The Problem of Learning Non-Taxonomic Relationships of Ontologies from Text

Ivo Serra[1], Rosario Girardi[1] and Paulo Novais[2]

[1] Federal University of Maranhão, Computer Science Departament, São Luís, Brazil
[2] University of Minho, Computer Science Departament, Braga, Portugal
ivocserra@gmail.com, rosariogirardi@gmail.com, pjon@di.uminho.pt

Abstract. Manual construction of ontologies by domain experts and knowledge engineers is a costly task. Thus, automatic and/or semi-automatic approaches to their development are needed. Ontology Learning aims at identifying its constituent elements, such as non-taxonomic relationships, from textual information sources. This article presents a discussion of the problem of Learning Non-Taxonomic Relationships of Ontologies and defines its generic process. Four techniques representing the state of the art of Learning Non-Taxonomic Relationships of Ontologies are described and the solutions they provide are discussed along with their advantages and limitations.

Keywords: Ontology, Ontology learning, Non-taxonomic relationships, Natural Language Processing.

1 Introduction

Manual construction of ontologies by domain experts and knowledge engineers is a costly task, thus automatic and/or semi-automatic approaches to their development are needed. Ontology Learning (OL) [4, 5] aims at identifying the constituent elements of an ontology, such as non-taxonomic relationships from textual information sources. Some techniques have been proposed for Learning Non-Taxonomic Relationships of Ontologies (LNTRO). All of them use Natural Language Processing (NLP) techniques [1, 7] to annotate the corpus with the information needed for the subsequent processing. Information Extraction (IE) techniques [12] are used to extract from the annotated corpus possible relationships and Machine Learning (ML) [19] or Statistic Techniques (ST) to make a refinement of the relationships outputted from the previous phases. This article discusses the problem of LNTRO, identifying its phases and what kind of techniques can be used to perform the activities of each phase. Four techniques of the state of the art on LNTRO are also described and the advantages and limitations of the solutions they adopt for each phase of LNTRO are discussed.

The paper is organized as follows. Section 2 introduces an ontology definition and the lexical realizations of non-taxonomic relationships. Section 3 defines the problem of LNTRO, its phases and what techniques can be used to approach each one. Section 4 describes four representative techniques of the state of the art on LNTRO and which solutions they adopt for each of its phases described in section 3. Finally, section 5 presents the conclusions discussing general and open research topics on LNTRO.

2 Non-Taxonomic Relationships of Ontologies

In order to define the problem of LNTRO it is necessary to formally characterize non-taxonomic relationships of ontologies and how they are realized in the text. The sub-sections 2.1 and 2.2 discuss these issues.

2.1 A Formal Definition of Ontology

An ontology is a formal and explicit specification of a shared conceptualization of a domain of interest [15]. *Conceptualization* refers to an abstract model of some phenomenon in the world. *Explicit*, means that the type of concepts used and the limitations of their use are explicitly defined. *Formal*, refers to the fact that the ontology should be machine readable. *Shared*, reflects the notion that an ontology captures consensual knowledge, that is, it's not private to some individual but accepted by a group. Currently, ontologies are applied in areas such as the communication of software agents [11], integration of information [2], composition of Web Services [22], description of contents to facilitate their recovery [16] in NLP [14], in the Semantic Web [3], in building knowledge-based systems [6] and in applications of knowledge management [17]. Formally, an ontology can be represented by a 6-tuple [13]:

$$O = (C, H, I, R, P, A). \tag{1}$$

where,

$C = C_C \cup C_I$ is the set of entities of the ontology. They are designated by one or more terms in natural language. The set C_C consists of classes, i.e., concepts that represent entities that describe a set of objects (for example, "Person" $\in C_C$) while the set C_I is constituted by instances, (for example "Anne Smith" $\in C_I$);

$H = \{kind\_of\ (c_1, c_2) \mid c_1 \in C_C, c_2 \in C_C\}$ is the set of taxonomic relationships between concepts, which define a concept hierarchy and are denoted by "kind_of(c_1, c_2)", meaning that c_1 is a subclass of c_2. For instance, "kind_of(Costumer, Person)";

