Control and Verification of Relations in the Creation of Ontology- Based Electronic Dictionaries for Language Learning

Antonio Vaquero, Francisco Álvarez, Fernando Sáenz

Universidad Complutense de Madrid, Facultad de Informática, Departamento de Sistemas Informáticos y Programación, C/ Prof. José García Santesmases, s/n, E-28040, Madrid, Spain

Universidad Autónoma de Sinaloa, Ángel Flores y Riva Palacios, s/n, C.P 80000, Culiacán, Sinaloa, México

vaquero@sip.ucm.es, francisco_alvarezm@fdi.ucm.es, fernan@sip.ucm.es

Abstract. In our previous work, we used a methodology based on software engineering principles to develop tools for building and querying electronic dictionaries based on a DAG-shaped taxonomy, with language learning purposes. However, the tools do not enforce any kind of control over the use of semantic relations during the DAG-shaped taxonomy construction. In order to teach specific semantic relations (e.g., is-a, part-of, etc.) the tools must control their appropriate usage, thus preventing learners from relying in their intuition to construct the taxonomy, and hence forefend them from making inappropriate and inconsistent modelling choices. Based on an analysis of common problems in existing lexical resources, we present a conceptual model as a first step towards the use of the methodology for the construction of dictionary building educational tools, where the usage of relations can be controlled by refining the semantics of concepts and relations.

1. Introduction

We have stated elsewhere [1,2] that language is an important value, but is nonetheless, inside the classroom, a weak knowledge domain because of the technical challenge it poses to both teachers and students. It can be argued that this knowledge deficiency can be mitigated using paper based resources (e.g., dictionaries, thesauri, etc.) or their electronic counterparts. However, as [3] point out, these resources put severe restrictions on access to words, and their conceptual counterparts: meanings. Thus, the appropriate language learning environments must be provided in order to palliate this language deficiency [4]. Moreover, the interfaces of these environments should be designed to provide an intuitive, coherent and effective way for accessing the desired information contained in the database.

We have developed several tools [1, 2] that overcome some of the deficiencies of most electronic dictionaries, by providing both a form-based and meaning-based access to the information [3, 5]. The tools are based on a decompositional theory of word meaning that has two levels of representation: a conceptual-semantic level and a syntactic-semantic level. The former is represented by a conceptual taxonomy or ontology that expresses the meanings of words, and the latter by words grouped in synonym clusters where lexical-syntactic information can be recorded. Furthermore, the tools follow an imposed constructivist model [7], that is, they provide means for gathering, representing ("externalizing" in the wording of [8]), structuring and creating navigational objects [9].

Our approach has several educational advantages. First, the ontology serves as a useful cognitive tool to extend memory and ease information processing, by allowing a learner to express meanings and relations directly [10]. Second, the grouping of words is also cognitively meaningful, as grouping makes clear the common features and differences between groups [10]. Third, semantic and lexical relations are an important aspect of natural language learning, because, as opposed to paper-based dictionaries, people do not store words in alphabetic order, rather they do it by meaning and relations (links/associations between words/concepts) [3]. In addition, this ontology-based approach is also becoming increasingly popular in the e-learning field, where topic maps are used for the classification, navigation or exploration of concepts, instances, relations and resources within a subject domain [11].

Nevertheless, the development of ontologies for educative purposes suffers from the same problems of their artificial intelligence counterparts: mainly the misuse and confusion of ontological relations [12]. Although the AI community has already recognized the importance of evaluating ontology technologies [13], this topic has been ignored by the Computers in Education community. For instance, in [11], the authors state that their tool for representing and structuring learning content uses several ontological relations (e.g., part-of, superclass-subclass, etc.); however, they do not describe any verification mechanism to control their usage and ensure that the intended meaning of these relations is clearly stated. The same applies to our tools [1, 2], although they allow representing a taxonomy based on a single implicit relation, the precise nature of the relation is not clear.

