Linguistic Knowledge Can Improve Information Retrieval. In Proceedings of the ANLP-2000, Seattle, WA, USA, 29 April–4 May 2000. Available online: http://ucrel.lancs.ac.uk/acl/A/A00/A00-1036.pdf (accessed on 30 October 2016).
64. Aguirre, S.; Quemada, J.; Salvachua, J. Mediadores e interoperabilidad en Elearning. In Proceedings of the V Encuentro Internacional Sobre Educación, Capacitación Profesional y Tecnologías de la Información, Barcelona, Spain, 16–18 June 2004. Available online: http://jungla.dit.upm.es/saguirre/publications/virtualEduca2004.pdf (accessed on 30 October 2016).
65. Duval, E.; Forte, E.; Cardinaels, K.; Verhoeven, B.; Van Durm, R.; Hendrikx, K.; Forte, M.W.; Ebel, N.; Macowicz, M.; Warkentyne, K.; et al. The Ariadne knowledge pool system. *Commun. ACM* **2001**, *44*, 72–78. [CrossRef]
66. Friesen, N. Final Raport on the "International LOM Survey". SC 36_WG4_N0109 ISO/IEC JTC1/SC36/WG4N0109, Canada. 2004. Available online: http://mdlet.jtc1sc36.org/doc/SC36_WG4_N0109.pdf (accessed on 30 October 2016).
67. Heath, B.; McArthur, D.J.; McClelland, M.K.; Vetter, R.J. Metadata Lessons from the iLuminia Digital Library. *Commun. ACM* **2005**, *48*, 68–74. [CrossRef]
68. Hepp, M. Possible Ontologies. How Reality Constrains the Development of Relevant Ontologies. *IEEE Internet Comput.* **2007**, *11*, 90–96. [CrossRef]
69. Lee, W.; Sugimoto, S. Towards Core Subject Vocabularies for Community-Oriented Subject Gateways. In Vocabularies in Practice, Proceedings of the International Conference on Dublin Core and Metadata Applications, Universidad Carlos III, Madrid, Spain, 13 September 2005; Méndez, E., Ed. Available online: http://dcpapers.dublincore.org/pubs/article/view/799/795 (accessed on 30 October 2016).
70. Greenberg, J.; Heidorn, B.; Seiberling, S.; Weakley, A.S. Growing Vocabularies for Plant Identification and Scientific Learning. In Vocabularies in Practice, Proceedings of the International Conference on Dublin Core and Metadata Applications, Universidad Carlos III, Madrid, Spain, 13 September 2005; Méndez, E., Ed. Available online: http://dcpapers.dublincore.org/index.php/pubs/article/viewArticle/808 (accessed on 30 October 2016).

71. Sierra, J.L.; Fernández-Valmayor, A. A Heritage Dissemination Approach for the Production and Maintenance of Repositories of Learning Objects. In Proceedings of the VIII Simposio Internacional de Informática Educativa, Salamanca, Spain, 13–16 September 2006. Available online: http://www.e-ucm.es/drafts/e-UCM_draft_57.pdf (accessed on 30 October 2016).

72. Dekkers, M.; Feria, L. (Eds.) Metada for Knowledge and Learning: DC-2006. In Proceedings of the International Conference on Dublin Core and Metada Applications, University of Colima, Department of Information Technologies, Colima, Mexico, 3–6 October 2006. Available online: http://dcpapers.dublincore.org/pubs/issue/view/29 (accessed on 30 October 2016).

73. Or-Bach, R. Educational benefits of metadata creation by students. *ACM SIGCSE Bull.* **2005**, *37*, 93–97. [CrossRef]

74. ISO 3166 Country Codes. Available online: http://www.iso.org/iso/home/standards/country_codes.htm (accessed on 30 October 2016).

75. Reed, S.; Lenat, D. Mapping Ontologies into Cyc. In Proceedings of the AAAI 2002 Conference Workshop on Ontologies for the Semantic Web, Edmonton, AB, Canada, 28–29 July 2002. Available online: https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-010.pdf (accessed on 30 October 2016).

76. Martínez de Sousa, J. *Diccionario de Lexicografía Práctica*; Bibliograf S.A.: Barcelona, Spain, 1995.

77. *Release AA2009: UMLS Knowledge Source—Documentation*; U.S. National Library of Medicine: Bethesda MD, USA, 2009. Available online: http://www.nlm.nih.gov/research/umls/umlsdoc.html (accessed on 30 October 2016).

78. Gruber, T.R. A translation approach to portable ontology specification. *Knowl. Adquis.* **1993**, *5*, 199–220. [CrossRef]

79. Borst, W.N.; Akkermans, J.M.; Top, J.L. Engineering Ontologies. *Int. J. Hum.-Comput. Stud.* **1997**, *46*, 365–406. [CrossRef]

80. Lamarca, M.J. *Hipertexto, el Nuevo Concepto de Documento en la Cultura de la Imagen*; Facultad de Ciencias de la Información, Dpto. de Biblioteconomía y Documentación, Universidad Complutense de Madrid: Madrid, Spain, 2007. Available online: http://www.hipertexto.info (accessed on 30 October 2016).