$I = \{is\_a\ (c_1, c_2) \mid c_1 \in C_I \wedge c_2 \in C_C\}$ is the set of relations between classes and its instances. For example, "is_a(Erick, Lawyer)";

$R = \{rel\ (c_1, c_2, ..., c_n) \mid \forall i, c_i \in C\}$ is the set of ontology relationships that are neither "kind_of" nor "is_a". For example "represent(Lawyer, Costumer)" and "represent(Erick, Anne Smith)";

$P = \{propC\ (c_k, tipo) \mid c_k \in C_C\} \cup \{propI\ (c_k, valor) \mid c_k \in C_I\}$ is the set of properties of ontology classes. "propC" defines the datatype of a property while propI defines its value. For instance, “subject(Case, String)” is a propC element while “subject(Case 12, adoption)” is a propI element.

$A = \{condition_x \Rightarrow conclusion_y (c_1, c_2,..., c_n) \mid \forall j, c_j \in C_C\}$ is a set of axioms, rules that allow checking the consistency of an ontology and infer new knowledge through some inference mechanism. The term $condition_x$ is given by $condition_x = \{(cond_1, cond_2, \ldots, cond_n) \mid \forall z, cond_z \cup H \cup I \cup R\}$. For example, “apply(defense_argument22, Case12) $\wedge$ similar(Case12, Case13) $\Rightarrow$ apply(defense_argument22, Case13)” is a rule indicating that these two legal cases are similar thus the same defense argument can be used in both cases.

2.2 Linguistic Realizations of Non-Taxonomic Relationships

Non-taxonomic relationships can be classified as domain independent or domain dependent. Domain independent relationships are of two subtypes ownership or aggregation. Aggregation is the "whole-part" relationship. For example, in the sentence "The car's wheel is out of order." there is a non-taxonomic relationship of aggregation between "car" and "wheel". The linguistic realization of the relationship of aggregation occurs in two forms: the possessive form of English (apostrophe) and the verb "to have" in any conjugation. However, the converse is not true, that is, the occurrence of such linguistic accomplishments does not imply a relationship of aggregation as will be explained in the next case. Ownership relationships are held as in the example: "Father and mother will wait for the court's decision." in which there is a relationship of ownership between "court" and "decision". The linguistic realization of this kind of relationship occurs in two forms: the possessive form of English (apostrophe) and the verb "to have" in any conjugation. However, the converse is not true, that is, the occurrence of such linguistic accomplishments does not imply a relationship of possession.

Domain dependent relationships are expressed by particular terms of an area of interest. For example, the sentence "The court will judge the custody in three days." holds the relationship "judge" between "court" and "custody" which is characteristic of the legal field. Table 1 summarizes the types of non-taxonomic relationships and their linguistic realizations.

Table 1. Types of non-taxonomic relationships and its linguistic realizations.

Type	Subtype	Linguistic realization	Sentences with non-taxonomically related concepts
Domain Independent	Aggregation	Possessive and the verb to have any conjugation	"The **car**’s **wheel** is out of order." "The dual core **UCP** has several **registers**."
	Ownership		"The couple will wait for the **court**’s **decision**."
Domain Dependent	—	Verbs of the domain	"The **court** will judge the **custody** in three days."

3 The problem of LNTRO

LNTRO is an approach to automate or semi-automate the extraction of these relationships from textual information sources. Non-taxonomic relationships correspond to the R set of an ontology (section 2.1). For example, "represents" is a non-taxonomic relationship between the classes "lawyer" and "client" in the legal domain.

LNTRO can generally be accomplished through the tasks of "Corpus construction", "Extraction of candidate relationships" (which in turn consists of the subtasks of "Corpus annotation" and "Extraction of relationships") and "Refinement" (Fig. 1).

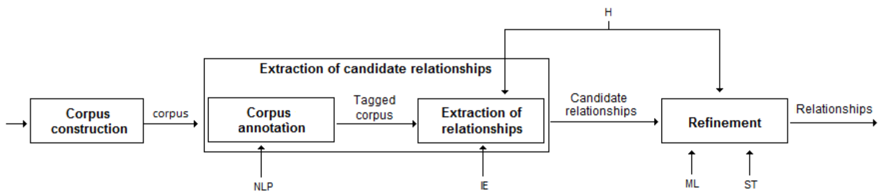

Fig. 1. The generic process of LNTRO.

