Definition of a Linguistic Resource for Opinion Mining

Franco Tuveria and Manuela Angioni

CRS4, Center of Advanced Studies, Research and Development in Sardinia,
Parco Scientifico e Tecnologico, Ed. 1, 09010 Pula (CA), Italy

Abstract. Many approaches to Opinion Mining are based on linguistic resources, lexicons or lists of words. The lack of suitable and/or available resources is one of the main problems in the process of opinion extraction and in general in the analysis of textual resources based on a linguistic approach. In this paper we describe FreeWordNet, a linguistic resource based on WordNet and useful in the automatic method we propose for the extraction of features in a general domain. In FreeWordNet each synset is enriched with a set of properties related to adjectives and adverbs and has a positive, negative or objective value associated. The properties associated to each synset support a better identification of the sentiment expressed in relation to the domain and give more details about the relevant terms or the expressions having an opinion associated.

1 Introduction

The linguistic approach to text analytics needs a detailed analysis of textual resources. A text implicitly contains the necessary knowledge to understand the meaning expressed. Several linguistic resources and knowledge bases support the automatic process of text analysis and understanding. Some of them are referred to the syntactic interpretation of the text in the parsing phase or play a relevant role in the conceptual interpretation of terms and in their sense disambiguation.

Many approaches to Opinion Mining and Sentiment Analysis are based on linguistic resources, lexicons or lists of words. In [1] is proposed a linguistic approach to Opinion Mining, based on a combination of adverbs and adjectives. Other approaches propose a methodology [2] for assign a polarity to word senses applying a Word Sense Disambiguation (WSD) process in building new resources based on WordNet [3] and oriented to Opinion Mining. Some of these approaches represent the base of the work presented in this paper. According to the definition of Opinion Mining given by [4], "Opinion Mining can be roughly divided into three major tasks of development of linguistic resources, sentiment classification, and opinion extraction and summarization". The lack of suitable and/or available resources is one of the main problems in an Opinion Mining process and in general in the analysis of textual resources based on a linguistic approach.

In Opinion Mining the feature extraction is the process of detection of relevant terms or expressions having opinions associated and identifying a domain. Knowing the polarity of words and their meanings can surely help to better identify the opin-

Tuveri F. and Angioni M..
Definition of a Linguistic Resource for Opinion Mining.
DOI: 10.5220/0004093300640073
In *Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science* (NLPCS-2012), pages 64-73
ISBN: 978-989-8565-16-7

ions related to the features expressed in sentences. The context described by sentences contributes to define the meaning of the terms, the relating features, the adjectives and the adverbs according to the domain in order to perform a better Word Sense Disambiguation. In this paper we describe a linguistic resource of adjectives and adverbs based on WordNet, called in the following FreeWordNet, where each synset is enriched with a set of properties related to adjectives and adverbs with a positive, negative or objective value associated. The properties help to better identify the sentiment expressed in relation to the domain and give more details about the opinion expressed. FreeWordNet is mainly involved in the development of an automatic process of feature extraction and especially in the steps of distinction and identification of subjective, objective or factual sentences and contributes in a basic way in the task of contextualization of the features.

The remainder of the paper is organized as follows: Section 2 refers to related works. Section 3 introduces the linguistic resource and the feature extraction system. Section 4 examines the work performed on for the human categorization of adjectives and adverbs, giving some details about the methodology followed, and considering some evaluations and the measures. Finally, Section 5 draws conclusions.

2 Related Works

Many approaches and resources have been proposed in determining the orientation of terms. In [5] the authors evidence that subjectivity is a property to be associated to word senses and that WSD can “directly benefit from subjectivity annotations”. In the conclusions they affirm that a very good agreement can be achieved between human annotators in labeling the polarity of senses.

Close to our work, SentiWordNet [6], [7] is one of the publicly available lexical resources, that extends WordNet thanks to a semi-automatic acquisition of the polarity of WordNet terms, evaluating each synset according to positive, negative and objective values. It provides the possibility to accept user feedback on the values assigned to synsets, allowing the building of a community of users in order to improve SentiWordNet. Despite its wide coverage SentiWordNet does not provide additional information we need to contextualize the content of the sentences, such as the properties we defined in FreeWordNet, able to better characterize the meaning of a term and its use in the context of the sentence.