81. Van Heijst, G.V.; Schreiber, A.T.; Wielinga, B.J. Using explicit ontologies in KBS development. *Int. J. Hum.-Comput. Stud.* **1997**, *46*, 183–292. [CrossRef]

82. Fensel, D.; van Harmelen, F.; Horrocks, I.; McGuinness, D.L.; Patel-Schneider, P.F. OIL: An Ontology Infrastructure for the Semantic Web. *IEEE Intell. Syst.* **2001**, *16*, 38–45. [CrossRef]

83. McGuinness, D.L. Ontologies Come of Age. In *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; MIT Press: London, UK, 2003. Available online: http://www.math.pku.edu.cn/teachers/linzq/teaching/stm/references/Ontologiescomeofage.pdf (accessed on 30 October 2016).

84. Margolis, E.; Laurence, S. Concepts. In *The Stanford Encyclopedia of Philosophy*, 2014 ed.; Zalta, E.N., Ed.; Available online: https://plato.stanford.edu/archives/spr2014/entries/concepts/ (accessed on 4 January 2017).

85. Staab, S.; Studer, R. (Eds.) *Handbook on Ontologies*; Springer: Berlin/Heidelberg, Germany, 2004.

86. Méndez, E. (Ed.) Vocabularies in Practice. In Proceedings of the International Conference on Dunlin Core and Metadata Applications, Universidad Carlos III, Madrid, Spain, 12–15 September 2005. Available online: http://dcpapers.dublincore.org/pubs/issue/view/28 (accessed on 30 October 2016).

87. Knight, C.; Gašević, D.; Richards, G. Ontologies to integrate learning design and learning content. *J. Interact. Media Educ.* **2005**. [CrossRef]

88. Hernández, H.; Saiz, M. Ontologías mixtas para la representación conceptual de objetos de aprendizaje. *Procesamiento del Lenguaje Natural* **2007**, *38*. Available online: http://www.sepln.org/revistaSEPLN/revista/38/11.pdf (accessed on 30 October 2016).

89. Jovanovic, J.; Gašević, D.; Verbert, K.; Duval, E. ALOCOM Ontology. Ontology of Learning Object Content Structure. In Proceedings of the 12th International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands, 18–22 July 2005.

90. Sowa, J.F. Ontology, Metadata, and Semiotics. In *Conceptual Structures: Logical, Linguistic, and Computational Issues*; Lecture Notes in AI #1867; Ganter, B., Mineau, G.W., Eds.; Springer: Berlin, Germany, 2000; pp. 55–81. Available online: http://users.bestweb.net/~sowa/peirce/ontometa.htm (accessed on 30 October 2016).

91. Carbonell, J.; Mitamura, T.; Nyberg, E.H. The Kant Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics). In Proceedings of the 4th Conference on Theoretical and Methodological Issues in Machine Translation, Montreal, QC, Canada; pp. 225–235. Available online: http://www.cs.cmu.edu/~jgc/publication/The_KANT_Prespective_A_Critique_ICTMI_1992.pdf (accessed on 30 October 2016).

92. Sowa, J.F. Concepts in the Lexicon: Introduction. Available online: http://www.jfsowa.com/ontology/lexicon.htm (accessed on 30 October 2016).

93. Fernández-Pampillón, A.M.; Matesanz del Barrio, M. Los diccionarios electrónicos: Hacia un nuevo concepto de diccionario. In *Nuevos Géneros Discursivos: Los Textos Electrónicos*; López-Alonso, C., Séré, A., Eds.; Estudios de Lingüística del Español (ELiEs): Barcelona, Spain, 2006; Volume 24. Available online: http://elies.rediris.es/elies24 (accessed on 30 October 2016).

94. Abiteboul, S.; Buneman, P.; Suciu, D. *Data on the Web. From Relations to Semistructured Data and XML*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 2000.

95. Blustein, J.; Noor, M. Personal glossaries on the WWW: An explanatory study. In Proceedings of the 2004 ACM Symposium on Document Engineering, New York, NY, USA, 28–30 October 2004; Vion-Dury, J., Ed.; ACM Press: New York, NY, USA, 2004; pp. 54–56.

96. Thayer, R.H. Software Engineering Glossary. *IEEE Softw.* **2003**, *20*, 9–397. [CrossRef]

97. Zernik, U. (Ed.) *Lexical Acquisition. Exploiting On-Line Resources to Build a Lexicon*; Lawrence Erlbaum Associates, Inc. Publishers: Mahwah, NJ, USA, 1991.

98. Cuesta, P.; Maña, M.J.; Cuevo, C. Aplicación de Técnicas de Recuperación de Información a un Glosario de Términos de Internet Desarrollado Utilizando Tecnologías JSP. In Actas de JBIDI, Primeras Jornadas de Bibliotecas Digitales. Valladolid, Spain, 6–7 November 2000; p. 129. Available online: http://infonautica.net/docs/jbidi/jbidi2000/12_2000.pdf (accessed on 30 October 2016).

99. Tabata, K.; Mitsumori, S. An Assertion-Based Information-Probe System: Document-Skeleton and Glossary-Skeleton Approach. *Inf. Knowl. Syst. Manag.* **2002**, *3*, 123–152.