It would be relatively easy to take our tools, and integrate validation mechanisms for relations like OntoClean [14], in them. However, we have created these tools following our own methodology based on relational database development and software engineering principles [15], and we intend to keep using it. With that in mind, we intend to modify the conceptual model we used in our latest work [2], and enhance it as a first step towards the use of the methodology for the development of our next generation of educative tools.

The rest of the paper is organized as follows. In section 2, we point out the taxonomy related problems found in some electronic dictionaries. In section 3, we depict a set of ideas intended to help developers to formally specify the meaning of concepts and relations in more detail. In section 4, a conceptual model that integrates the aforementioned ideas is introduced and described. Finally, in section 5, some conclusions and future work are outlined.

2. Electronic Dictionaries for People and Machines

Dictionaries are a vital component of any natural language processing system, whether natural or artificial. In their modern electronic form, dictionaries have tremendous potential, provided they are built in a way that allows for their use not only by experts or machines, but also by ordinary language users [5]. However, despite the enormous interest in electronic dictionaries in general, little attention has been paid to the language users [5, 16].

One of the important features that is present in any dictionary (electronic or not) is the presence of semantic relations linking lexical items, and thus forming a semantic network . However, the networks or hierarchies present in most electronic dictionaries are flawed. Thus, it is fundamental to identify these flaws in order to avoid past and present mistakes and create a sound conceptual model that leads to an electronic dictionary where these errors can be avoided.

In this section, we will make a brief review of some of these flaws. Nonetheless, because of the lack of space we will focus on machine-readable dictionaries (MRD) and WordNet, as both are commonly cited and reused resources for computer-assisted language learning [5, 16].

2.1 Machine-Readable Dictionaries

In the electronic versions of paper-based dictionaries (MRD) semantic relations are implicit and are difficult to detect by a language user, because the information contained in these dictionaries is garbled in some way [17]. This makes these resources near to useless for vocabulary learning, mainly because the rich associative semantic network that a dictionary can contain remains inaccessible to them, and because that same network is so flawed that even a state-of-the-art NLP application could never straighten it up [17].

Consequently, once a hierarchy has been extracted from a MRD, it is noticed that, at the higher levels, some concepts simply lack a term to designate them exactly. This lack of clear-cut terms for higher level concepts generates (at least) [17]: circular definitions yielding hierarchies containing loops, and ruptures in knowledge representation (e.g., a utensil is a container) that could mislead a language learner to wrong inferences.

2.2 WordNet

WordNet is a resource that is a computational reflection of the lexical memory of a native speaker. Although [5] argues that such an organization is useful for language learning, a close look at WordNet’s structure points out to a different direction.

In WordNet, the is-a, part-of and member-of relations between synsets (WordNet’s representation of concepts) in WordNet are not used in a consistent way: sometimes they are redundant and sometimes they are broken [18]. For example, according to the taxonomic relations linking the hypernyms of “fever” in WordNet 1.6, “fever” ends up being categorized as “psychological feature” [19]. Moreover, the semantic relations in WordNet do not always reflect its purpose, as they do not always reflect an ontological relation [12]

Furthermore, all the notions (e.g., concepts, taxonomic relations, lexical relations, words) in WordNet are conflated into a unique structure. Thus, it is hard to tell, by looking at its structure, the difference between a concept and the word that describes it. This could impair the learning process, because, if the NLP community has spent years trying to learn how to use WordNet, for a novel language learner, it will be difficult if not impossible.

2.3 Relationship Control in Ontology-Based Electronic Dictionaries

Based on the above empirical evidence, we claim that a dictionary, in order to be of any utility to a language learner, needs to have a clear-cut structure, where all the subtle things that are normally implicit or clouded are explicitly represented at their appropriate level. Moreover, if a learning tool is to constructively allow a learner to create an ontology, this tool must enforce the use of relation control mechanisms, so that it is not created relying essentially on the intuition of the developer.