Another lexical resource consisting of WordNet senses automatically annotated by positive and negative polarity, is Q-WordNet [8] that tries to maximize the linguistic information contained in WordNet, taking advantage of the human effort given by lexicographers and annotators instead of applying supervised classifiers.

WordNet-Affect [9] has been developed starting from WordNet, assigning one or more affective labels (a-labels) to a subset of synsets representing affective concepts that contribute to precise the affective meaning. For example, the a-label Emotion represents the affective concepts related to emotional state. Other concepts are not emotional-affective but represent moods, situations eliciting emotions, or emotional responses. It is available for free only for non-profit institution. The same staff developed WordNet Domains [10], a resource that maps the WordNet synsets to a subset of categories of the Dewey Decimal Classification System. The idea of mapping synsets and categories has in part inspired the development of FreeWordNet.

A further resource built as a “Gold Standard” is MicroWnOp [11], used to validate SentiWordNet. It is a carefully balanced set of 1,105 WordNet synsets manually annotated according to their degrees of polarity with the three scores summing up to 1. MicroWnOp has been adopting two criteria: the opinion relevance, that means that the synsets should be relevant to represent the opinion topic, and the WordNet representativeness, respecting the distribution of the synsets among the four parts of speech.

3 FreeWordNet: A Linguistic Resource for Opinion Mining

The development of an Opinion Mining system able to automatically extract features, independently by the domain, evidenced the need for additional characteristics to the existing resources, such as SentiWordNet and Q-WordNet. Such resources identify the polarity values of WordNet terms but do not provide any information about the context of use of their meanings. In FreeWordNet the meanings expressed by adjectives and adverbs have been extended with polarity values and some properties useful in the steps of the feature extraction process, as described in the following. These properties associated to each synset help to better identify the sentiment expressed in relation to a given domain, provide more details about the features and characterize the content of the sentences in which they are used.

In FreeWordNet adjectives and adverbs have two different levels of categories. The first level of categorization is automatically performed by a Semantic Classifier [12] able to categorize text documents using the categories of WordNet Domains and providing as result a set of categories and weights. In FreeWordNet the Classifier categorizes the glosses of the terms, assigns to each synset a set of categories, such as Person or Gastronomy*,* useful to associate features to adjectives and adverbs.

The second level of categories, related to the human categorization, defines sets of 14 and 7 properties, respectively for adjectives and adverbs, as showed in Table 1 and Table 2 with the polarity evaluation. The idea is that adjectives and adverbs could be grouped in categories according to their meanings. In [13] the authors defined as subjective expressions the: “words and phrases used to express mental and emotional states, such as speculations, evaluations, sentiments and beliefs”. This definition has been extended in FreeWordNet by adding e.g. adjectives related to human senses, like the sense of *Touch* and the sense of *Taste*.

Regarding the adverbs, the properties defined in FreeWordNet consider their meaning, the position or the strength. Based on their characteristics, in FreeWordNet have been considered adverbs of manner, adverbs of place, adverbs of time, adverbs of quantity or degree, of affirmation, negation or doubt (grouped as AND adverbs), adverbs as intensifiers or emphasizers and adverbs used in adversative and in consecutives sentences, as listed in Table 2. Only the adverbs of manner may be positive or negative. The adverbs of degree give the idea about the intensity with which something happens or have an impact on sentiment intensity. The other give additional information to the analysis related to the location or the direction, the time.

Table 1. Properties of adjectives.

Adjectives	Pos.	Neg.	Net.	Tot.
Emotion	52	73	3	128
Moral/Ethic	45	155	2	202
Character	355	584	220	1159
Weather	7	26	6	39
Color	0	9	42	51
Quantity	16	0	9	25
Appearance	41	83	46	170
Material	22	11	54	87
Shape	0	0	30	30
Touch	3	13	6	22
Taste	40	41	5	86
Dimension	11	2	60	73
Chronologic	3	0	30	33
Geographic	0	10	19	29
Others	29	17	87	133
Total	*624*	*1024*	*619*	*2267*

Table 2. Properties of adverbs.

