Using Domain-Specific Verbs for Term Classification

Irena Spasić
Computer Science
University of Salford, UK
I.Spasic@salford.ac.uk

Goran Nenadić
Department of Computing
UMIST, UK
G.Nenadic@umist.ac.uk

Sophia Ananiadou
Computer Science
University of Salford, UK
S.Ananiadou@salford.ac.uk

Abstract

In this paper we present an approach to term classification based on verb complementation patterns. The complementation patterns have been automatically learnt by combining information found in a corpus and an ontology, both belonging to the biomedical domain. The learning process is unsupervised and has been implemented as an iterative reasoning procedure based on a partial order relation induced by the domain-specific ontology. First, term recognition was performed by both looking up the dictionary of terms listed in the ontology and applying the C/NC-value method. Subsequently, domain-specific verbs were automatically identified in the corpus. Finally, the classes of terms typically selected as arguments for the considered verbs were induced from the corpus and the ontology. This information was used to classify newly recognised terms. The precision of the classification method reached 64%.

1 Introduction

Basic notions used when describing a specific problem domain are concepts, classes and attributes (or features). The identification of concepts, linguistically represented by domain-specific terms (Maynard and Ananiadou, 2000), is a basic step in the automated acquisition of knowledge from textual documents. Textual documents describing new knowledge in an intensively expanding domain are swamped by new terms representing newly identified or created concepts. Dynamic domains, such as biomedicine, cannot be represented by static models, since new discoveries give rise to the appearance of new terms. This makes the automatic term recognition (ATR) tools essential assets for efficient knowledge acquisition.

However, ATR itself is not sufficient when it comes to organizing newly acquired knowledge. Concepts are natively assorted into groups and a well-formed model of a domain, represented through terms and their relations, needs to reflect this property consistently. Dynamic domain models should be able to adapt to the advent of new terms representing newly discovered or identified concepts. In other words, newly extracted terms need to be incorporated into an existing model by associating them with one another and with already established terms preferably in an automated manner. This goal may be achieved by relying on *term clustering* (the process of linking semantically similar terms together) and *term classification* (the process of assigning terms to classes from a predefined classification scheme). In particular, classification results can be used for efficient and consistent term management through populating and updating existing ontologies in expanding domains such as biomedicine. In this paper, we compare some of the term classification approaches and introduce another approach to this problem.

The paper is organised as follows. In Section 2 we provide a brief overview of the existing term classification approaches and suggest the main idea of our approach to this problem. Section 3 describes the learning phase of our classification method. Further, Section 4 provides details on the classification algorithm. Finally, in Section 5 we describe the evaluation strategy and provide the results, after which we conclude the paper.

2 Term Classification Approaches

Similarly to general classification algorithms, the existing term classification approaches typically rely on learning techniques. These techniques are most often statistically based (e.g. hidden Markov

models, naive Bayesian learning, etc.). Other techniques include decision trees, inductive rule learning, support-vector machines (SVMs), etc. We, on the other hand, suggest the use of a genetic algorithm as a learning engine for the classification task. Let us now discuss some approaches to the automatic classification of biomedical terms.

Nobata et al. (2000) implemented a statistical method for term classification. In their approach, each class was represented by a list of (single) words. The first step was to estimate the conditional probability $P(c \mid w)$ of each word w being assigned to a specific class c, based on the assumption that each word occurrence is independent of its context and position in the text. Further, yet another strong restriction was made by assuming that there was one-to-one correspondence between terms and their classes. In addition, this approach is not applicable to "unknown" terms, i.e. terms containing words for which no classification probabilities had been determined. A special class, referring to "other", was introduced to cover such words. Bearing in mind the increasing number of new terms, such an approach is bound to produce skewed results, where many of the terms would simply be classified as "other".