The task of "Corpus construction" consists of selecting documents on the domain we expect to extract relationships from. This is usually a costly task and the outcome of any LNTRO technique depends on its quality.

The "Extraction of candidate relationships" task aims at identifying a set of possible relationships. It has the corpus built in the previous phase as input and candidate relationships as its product. It is composed of two sub-activities: "Corpus annotation" and "Extraction of relationships". The "Corpus annotation" task consists of applying tags to the text with NLP techniques that are necessary for the next steps in LNTRO. The "Extraction of relationships" task consists of searching in the annotated corpus for evidences suggesting the existence of relationships. For example, Maedech [18] considers the existence of two instances of ontology concepts in a sentence as evidence that they are non-taxonomically related. For Villaverde et al. [24] a relationship is identified by the presence of two concepts of an ontology in the same sentence with a verb between them. This sub-task can also receive the concepts of the ontology (C set) as input. In this case the search space for relationships is reduced and there is a potential for achieving greater precision in the extraction of relationships.

Relationships from the previous task should not be recommended to the specialist, since there is usually a substantial amount of them that do not correspond to good suggestions. For this reason Machine Learning (ML) or Statistic Techniques (ST) can be used in the "Refinement" phase. The ontology taxonomy (set H) can also be given as input. In this case the LNTRO technique is able to suggest to the specialist the best possible level in the hierarchy where to add the relationship. This functionality is explained in section 4.2. Table 2 summarizes the phases of LNTRO.

Table 2. Phases of LNTRO.

<table>
<tr><th colspan="2">Phases</th><th colspan="2">Description</th></tr>
<tr><td colspan="2">Corpus construction</td><td colspan="2">Selection of documents in quantity and quality required for LNTRO</td></tr>
<tr><td rowspan="2">Extraction of candidate relationships</td><td>Corpus annotation</td><td>Annotate the corpus using NLP techniques required for the continuity of LNTRO</td><td rowspan="2">Extraction of an initial set of relationships</td></tr>
<tr><td>Extraction of relationships</td><td>Application of the algorithm for extraction of relationships from the annotated corpus</td></tr>
<tr><td colspan="2">Refinement</td><td colspan="2">Application of ML or ST techniques to suggest the most probably relationships.</td></tr>
</table>

4 Techniques for LNTRO

In the following sections, four state of the art techniques for LNTRO are presented. The solutions adopted to approach the generic phases of LNTRO are highlighted and their positive aspects and limitations are discussed.

4.1 LNTRO based on the Extraction of Association Rules

This technique described in [24] has two phases "Identification of occurrences of relationships" and "Mining associations". The "Identification of occurrences of relationships" receives a corpus and a set of concepts of an ontology and outputs a set of tuples in the form $<c_1, v, c_2>$, where c_1 and c_2 are the ontology concepts and v is a verb. Initially using Wordnet [10] each ontology concept is extended with its synonyms to increase the recall of the search. Then the POS-tagging is performed in order to identify the verbs. For sentences that satisfy the following two conditions a tuple (c_1, v, c_2) is generated: (a) sentences that have exactly two concepts and a verb between them and (b) the two concepts are at a maximum distance of D terms. "D" is a parameter whose value is defined experimentally by the specialist and corresponds to the maximum number of terms that must exist between two concepts for them to be considered related. For example if D = 3 then for the sentence "The Judged the court in custody three days." a tuple <court, judge, custody> is generated since there are two terms between the concepts. However, for the sentence "The court of North Dakota will judge the three days in custody." no tuple is generated.

Once a set of candidate relationships (set of tuples outputted from the previous phase) is obtained, "Mining associations" can be performed that aim at refining the results of the previous phase before they are suggested to the specialist. For this purpose an algorithm to extract association rules [23] is used. The product of this phase are non-taxonomic relationships represented by association rules in the form $<c_1 \land c_2>\rightarrow<v>$, which have values of support and confidence greater than the minimum defined experimentally by the specialist.