Therefore, the ideas used in ontological engineering to properly structure taxonomies [14, 20], can and should be used in learning environments such as ours, in order to teach students the appropriate meaning of basic semantic relations, and thus help them improve their conceptual modelling and reasoning skills. For example, in [21] the authors present a Q/A system designed to assist a modeler in properly structuring a taxonomy, by asking a series of questions.

3. Refining the Semantics of Concepts and Relations

In order to give our first step towards the enhancement of the conceptual model, we need to clearly state what are the elements that will be abstracted and represented in our upgraded conceptual model, and help us avoid the problems described in section 3.

These elements are properties of concepts, concrete relations, algebraic and intrinsic properties of relations, and implicit relations. They will help an ontology developer to specify, for concepts and relations, formal and informal semantics that clarify the intended meaning of both entities in order to avoid the problems discussed in section 2. Informal semantics are the textual definitions for both concepts and relations as opposed to formal semantics that are represented by the properties of concepts and relations.

However, the fact that these elements will be part of the enhanced conceptual model does not imply that they are an imposition but rather a possibility, a recommendation that is given to each developer.

3.1 Properties of Concepts

These are formal semantic specifications of those aspects that are of interest to the ontology developer. In particular, these specifications may be the metaproperties of [14] (e.g., R, I, etc.)

3.2 Concrete Relations

Instead of relations with an unclear meaning (e.g. subsumption), we propose the use of relations with well-defined semantics, up to the granularity needed by the ontology developer. We call these, concrete relations (e.g., is-a, part-of, etc.) For a given application, on a given domain, a set of concrete relations is needed. Perhaps the is-a relation is enough. Maybe it is only necessary the part-of one. Perchance both of them, or mayhap it is necessary to consider more concrete relations.

Relations	**Definitions**	**Examples**
C *is-a* C_1	Every C at any time is at the same time a C_1	*myelin is-a lipoprotein*
C *part-of* C_1	Every C at any time is part of some C_1 at the same time	*nucleoplasm part-of nucleus*

Table 1. Definitions and Examples of Relations

Relations	**Transitive**	**Symmetric**	**Reflexive**
is-a	+	-	+
part-of	+	-	+

Table 2. Algebraic Properties of some Concrete Relations

3.3 Algebraic Properties of Relations

The meaning of each relation between two concepts must be established, supported by a set of algebraic properties from which, formal definitions could be obtained (e.g., transitivity, asymmetry, reflexivity, etc.) This will allow reasoning applications to automatically derive information from the resource, or detect errors in the ontology [20]. Moreover, the definitions and algebraic properties will ensure that the corresponding and probably general-purpose relational expressions are used in a uniform way [20]. Tables 1 and 2 (taken from [20]) show a set of concrete relations with their definitions and algebraic properties.

3.4 Intrinsic Properties of Relations

How do we assess, for a given domain, if a specific relation can exist between two concepts? The definitions and algebraic properties of relations, although useful, are not enough. As [12,14] point out, we need something more. Thus, for each relation, there must be a set of properties that both a child and its parent concept must fulfill for a specific relation to exist between them. We call these properties, intrinsic properties of relations.

4. Designing an Ontology-Based Monolingual Electronic Dictionary

In this section, we present an application of our proposal: an E-R schema of an ontology-based electronic dictionary (shown in figure 1), upgraded from the one in [2]. However, we will focus only on the ontological side and leave the rest for future papers.