While Nobata et al. (2000) statistically processed the information found inside the terms, Collier et al. (2001) applied statistical techniques to the information found outside the terms. A hidden Markov model based on *n*-grams (assuming that a term's class may be induced from the previous *n*-1 lexical items and their classes) was used as a theoretical basis for their classification method. The method relied on the orthographic features including numerals, capital and Greek letters, special characters (such as `-`, `/`, `+`, etc.), parenthesis, etc. In the biomedical domain, such features often provide hints regarding the class of a specific term. Each unclassified term was assigned a class of the most similar (with respect to the orthographic features) term from the training set. This approach encountered the minority class prediction problem. Namely, the best classification results in terms of recall and precision were achieved for the most frequent class of terms in their training corpus, while the worst results were those achieved for the least frequent class.

Hatzivassiloglou et al. (2001) proposed a method for unsupervised learning of weights for context elements (including words as context constituents and the corresponding positional and morphological information) of known terms and using these weights for term classification. Three well-known learning techniques were used: naive Bayesian learning, decision trees, and inductive rule learning. Simplified classification experiments in which a classification algorithm was choosing between two or three options respectively were conducted. The precision of binary classification was around 76% for all three learning algorithms, and the precision dropped to approximately 67% when choosing between three options. If the proposed techniques were to be applied for general classification where the number of options is arbitrary, the precision is expected to decrease even further.

Nenadic et al. (2003b) conducted a series of large-scale experiments with different types of features for a multi-class SVM. These features included document identifiers, single words, their lemmas and stems, and automatically recognised terms. The results indicated that the performance was approximately the same (around 60% in the best case) when using single words, lemmas or stems. On the other side, terms proved to be better (more than 90% precision) than single words at lower recall points (less than 10%), which means that terms as features can improve the precision for minority classes. The best results were achieved with document identifiers, but such features cannot be used on the fly in new documents.

Spasic et al. (2002) used a genetic algorithm (GA) based on a specific crossover operator to explore the relationships between verbs and the terms complementing them. The GA performed reasoning about term classes allowed to be combined with specific verbs by using an existing ontology as a seed for learning. In this paper, we use the results of the proposed methodology as a platform for term classification. In the following section we briefly overview the method for the acquisition of verb complementation patterns.

3 Verb Complementation Patterns

By looking at the context of an isolated verb occurrence it is difficult to predict all term classes that can be combined with the given verb. On the other hand, the whole "population" of terms complementing a specific verb is likely to provide a certain conclusion about that verb with respect to its

complementation patterns. This was a primary motivation for Spasic et al. (2002) to use a GA as it operates on a population of individuals as opposed to a single individual. This fact also makes the approach robust, since it does not rely on every specific instance of verb-term combination to be correctly recognised.

As not all verbs are equally important for the term classification task, we are primarily interested in domain-specific verb complementation patterns. In our approach, a complementation pattern of a domain-specific verb is defined as a disjunction of terms and/or their classes that are used in combination with the given verb. The automatic acquisition of these patterns is performed in the following steps: term recognition, domain-specific verb extraction, and the learning of complementation patterns. Let us describe each of these steps in more detail.

3.1 Term Recognition

First, a corpus is terminologically processed: both terms present in the ontology and the terms recognised automatically are tagged. Terms already classified in the ontology are used to learn the classes allowed by the domain-specific verbs, while the new terms are yet to be classified based on the learnt classes. New terms are recognized by the *C/NC-value method* (Frantzi et al., 2000), which extracts multi-word terms. This method recognises terms by combining linguistic knowledge and statistical analysis. Linguistic knowledge is used to propose term candidates through general term formation patterns. Each term candidate t is then quantified by its termhood *C-value*(t) calculated as a combination of its numerical characteristics: length $|t|$ as the number of words, absolute frequency $f(t)$ and two types of frequency relative to the set $S(t)$ of candidate terms containing a nested candidate term t (frequency of occurrence nested inside other candidate terms and the number of different term candidates containing a nested candidate term):

$$C-value(t) = \begin{cases} \ln|t| \cdot f(t), & \text{if } S(t) = \varnothing \\ \ln|t| \cdot (f(t) - \frac{1}{|S(t)|} \sum_{s \in S(t)} f(s)), & \text{if } S(t) \neq \varnothing \end{cases}$$

Obviously, the higher the frequency of a candidate term the greater its termhood. The same holds for its length. On the other side, the more frequently the candidate term is nested in other term candidates, the more its termhood is reduced. However, this reduction decreases with the increase in the number of different host candidate terms as it is hypothesised that the candidate term is more independent if the set of its host terms is more versatile.