For example, in the sentence "Our data suggests that lipoxygenase metabolites activate ROI formation which then induce IL-2 expression via NF-kappa B activation", Lipoxygenase (Li) and Reactive Oxygen Intermediates (ROI) are concepts and Activate (Ac) is a verb. In the first phase the tuple <Li, ROI, Ac> is generated representing the fact that the two extraction conditions described previously were satisfied. In the second phase if the rule <Li, ROI>→<Ac> has values of support and confidence greater than or equal to the minimum support and confidence, it is recommended to the specialist. Table 3 shows which solutions have been adopted for each one of the generic phases for LNTRO as defined in section 3.

Table 3. Solutions for LNTRO based on the Extraction of Association Rules.

Phase	Adopted solution
Corpus construction	Ad-hoc. A corpus already available in the medical field (Genia) was used in its experiment.
Corpus annotation	POS-tagging
Extraction of relationships	Uses the algorithm already described to extract candidate relationships in the form of tuples (<c_1, c_2, v>)
Refinement	Uses a technique known as the "Extraction of Association Rules" to suggested non-taxonomic relationships in the form of rules (<$c_1 \wedge c_2$>→ <v>)

A positive aspect of this proposal is that it labels with verbs the relationships between two concepts found in each sentence. In addition, the search space for relations is restricted since ontology concepts are given as input to the technique, thus potentially leading to better results. The technique only extracts concepts from the text; concept instances are ignored. Furthermore, it doesn't use stemming, a NLP technique that could lead to better recall values. Moreover, one restriction is the fact that no treatment is given to the possessive form "'s" that is one of the linguistic realizations of non-taxonomic relationships which can be present in the corpus with reasonable frequency. In addition, the authors refer to the verbs as single words when in fact, in most of the cases, they appear in the form of verbal phrases. In Genia [20], the corpus used to illustrate and evaluate the technique, coincidentally most of the verbal phrases are composed of a single term, which is a uncommon fact. Therefore, to be applied to corpora without this characteristic, the technique should be updated either to work with verb phrases or with the information of which verb, among those of the verb phrase, should be used as the label of the relationship. For the evaluation of the technique, the recall and precision measures are used, fact that we consider too restrictive for a noisy area like AO.

4.2 LNTRO based on the Extraction of Generalized Association Rules

Maedech and Staab [18] propose a process similar to that of Villaverde et al. [24], with the difference that it uses an algorithm of generalized association rules [18] to suggest the possible most appropriate hierarchical level for the relationship and works with texts in German.

The technique has two phases "Text processing" and "Mining associations". In the first phase, the objective is to extract pairs of concepts from the text that correspond to candidate relationships. For this purpose, the title and the sentence heuristics are used. The first one says that a pair of related concepts should be created for every concept in the text with every concept in the title. This heuristic is based on the intuition that the concepts that appear in the text body are related to the concepts that appear in the title. The second one sets up a tuple for each pair of concepts that are present in the same sentence.

In the second phase relationships in the form of pairs of concepts from the previous phase are submitted to an algorithm for Mining Generalized Association Rules [18]. The goal is to extract non-taxonomic relationships in the form of association rules and suggest the best possible level in the hierarchy where to add the relation-ships.

After applying the Extraction of Generalized Association Rules, the rule *area → hotel* is discarded because *area -> accommodation* is an ancestral rule (its concepts are in the same or higher levels in the ontology taxonomy) and has values for support or confidence greater or equal than the descendent rule. The same happens to the rules *room -> television* and *room → furnishing* (Table 4). The solutions adopted for each one of the generic phases for LNTRO are shown in Table 5.

Table 4. Extracted Relationships [18].

Discovered relations	Confidence	Suport
(area → accommodation)	0,38	0,04
~~(area → hotel)~~	~~0,1~~	~~0,03~~
(room → furnishing)	0,39	0,03
~~(room → television)~~	~~0,29~~	~~0,02~~
(accommodation → address)	0,34	0,05
(restaurant → accommodation)	0,33	0,02

Table 5. Solutions for LNTRO based on the Extraction of Generalized Association Rules.

Phase	Adopted solution
Corpus construction	Ad-hoc. A corpus already available in the touristic domain (Lonely Planet) was used in its experiment.
Corpus annotation	Uses chunking, stemming and REN.
Extraction of relationships	Uses sentence and title heuristics to extract candidate relationships as concept pairs (CP = $\{(a_{i,1}, a_{i,2}) \mid a_{i,j} \in C\}$).
Refinement	Uses a technique known as mining generalized association rules [ref] to recommend relationships as association rules in the form $c_1 \rightarrow c_2$.