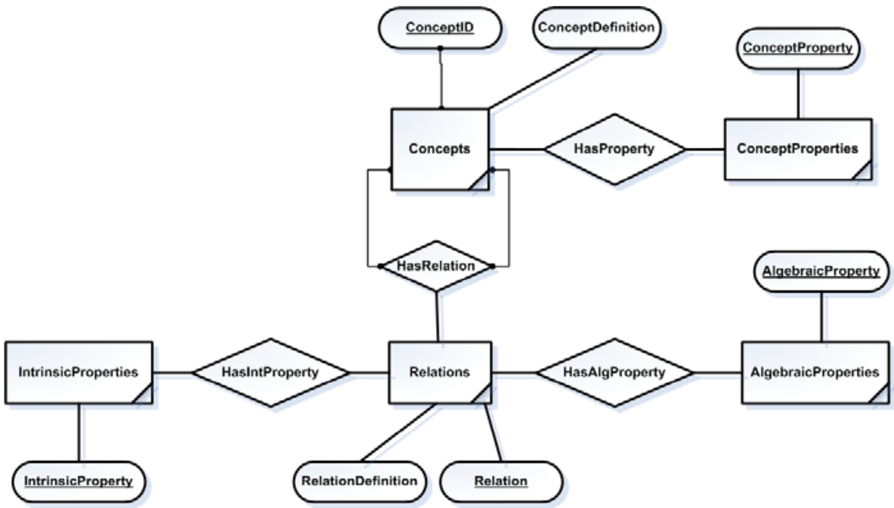

Fig. 1. Conceptual Model for an Ontology-Based Electronic Dictionary

4.1 Entity Sets

The entity set Concepts denotes the meaning of words, and it has two attributes: ConceptID (artificial attribute intended only for entity identification, which shall be explained later), and ConceptDefinition, intended for the textual definition of the meaning. The entity set ConceptProperties represents the set of properties described in section 3.1, and it has one attribute: ConceptProperty used to represent each property.

The entity set Relations represents the set of relations that can exist in an ontology, and it has two attributes: Relation that captures the textual name of each relation (e.g., is-a, part-of, etc.), and RelationDefinition for the textual definition of relations as seen in table1.

The entity set AlgebraicProperties represents the properties of relations as seen in table 2, and it has one attribute: AlgebraicProperty that denotes each algebraic property. The entity set IntrinsicProperties conveys the set of properties mentioned in section 3.4 and has one attribute: IntrinsicProperty representing each intrinsic property.

4.2 Relationship Sets

The relationship set HasProperty is used to assign properties to concepts. The ternary relationship set HasRelation is used to represent that two concepts in an ontology can be linked by a given relation. The relationship set HasAlgProperty is used to convey that relations could have attached a set of algebraic properties; the same applies for the relationship set HasIntProperty but for intrinsic properties.

5. Conclusions and Future Work

We have made a first step towards the use of our methodology for the creation of tools to build and query electronic dictionaries based on DAG-shaped taxonomies. We have acknowledged the need for control and verification mechanisms in the creation and usage of taxonomic relationships, and pointed out its importance in learning environments that use taxonomies to organize knowledge. Thus, we have proposed the use of ontological techniques to improve the semantic cleanness of taxonomies for tasks that go from word meaning categorization to learning content classification. This implied that our previous conceptual model [2] had to be modified.

The new model can account for any number of ontological relationships (as long as they are binary). Furthermore, we have incorporated as part of the model, a set of ideas that are needed to design ontology-based electronic dictionaries where the use of semantic relations and the addition of concepts can be controlled.

Nevertheless, there are a few issues that we need to consider before going through the logical and physical design stages of the methodology, and thus develop the interface(s). First, how to implement the Q/A system as part of tools, so that the notions described in section 3 are transparent to the user. Second, how to improve the friendliness of the interface by providing the user with facilities that allow him to quickly find the needed information (perhaps integrating the ideas of [3, 5, 11]).

References

1. Vaquero, A., Sáenz, F. and López, C. 2003. Herramientas para la Creación de Diccionarios Monolingües con Objetivos Pedagógicos. Challenges 2003 - 5º SIIE, Braga (Portugal), September, 2003.