Term distribution in top-ranked candidate terms is further improved by taking into account their context. The relevant context words, including nouns, verbs and adjectives, are extracted and assigned weights based on how frequently they co-occur with top-ranked term candidates. Subsequently, context factors are assigned to candidate terms according to their co-occurrence with top-ranked context words. Finally, new termhood estimations (NC-values) are calculated as a linear combination of the C-values and context factors.

Nenadic et al. (2003a) modified the C/NC-value to recognise acronyms as a special type of single-word terms, and, thus, enhanced the recall of the method. On the other hand, the modified version incorporates the unification of term variants into the linguistic part of the method, which also improved the precision, since the statistical analysis is more reliable when performed over classes of equivalent term variants instead of separate terms.

3.2 Domain-Specific Verb Recognition

Verbs are extracted from the corpus and ranked based on the frequency of occurrence and the frequency of their co-occurrence with terms. A stop list of general verbs frequently mentioned in scientific papers independently of the domain (e.g. `observe`, `explain`, etc.) was used to filter out such verbs. The top ranked verbs are selected and considered to be domain-specific. Moreover, these verbs are also corpus-specific (e.g. `activate`, `bind`, etc.). Table 3 provides a list of such verbs, which were used in the experiments.

3.3 Complementation Pattern Learning

In order to learn a verb complementation pattern for each of the selected verbs separately, terms are collected from the corpus by using these verbs as anchors. A GA has been implemented as an itera-

tive reasoning procedure based on a partial order relation induced by the domain-specific ontology.[1] In each iteration pairs of verb complementation patterns represented as sets of terms and term classes are merged. This operation involves the substitution of less general terms/classes by their more general counterparts, if there is a path in the ontology connecting them. Otherwise, the disjunction of the terms is formed and passed to the next iteration. Figure 1 depicts the process of learning a verb complementation pattern.

Since the partial order relation induced by the ontology is transitive, the order in which terms are processed is of no importance. The final verb complementation patterns are minimal in the sense that the number of terms in a verb complementation pattern and the depth of each individual term in the ontology are minimised.

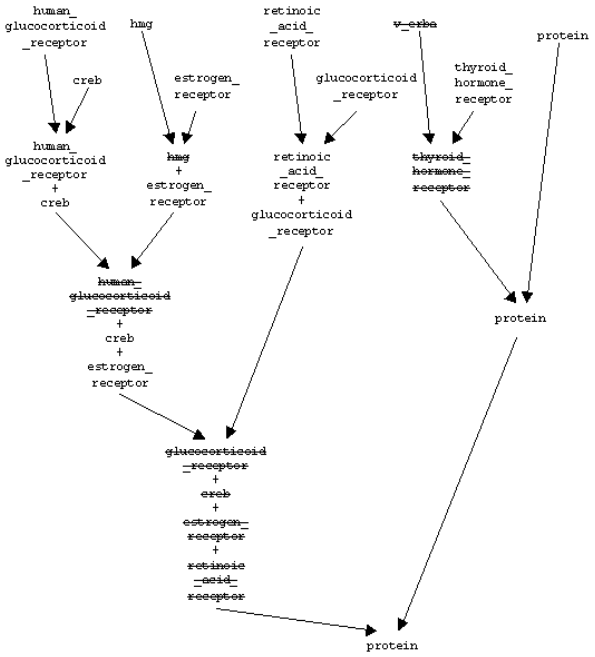

Figure 1. Learning the complementation pattern for the verb `bind`

4 Term Classification Method

The verb complementation patterns have been obtained by running the GA on a set of terms some of which were present in an ontology, which is used during the learning process. The newly recognised terms (i.e. the ones not found in the ontology) will remain included in the final verb complementation patterns as non-classified terms, since at this point it is not known which classes could replace them. All elements of the final verb complementation patterns can be thus divided into two groups based on the criterion of their (non)existence in the ontology. The elements already present in the ontology are candidate classes for the newly recognised terms. Let us now describe the classification method in more detail.