A positive aspect of this proposal is the use of the algorithm for the Extraction of Generalized Association Rules that suggests the best possible level in the ontology taxonomy where the relationship should be added. On the other hand, a limitation is the fact that the technique does not label the relationships but, only indicates what classes are related. Furthermore, it uses gazetteers lists to associate instances in the text with ontology classes. This makes the effectiveness of the technique dependent on the extent of these lists.

4.3 LNTRO based on Queries on Web Search Engines

Sanchez and Moreno [21] propose an automatic technique for LNTRO that is able to learn verbs from a domain, extract related concepts and label them using the Web instead of a traditional corpus as a source for the construction of an ontology. Despite being diverse and unstructured, according to the authors, the redundancy of information in an environment as vast as the Web is a measure of its relevance and veracity. The first phase is the extraction and selection of verbs that express relationships characteristic of the domain. Based on morphological and syntactic analysis, verbs that have a relationship with the domain keyword are extracted. Then, the degree of relationship between each verb and the domain is measured. To do so, statistical measures are made about the term distribution on the web. The obtained values are used to rank the list of candidate verbs. This lets one choose the labels of non-taxonomic relationship that are closely related to the domain. The domain related verbs are used to discover non-taxonomic related concepts. To do so it queries the web with the patterns "keyword domain-verb" or "domain-keyword verb" that returns a corpus related to the specified query. The goal is to search the content of documents to find concepts that proceed ("*High sodium diets* are associated with hypertension") or succeed ("Hypertension is caused by *hormonal problems*") the constructed patterns. These concepts are candidate to be non-taxonomically related to the original keyword. Table 6 shows which solutions have been adopted for each one of the generic phases for LNTRO as defined in section 3.

Table 6. Solutions for LNTRO based on Queries on Web Search Engines.

Phase	Adopted solution
Corpus construction	Based on documents returned by a Web search engine.
Corpus annotation	Chunking.
Extraction of relationships	Extracts verb phrases and noun phrases as labels and concepts of relationships respectively.
Refinement	Statistical processing based on the result of queries in a web search engine.

A positive aspect in this proposal is that specialists do not have to deal with the construction or selection of corpora, a generally laborious task. They are automatically created with the help of a web search engine. In addition, the process is domain independent and fully automatic. On the other hand, one limitation is that learning relationships is dependent of learning concepts and vice versa which makes the process less flexible.

4.4 LNTRO based on logistic regression

Fader, Soderland and Etzioni [9] propose a technique that is domain independent and extracts non-taxonomic relationships from corpora in English. It uses a syntactic and a lexical constraint.

The syntactic constraint requires that verbal phrases match the following patterns: a verb (e.g. "invented"), a verb immediately followed by a preposition (e.g. "located in"), or a verb followed by nouns, adjectives or adverbs ending with a preposition (e.g. "has atomic weight of"). The syntactic constraint reduces "uninformative" extractions, for example, for the sentence "Faust made a deal with the Devil" the tuple <Faust, made, devil> corresponds to a non informative extraction. The relationship extracted using the syntactic patterns would be <Faust, made a deal with, devil>, which is a valid relationship. However, it allows the extraction of relationships considered too "specific". As an example, let us consider the sentence "The Obama administration is offering only modest greenhouse gas reduction targets at the conference". The syntactic patterns will match the phrase "is offering only modest greenhouse gas reduction targets at". Thus, there are phrases that satisfy the syntax constraint, but are not relationships. To overcome this limitation the lexical constraint is used to separate sentences that represent real relationships from those very specific ones, such as the example sentence. The restriction is based on the intuition that a valid relational sentence must have many different arguments in a large corpus. The example sentence is specific to the pair of arguments "Obama administration" and "conference", so it is unlikely to represent a relationship. The lexical restriction is implemented by a repository of verb phrases that are considered sufficiently generic (have many different arguments). The repository is manually built and whenever a verb phrase meets any of the syntactic patterns it is checked against it. Verb phrases not present in the repository are not recommended as relationships.