2. Vaquero, A., Alvarez, F., and Sáenz, F. 2005. Learning Linguistic Concepts through the Construction of Dictionaries with a DAG shaped Taxonomy. In Proceedings of the EduTech 2005 workshop. Perth, Western Australia.
3. Bilac, S. and Zock, M. 2003. Towards a User-Friendly Dictionary Interface. In Proceedings of the Papillon 2003 Workshop, 3-5 July 2003, NII, Sapporo, Japan.
4. Zeltzen, D. and Addison, R. K. 1997. Responsive Virtual Environments, Communications of the ACM, 40(8), August.
5. Zock, M. 2002. Electronic Dictionaries for Men, Machines or for both? Third Papillon seminar, NII, Tokyo, Japan.
7. Cabrera, A., 1995, Informática educativa: La revolución constructivista, Informática y Automática, 28(1).
8. Vygotsky, L. 1986. Thought and Language, The MIT press.
9. Zeiliger, R., Belisle, C., Cerratto, T. 1999. Implementing a Constructivist Approach to Web Navigation Support. In proceedings of the ED-MEDIA'99 Conference, AACE, Seattle,Wa., USA. 1999.
10. Tversky, B. 2000. Some ways that maps and diagrams communicate. In Freksa, C., Brauer, W., Habel, C., and Wender, K. F.(eds). Spatial cognition II: Integrating abstract theories, empirical studies, formal models, and practical applications. Springer-Verlag, Berlin.
11. Dicheva, D., Dichev, C., Wang. D. 2005. Visualizing Topic Maps For E-Learning, Workshop on Applications of Semantic Web in E-Learning (SW-EL@ICALT'05), in conjunction with 5th IEEE International Conference on Advanced Learning Technologies, ICALT 2005, July 5-8, 2005, Kaohsiung, Taiwan, 950-951.
12. Guarino, N. 1998. Some Ontological Principles for Designing Upper Level Lexical Resources. A. Rubio et al. (eds.), Proceedings of the First International Conference on Language Resources and Evaluation, Granada, 28-30 May, Paris: ELRA, 527-534.
13. York, S., Gómez-Pérez, A., Daelemans,W., Reinberger, M., Guarino, N., Fridman, N. 2004. Why Evaluate Ontology Technologies? Because It Works! IEEE Intelligent Systems 19(4): 74-81 (2004).
14. Guarino, N. and Welty, C. 2004. An Overview of OntoClean. In Steffen Staab and Rudi Studer, eds., The Handbook on Ontologies. Pp. 151-172. Berlin:Springer-Verlag.
15. Sáenz, F. and Vaquero, A. Applying Relational Database Development Methodologies to the Design of Lexical Databases. Database Systems 2005, IADIS Virtual Multi Conference on Computer Science and Information Systems (MCCSIS 2005), ISBN 972-8939-00-0, April, 2005.
16. Md Maruf Hasan, Kazuhiro Takeuchi, Virach Sornlertlamvanich and Hitoshi Isahara. 2004. Putting NLP Tools into Action: Incorporating NLP Tools in Web-based Language Learning Environment. In Proceedings of the Asian Symposium on Natural Language Processing to Overcome Language Barriers, IJCNLP-04, Sanya City, Hainan Island, China, pp. 85-90.
17. Ide, N., and Veronis, J. 1993. Extracting Knowledge Bases from Machine-Readable Dictionaries: Have we wasted our time? In Proceedings of the First International Conference on Building and Sharing of Very Large-Scale Knowledge Bases. Tokyo.
18. Martin, P. 2003. Correction and Extension of WordNet 1.7. ICCS 2003, 11th International Conference on Conceptual Structures,(© Springer Verlag, LNAI 2746), pp. 160-173), Dresden, Germany, July 21-25.
19. Burgun, A. and Bodenreider, O. 2001. Aspects of the Taxonomic Relation in the Biomedical Domain. In: C Welty C. Second international conference on formal ontologies in information systems, Ogunquit, Maine.
20. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall CJ, Neuhaus F, Rector A, Rosse C. 2005. Relations in Biomedical Ontologies. Genome Biology, 6(5).
21. Nicola N. and Welty C. 2000. Ontological Analysis of Taxonomic Relationships. In Proceedings of ER-2000: The International Conference on Conceptual Modelling. Springer-Verlag LNCS.