Let $V = \{v_1, v_2, \ldots, v_n\}$ be a set of automatically identified domain-specific verbs. During the phase of learning verb complementation patterns, each of these verbs is associated with a set of classes and terms it co-occurs with. Let $C_i = \{c_{i,1}, c_{i,2}, \ldots, c_{i,m_i}\}$ denote a set of classes assigned automatically to the verb v_i ($1 \le i \le n$) by a learning algorithm based on the information found in the corpus and the training ontology. As indicated earlier, we define such set to be a verb complementation pattern for the given verb.

4.1 Statistical Analysis

As we planned to use verb complementation patterns for term classification, we modified the original learning algorithm (Spasic et al., 2002) by attaching the frequency information to terms and their classes. When substituting a less general class by its more general counterpart,[2] the frequency information is updated by summing the two respective frequencies of occurrence. In the final verb complementation pattern, each class $c_{i,j}$ has the frequency feature $f_{i,j}$, which aggregates the frequency of co-occurrence with v_i ($1 \le i \le n$; $1 \le j \le m_i$) for the given class and its subclasses. The frequency information is used to estimate the class probabilities given a verb, $P(c_{i,j} \mid v_i)$:

$$p_{i,j} = \frac{f_{i,j}}{\sum_{l=1}^{m_l} f_{i,l}}$$

[1] The partial order relation is based on the hierarchy of terms/classes: term/class t_1 is in relation with t_2, if there is a path in the ontology from t_2 to t_1. In that case, we say that t_2 is *more general* than t_1.

[2] The ontology used for learning allowed multiple inheritance only at the leaf level, that way incurring no ambiguities when substituting subclass by its superclass. The multiple inheritance at the leaf level was resolved by mapping each term to all its classes, which were then processed by a GA.

Unclassified terms remain present in the final verb complementation patterns, and, like classes, they are also assigned the information on the frequency of co-occurrence with the given verb. When classifying a specific term, this information is used to select the verb based on whose pattern the term will be classified. Precisely, the verb the given term most frequently co-occurs with is chosen, as it is believed to be the most indicative one for the classification purpose.

4.2 Term Similarity Measure

A complementation pattern associated with the chosen verb typically contain several classes. In order to link the newly recognised terms to specific candidate classes, we used a hybrid term similarity measure, called the *CLS similarity measure*. It combines contextual, lexical and syntactic properties of terms in order to estimate their similarity (Nenadic et al., 2002).

Lexical properties used in the CLS measure refer to constituents shared by the compared terms. The rationale behind the lexical term similarity involves the following hypotheses: (1) Terms sharing a head are likely to be hyponyms of the same term (e.g. `progesterone receptor` and `oestrogen receptor`). (2) A term derived by modifying another term is likely to be its hyponym (e.g. `nuclear receptor` and `orphan nuclear receptor`). Counting the number of common constituents is a simple and straightforward approach to measuring term similarity, but it falls short when it comes to single-word terms and those introduced in an ad-hoc manner. Thus, properties other than lexical need to be included.

We use syntactic properties in the form of specific lexico-syntactical patterns indicating parallel usage of terms (e.g. `both Term and Term`). All terms used within a parallel structure have identical syntactic features and are used in combination with the same verb, preposition, etc., and, hence, can be regarded as similar with high precision. However, patterns used as syntactic properties of terms have relatively low frequency of occurrence compared to the total number of terms, and in order to have a good recall, a large-size corpus is needed. In order to remedy for small-size corpora, other contextual features are exploited.