The technique has three phases as follows. The phases of "Extraction of relationships" and "Extraction of arguments" have high recall but low precision. Thus a refinement is required in order to reveal the most probable relationships among all extracted from the application of the syntactic and lexical constraints.

- Extraction of relationships: For each verb "v" in a sentence "s", find the longest sequence of words "r" such that (a) "r" is initiated by "v", (b) "r" satisfies the syntactic constraint and (c) "r" satisfies the lexical constraint.
- Extraction of arguments: For each verb phrase "r" identified in the previous step, find the noun phrase "x" closer to the left of "r" in the sentence such that "x" is not a relative pronoun, adverb "Who" or existential "there". Find the noun phrase "y" closer to the right of "r" in "s". If a pair (x, y) has been found, return (x, r, y) as an extracted relationship.
- Refinement: In this phase a logistic regression classifier [19] is used to rank relationships according to the probability of representing valid relationships. Table 7 shows which solutions have been adopted for each one of the generic phases of LNTRO as defined in section 3.

Table 7. Solutions for LNTRO based on Logistic Regression.

Phase	Adopted solution
Corpus construction	Ad-hoc construction.
Corpus annotation	Chunking.
Extraction of relationships	Uses the algorithms described in the phases "Relations extraction" and "Extraction of arguments".
Refinement	Uses a logistic regression classifier to assign a probability to each relationship.

The technique extracts relationships with labels in the form of verb phrases that comply with the syntactic and lexical constraints. This way, extracted relationships have the maximum semantics. Moreover, the technique is capable of extracting relationships from very small corpus, such as a single sentence and from any area of knowledge (domain independent).

A limitation of this approach is that it uses a manually built repository of verb phrases, containing those that are considered sufficiently generic to be present in sentences relating various concepts. If a relationship potentially "valid" is represented by a verb phrase that is not in the repository, it will be discarded.

5 Concluding Remarks

This work approached the LNTRO problem, its phases and the knowledge areas which provide solutions to them. Four techniques of the state of the art on LNTRO were presented and the solutions each one adopted for the phases of LNTRO were highlighted. Advantages and limitations of each of the techniques were also discussed. To end our considerations on LNTRO, we now discuss some relevant issues and point out a line of research.

The corpus used for LNTRO may contain classes, instances of classes or both. For the first case, a search is performed for classes in the text. This search can include the synonyms of the concepts and/or their stems, thus increasing the recall of the extracted concepts from the corpus. If the corpus has only instances of classes it is necessary to use Named Entity Recognition (NER). If the corpus has both classes and instances, all these solutions can be used together.

Non-taxonomic relationships are generally represented by a pair of concepts and optionally a label. The first representation has the disadvantage of being semantically poorer because we know which classes are related but do not have a name giving a meaning to the relationship. The second is the representation that has the highest semantics since the relationships are constituted by a pair of concepts and a label. The label is generally a verb phrase found between the two concepts in a sentence.

In the "Corpus annotation" phase the following NLP techniques are generally used: tokenization which separates the text into tokens, and is a prerequisite for any other NLP technique; sentence splitter separates the text into sentences since for most, if not all, the LNTRO techniques relationships are found between terms in the same sentences; NER when the input corpus has instances of ontology classes and chuncking which identifies sintagmas. In the context of LNTRO the relevant ones are noun phrases, considered as concepts by techniques that do not receive them as input and verb phrases generally used as the labels for relationships.

LNTRO techniques that use an ontology taxonomy (H set) as input can suggest the best level in the hierarchy where to insert the relationship. Those that receive only the ontology concepts (C set) have the search space for relationships reduced and have the potential of obtaining better results when compared to those that don't receive this input. Techniques that don't receive any of these sets as input often consider noun phrases as concepts.

Techniques on LNTRO are usually evaluated comparing their results against reference ontologies [8]. However, comparing them when executed under similar conditions is a work that still has to be done.