Adverbs	Pos.	Neg.	Net.	Tot.
Time	0	0	7	7
Manner(things)	18	25	8	51
Manner(person)	166	205	5	376
Place	0	0	3	3
Intensifiers	0	0	38	38
Quantity	0	0	6	6
AND	1	0	0	1
Total	*185*	*230*	*67*	*482*

Together the two levels of categories define the context of use of the terms in a sentence. For example, in the tourism domain, if an adjective has *Moral/Ethic* property and is categorized as *Person*, we can expect the context to be related to personnel evaluation. Likewise, in the sentence “the breakfast was good” the adjective “good” has *Taste* property, is categorized as *Gastronomy* and is referred to the feature “breakfast”. The categories and the properties of adjectives and adverbs provide the possibility to separate sentences having polarity valence from the others. We agree with [8] considering in the polarity classification not only word sense, sentence, or text depending on subjectivity but even on polarity factual detection. In fact there might be word senses or sentences objectively having polarity valence. Factual sentences with a polarity valence give information about situations, facts that could be evaluated as positive or negative according to objective criteria related to their description, while subjective sentences give personal opinions about features, facts, etc. FreeWordNet allows us to have a distinction between adjectives having meaning that we consider subjective and adjectives having factual valence. Categories like *Emotion*, *Moral/Ethic*, *Character*, *Taste*, *Touch*, *Appearance* have a subjective valence. *Weather*, *Color*, *Quantity*, *Material*, *Dimension*, *Chronologic*, and *Geographic* express factual polarity values as showed in Table 1. An example is the following sentence: “The season is arid”. This is not a subjective but a factual sentence and expresses a negative meaning from the life point of view. The adjective “arid”, having synset 02552415, in WordNet 3.0, and meaning specified by this gloss: “lacking sufficient water or rainfall", has a negative factual value and has been classified with the property *Weather*.

3.1 The Feature Extraction System

In order to provide more details about the motivation of the creation of FreeWordNet, we describe the automatic method for the extraction of the features from a corpus of reviews in a general domain and the role of FreeWordnet in the process. The proposed method, as depicted in Figure 1, is based on a linguistic approach to the semantic analysis of the opinions expressed in a set of reviews.

The first step of the process is the creation of the corpus of reviews related to a specific domain. In this paper, we do not mind about the way the reviews are gathered from the sources of information. Sentences having orthographic errors are discarded or corrected. Only well-built sentences have been selected and inserted in the corpus in order to avoid introducing errors and to facilitate the syntactic parser activities.

The analysis of the corpus is performed by a set of two modules including, at a top level, a Semantic Classifier and a Sentence Analyzer. The Semantic Classifier performs a thorough syntactic analysis of the sentences and a phrase chunking process through the TreeTagger [14] parser and chunker, able to annotate the text with part-of-speech tags and lemma information. The parser identifies into each sentence its sub-constituents. A Java class wraps the evaluation provided by TreeTagger and, analyzing the parts of speech, identifies the associations between nouns and their related information. Such analysis is used in the semantic categorization process of the corpus of reviews. The text categorization process provides as result a set of categories and weights that define the domain for the corpus of reviews. For example, considering a set of reviews about a hotel, the domain is characterized by categories such as Tourism, Person, Gastronomy, and by their weights.

The Semantic Classifier performs the corpus categorization evaluating the categories and their weights for each sentence. The categories and their weights are compared with the categories describing the domain of the corpus in order to decide if a sentence is relevant. For example, analyzing reviews about tourism and especially reviews about hotels, we expect to examine sentences containing opinions about geographical locations, buildings, rooms, staff and food. Moreover, the categorization of each sentence of the reviews is managed by the Sentence Analyzer in order to distinguish between subjective and objective sentences, with or without orientation, and in particular in order to detect factual sentences having polarity value. In this phase two sets of categories related to the synsets are used: the semantic one, performed automatically by the Semantic Classifier, and the human one, given by the properties of FreeWordNet. The first set of categories allows excluding sentences not belonging to the domain of the corpus. As said, the properties of FreeWordNet related to the *Moral/Ethic* or *Emotional* sphere imply subjective values, while others identifying e.g. *Chronologic* or *Shape* properties imply factual valence. In such a way, we consider only subjective sentences or factual sentences having polarity valence. The pre-processing of the corpus of textual resources has been performed in order to acquire different levels of information, related to the whole corpus, to the sentences or to each term. All the information involved in the categorization process is still used in the feature extraction phase in order to perform the disambiguation of the terms and to extract relations between features, adjectives and adverbs.