Context patterns (CPs) in which terms appear are used as additional features for term comparison. CPs consist of the syntactic categories and other grammatical and lexical information (e.g. `PREP NP V:stimulate`). They are ranked according to a measure called CP-value (analogue to C-value for ATR). The ones whose CP-value is above a chosen threshold are deemed significant and are used to compare terms. Each term is associated with a set of its CPs, and contextual similarity between terms is then measured by comparing the corresponding sets. Automatically collected CPs are indeed domain-specific, but the method for their extraction is domain independent.

4.3 Term-Class Similarity

The CLS similarity measure applies to pairs of terms. However, in case of multiple choices provided by the verb complementation patterns, we need to compare terms to classes. In order to do so, we use the similarity between the given term and the terms belonging to the classes. The selection of terms to be compared is another issue. One possibility is to use the full or random set of terms (belonging to the given class) that occur in the corpus. Alternatively, some ontologies provide a set of prototypical instances for each class, which can be used for comparison of terms and classes.[3] More formally, if c is a class, e_1, e_2,..., e_k are terms representing the class, and t is a term, then the similarity between the term t and the class c is calculated in the following way:

$$Ex(t,c) = \max_{i \in \{1,\ldots,k\}} \frac{CLS(t,e_i)}{\sqrt{\sum_{j=1}^{k} CLS^2(t,e_j)}}$$

This example-based similarity measure maximises the value of the CLS measure between the term and the instances representing the class. In addition, the values of the CLS measure are mapped into the interval (0,1) by performing vector normalisation in order to make them comparable to the class probability estimations.

4.4 Term Classification

Finally, given the term t and the verb v_i it most frequently co-occurs with, a score is calculated for

[3] For example, in the UMLS ontology each class is assigned a number of its prototypical examples represented by terms.

each class $c_{i,j}$ from the set C_i according to the following formula:

$$C(t, c_{i,j}) = a \cdot p_{i,j} + (1-a) \cdot Ex(t, c_{i,j}) \quad (1)$$

where a ($0 \le a \le 1$) is a parameter, which balances the impact of the class probabilities and the similarity measure.[4] A class with the highest $C(t, c_{i,j})$ score is used to classify the term t. Alternatively, multiple classes may be suggested by setting a threshold for $C(t, c_{i,j})$.

At this point, let us reiterate that the final verb complementation patterns are minimal in the sense that the number of terms in a verb complementation pattern and the depth of each individual term in the ontology are minimised. The latter condition may cause the classification to be crude, that is – new terms will be assigned to classes close to the root of the ontology. For more fine-grained classification results, the classes placed close to the root of the ontology should be either removed from the initial verb complementation patterns, thus being unable to override the classes found lower in the hierarchy or in other way prevented from substituting less general terms. The depth up to which the terms are to be blocked may be empirically determined.

5 Experiments and Evaluation

5.1 Resources

The resources used for the experiments include an ontology and a corpus, both belonging to the domain of biomedicine. We used an ontology, which is a part of the UMLS (Unified Medical Language System) knowledge sources (UMLS, 2002). UMLS integrates biomedical information from a variety of sources and is regularly updated. Knowledge sources maintained under the UMLS project include: METATHESAURUS linking term variants referring to the same concepts; SPECIALIST LEXICON providing syntactic information for terms, their component words, and general English words; and SEMANTIC NETWORK containing information about the classes to which all METATHESAURUS concepts have been assigned.

The knowledge sources used in our term classification experiments include METATHESAURUS and SEMANTIC NETWORK. As the number of terms in METATHESAURUS was too large (2.10 million terms) and the classification scheme too broad (135 classes) for the preliminary experiments, we made a decision to focus only on terms belonging to a subtree of the global hierarchy of the SEMANTIC NETWORK. The root of this subtree refers to substances, and it contains 28 classes.

The corpus used in conjunction with the above ontology consisted of 2082 abstracts on nuclear receptors retrieved from the MEDLINE database (MEDLINE, 2003). The majority of terms found in the corpus were related to nuclear receptors and other types of biological substances, as well as the domain-specific verbs extracted automatically from the corpus in the way described in Section 3.