References

1. Allen, J.: Natural Language Understanding. Redwood City, CA: The Benjamin/Cummings Publishing Company, Inc. (1995)
2. Alexiev, V., Breu, M., de Bruijn, J., Fensel, D., Lara, R., Lausen, H.: Information Integration with Ontologies: Experiences from an Industrial Showcase, Wiley (2005)
3. Bontcheva, K., Cunningham, H.: The Semantic Web: A New Opportunity and Challenge for Human Language Technology, In Proceedings of the Workshop on Human Language Technology for the Semantic Web and Web Services, 2nd International Semantic Web Conference, Sanibel Island (2003)
4. Buitelaar, P., Cimiano, P., Magnini, B.: Ontology Learning from Text: An Overview. DFKI, Language Technology Lab. AIFB, University of Karlsruhe. ITC-irst (2003)
5. Buitelaar, P., Cimiano, P., Magnini, P.: Ontology Learning from Text: Methods, Evaluation and Applications, IOS Press, Amsterdam, The Netherlands (2006)
6. Cimiano, P., Volker, J., Studer, R.: Ontologies on Demand? – A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text. In Information, Wissenschaft und Praxis 57 (6-7): 315-320 (2006)
7. Dale, R., Moisl, H., Somers, H. L.: Handbook of natural language processing. CRC (2000)
8. Dellschaft, K., Staab, S.: On how to perform a gold standard based evaluation of ontology learning. In: Proceedings of the 5th International Semantic Web Conference, pp. 228 – 241, Athens. Springer (2006)
9. Fader, A., Soderland, S., Etzioni, O.: Identifying Relations for Open Information Extraction. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland (2011)
10. Fellbaum, C.: WordNet: An Electronic Lexical Database. Cambridge: MIT Press. 23-24 p. (1998)
11. Finin, T., Fritzson, R., McKay, D., McEntire, R.: KQML as an Agent Communication Language. In: Proceedings of the 3rd International Conference on Information and Knowledge management, p. 456-463 (1994)

12. Freitag, D.: Information extraction from HTML: Application of a general machine learning approach. In Proceedings of the 15th Conference on Artificial Intelligence. pp. 517-523, (1998)
13. Girardi, R.: Guiding Ontology Learning and Population by Knowledge System Goals. In: Proceedings of International Conference on Knowledge Engineering and Ontology Development, Ed. INSTIIC, Valence, pp. 480 – 484 (2010)
14. Girardi, R., Ibrahim, B.: Using English to retrieve software. Journal Of Systems Software: Special Issue on Software Reusability, New York, Elsevier. v. 30, n. 3, p. 249 - 270, (1995).
15. Gruber, T. R.: Toward Principles for the Design of Ontologies used for Knowledge Sharing, International Journal of Human-Computer Studies. Nº 43, pp. 907-928 (1995)
16. Guarino, N., Masolo, C., Vetere, C.: Ontoseek: Content-based Access to the web. IEEE Intelligent Systems, vol. 14 (3), pp. 70-80 (1999)
17. Jurisica, I., Mylopoulos, J., Yu, E.: Using ontologies for knowledge management: an information systems perspective. In: Knowledge: creation, organization and use – Proceedings of the 62nd annual meeting of the American Society for Information Science, Washington, DC, pp 482–496 (1999)
18. Maedche, A., Staab, S.: Mining non-taxonomic conceptual relations from text. In Knowledge Engineering and Knowledge Management. Methods, Models and Tools: 12th International Conference. Proceedings. 189-202. Berlin: Springer (2000)
19. Mitchell, T.: Machine Learning, McGraw Hill. (1997)
20. Rinaldi, F. et al.: Mining relations in the GENIA corpus. Proceedings of the Second European Workshop on Data Mining and Text Mining for Bioinformatics: 61 - 68. (2004)
21. Sanchez, D., Moreno, A.: Learning non-taxonomic relationships from web documents for domain ontology construction. Data and Knowledge Engineering, 64(3), p. 600-623 (2008)
22. Sirin, E., Hendler, J., Parsia, B.: Semi-automatic composition of web services using semantic descriptions. In: Proceedings of the ICEIS Workshop on Web Services: Modeling, Architecture and Infrastructure (2002)
23. Srikant, R., Agrawal, R.: Mining generalized association rules. In Proc. of VLDB' 95, pages 407-419 (1995).
24. Villaverde, J., Persson, A., Godoy, D., Amandi, A.: Supporting the discovery and labeling of non-taxonomic relationships in ontology learning. Expert Syst. Appl. 36(7): 10288-10294 (2009)