5.2 Evaluation Framework

When retrieving terms found in the context of domain-specific verbs (see Section 3 for details) both terms found in the ontology and terms recognised on the fly by the C/NC-value method should be extracted. However, for the purpose of evaluation, only terms classified in the ontology were used. In that case, it was possible to automatically verify whether such terms were correctly classified by comparing the classes suggested by the classification method to the original classification information found in the ontology.

During the phase of retrieving the verb-term combinations, some of the terms were singled out for testing. Namely, for each verb, 10% of the retrieved terms were randomly selected for testing, and the union of all such terms formed a testing set (138 terms) for the classification task. The remaining terms constituted a training set (1618 terms) and were used for the learning of complementation patterns.

5.3 Results

Based on the training set, domain-specific verbs were associated with the complementation patterns given (see Table 1 for examples). Then, each term from the training set was associated with the verb

[4] Note that when $a = 0$, the classification method resembles the nearest neighbour classification method, where the examples are used as a training set. On the other hand, when $a = 1$, the method is similar to naive Bayesian learning. However, in both cases the method represents a modification of the mentioned approaches, as the classes used in formula (1) are not all classes, but the ones learned by the GA.

it most frequently co-occurred with. The complementation pattern learnt for that verb was used to classify the term in question.

Verb	Complementation pattern
activate bind	Immunologic Factor Receptor Enzyme Hormone Organic Chemical Hazardous or Poisonous Substance Pharmacologic Substance

Table 1. Learnt verb complementation patterns

Since the UMLS ontology contains a number of prototypical examples for each class, we have used these class representatives to compare unclassified terms to their potential classes as indicated in Section 4. Table 2 shows the results for some of the terms from the testing set and compares them to the correct classifications obtained from the ontology.

Term	Suggested class	Correct classes
4 hydroxy-tamoxifen	Organic chemical	Organic chemical
benzoic acid	Organic chemical	Organic chemical Pharmacologic substance
testoster-one	Pharmacologic substance	Steroid Pharmacologic substance Hormone

Table 2. Examples of the classification results

Note that in UMLS one term can be assigned to multiple classes. We regarded a testing term to be correctly classified if the automatically suggested class was among these classes. Table 3 provides information on the performance of the classification method for each of the considered verbs separately and for the combined approach in which the verb most frequently co-occurring with a given term was used for its classification. The combined approach provided considerably higher recall (around 50%) and a slight improvement in precision (around 64%) compared to average values obtained with the same method for each of the verbs separately. The classification precision did not tend to very considerably, and was not affected by the recall values. The recall could be improved by taking into account more domain-specific verbs, while the improvement of precision depends on proper tuning of: (1) the module for learning the verb complementation patterns, and (2) the similarity measure used for the classification. Another possibility is to generalize the classification method by relying on domain-specific lexico-syntactic patterns instead of verbs. Such patterns would have higher discriminative power than verbs alone. Moreover, they could be acquired automatically. For instance, the CP-value method can be used for their extraction from a corpus (Nenadic et al., 2003a).

Verb	Recall	Precision	F-measure
activate	19.28	66.59	29.90
bind	29.30	66.53	40.68
compete	3.58	63.16	6.78
conserve	2.41	61.82	4.64
inhibit	16.62	62.81	26.28
interact	13.16	64.31	21.85
mediate	11.68	62.75	19.69
modulate	10.44	64.13	17.96
repress	6.18	62.91	11.25
stimulate	9.39	63.25	16.35
Average:	12.20	63.83	20.48
Combined:	**49.88**	**64.18**	**56.13**

Table 3. The performance of the classification method

The values for precision and recall provided in Table 3 refer to the classification method itself. If it were to be used for the automatic ontology update, then the success rate of such update would also depend on the performance of the term recognition method, as the classification module would operate on its output. We used the C/NC-value method for ATR; still any other method may be used for this purpose. We have chosen the C/NC-value method because it is constantly improving and is currently performing around 72% recall and 98% precision (Nenadic et al., 2002).

6 Conclusion

Efficient update of the existing knowledge repositories in many rapidly expanding domains is a burning issue. Due to an enormous number of terms and the complex structure of the terminology, manual update approaches are prone to be both inefficient and inconsistent. Thus, it has be-

come absolutely essential to implement efficient and reliable term recognition and term classification methods as means of maintaining the knowledge repositories. In this paper, we have suggested a domain independent classification method as a way of incorporating automatically recognised terms into an existing ontology. For the preliminary experiments, we used the UMLS ontology in the domain of biomedicine, but the method can be easily adapted to use other ontologies in any other domain.

The classification method makes use of the contextual information. Not all word types found in the context are of equal importance in the process of reasoning about the terms: the most informative are verbs, noun phrases (especially terms) and adjectives. The presented term classification approach revolves around domain-specific verbs. These verbs are used to collect unclassified terms and to suggest their potential classes based on the automatically learnt verb complementation patterns.

Note that not every term appearing in a corpus is guaranteed to be classified by the proposed classification method due to the fact that a term need not occur as a complement of a domain-specific verb. Still, for a large number of terms the classification method is expected to obtain the classification information, as it is highly probable (though not certain) for a term to occur in a context of a domain-specific verb. The main goal of the method is to provide aid for the automatic ontology update by populating newly recognised terms into an existing ontology, rather than classifying arbitrary term occurrences in the corpus.

The presented classification method can be easily modified to use lexical classes other than verbs as a criterion for classification. Even more, it can be further generalised to use a combination of lexical classes, which can be specified as a set of lexico-syntactic patterns. Further experiments with the generalisation of the classification method by basing it on a set of domain-specific lexico-syntactic patterns instead of domain-specific verbs are expected to demonstrate better performance in terms of recall and precision. These facts suggest that our classification approach, in combination with the C/NC-value method, could be reliably used as a (semi)automatic ontology maintenance procedure.

References

Nigel Collier, Chikashi Nobata and Junichi Tsujii. 2001. *Automatic Acquisition and Classification of Terminology Using a Tagged Corpus in the Molecular Biology Domain*. Journal of Terminology, John Benjamins.

Katerina Frantzi, Sophia Ananiadou and Hideki Mima. 2000. *Automatic Recognition of Multi-Word Terms: the C-value/NC-value Method*. International Journal on Digital Libraries 3(2):115-130.

Vasileios Hatzivassiloglou, Pablo Duboue and Andrey Rzhetsky. 2001. *Disambiguating Proteins, Genes, and RNA in Text: A Machine Learning Approach*. Bioinformatics, 1(1):1-10.

Diana Maynard and Sophia Ananiadou. 2000. *Identifying Terms by their Family and Friends*. Proceedings of COLING 2000, Saarbrucken, Germany, 530-536.

MEDLINE. 2003. National Library of Medicine. Available at: http://www.ncbi.nlm.nih.gov/PubMed/

Goran Nenadic, Irena Spasic and Sophia Ananiadou. 2002. *Automatic Acronym Acquisition and Term Variation Management within Domain-Specific Texts*. Proceedings of LREC-3, Las Palmas, Spain, 2155-2162.

Goran Nenadic, Irena Spasic and Sophia Ananiadou. 2003a. *Automatic Discovery of Term Similarities Using Pattern Mining*. To appear in Terminology.

Goran Nenadic, Simon Rice, Irena Spasic, Sophia Ananiadou and Benjamin Stapley. 2003b. *Selecting Features for Text-Based Classification: from Documents to Terms*. Proceedings of ACL Workshop on Natural Language Processing in Biomedicine, Sapporo, Japan.

Chikashi Nobata, Nigel Collier and Junichi Tsujii. 2000. *Automatic Term Identification and Classification in Biology Texts*. Proceedings of the Natural Language Pacific Rim Symposium (NLPRS'2000), 369-375.

Irena Spasic, Goran Nenadic and Sophia Ananiadou. 2002. *Tuning Context Features with Genetic Algorithms*. Proceedings of 3rd International Conference on Language, Resources and Evaluation, Las Palmas, Spain, 2048-2054.

UMLS. 2002. UMLS Knowledge Sources. National Library of Medicine, 13th edition.