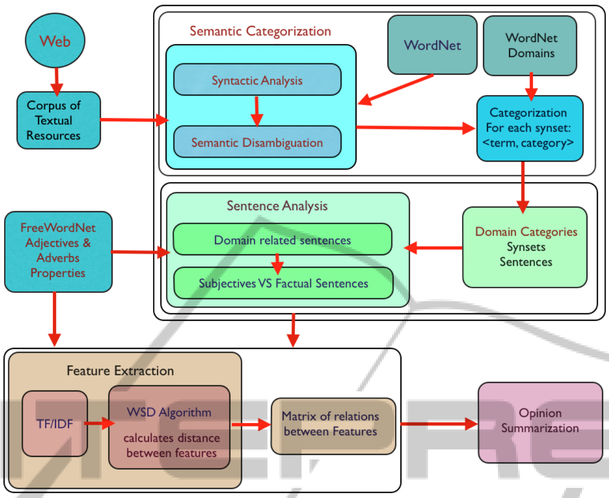
Web
Corpus of Textual Resources
Semantic Categorization
Syntactic Analysis
Semantic Disambiguation
WordNet
WordNet Domains
Categorization For each synset: <term, category>
FreeWordNet Adjectives & Adverbs Properties
Sentence Analysis
Domain related sentences
Subjectives VS Factual Sentences
Domain Categories Synsets Sentences
Feature Extraction
TF/IDF
WSD Algorithm
calculates distance between features
Matrix of relations between Features
Opinion Summarization

Fig. 1. The schema of the feature extraction process.

The feature extraction process consists of two main phases. The first step involves the application of a tf-idf function to the nouns contained in the corpus of sentences having polarity orientation, obtaining as result a first list of candidate features. In the second step the WSD algorithm processes the feature terms in order to perform their disambiguation, excluding synonyms and terms not referred to the domain categories. The features are now identified by their synsets.

The WSD algorithm is inspired to the measure of similarity proposed by Leacock-Chodorow [15] and calculates the semantic distance between the synsets related to the features using the semantic net of WordNet. The algorithm also uses the common categories between the synsets of each pair of terms and provides a weight to each synset based on the number of synsets related to each term. The minimum distance between the synsets is used in order to assign the most probable meaning to each term. Finally the algorithm defines a matrix of all the possible relations between the synsets of the features. The rows and columns of said matrix are the extracted features. The matrix contains as weights the values of distance that measure the strength of the relations existing between two features. The higher the weight, the stronger the relation. The matrix allows grouping features by means of the strength of their relations. The adjectives having a polarity associated and the adverbs of manner and the intensifier ones are put in relation with the feature terms they are referred to in the same sentence of the review. The property and the polarity value of each adjective enrich the information about the related feature. The presence of intensifier adverbs contributes to determine the grade of the expressed opinion. An example of result of the semantic categorization is given by the processing of the sentence: “The arid climate is characterized by a high evaporation and lack of rainfalls”. The Semantic Classifier categorizes the sentence and identifies the most relevant categories (Meteorology 75%, Psychology 25%). The WSD of adjectives performed by the Sentence

Analyzer, assigning the correct property to adjectives and adverbs, identifies the subjective sentences. The adjective “arid” has two meanings, each one related to a different gloss and categorized with different sets of categories. Both the glosses of the adjective “arid” have been analyzed and classified in FreeWordNet. The right sense of the adjective is chosen by considering the matching of the most relevant categories with the categories of both the glosses. The categories related to the synset “302462790” having gloss “lacking sufficient water or rainfall; an arid climate”. The main category associated to the synset, Meteorology, matches the main category of the sentence, providing in such way information about the most probable meaning of the adjective. Its property and the value are used in the polarity evaluation of the sentence.

4 Human Categorization of Adjectives and Adverbs

The human categorization of adjectives and adverbs included in FreeWordNet according to a set of properties has been manually performed as described in the following.

The set of representative properties has been defined in order to put in evidence the most evident characteristics of adjectives and adverbs. We identified 14 properties plus another category collecting adjectives not concerning the other previous categories as described in Table 1. The adverbs have been distinguished in 7 properties, as described in Table 2. The analysis started considering a collection of public resources of terms having polarity valence and publicly available on the Web. The list of terms obtained and selected by their polarity information has been used as starting point for the definition of the database of adjectives and adverbs. Groups of adjectives and adverbs able to guarantee the opinion relevance have been chosen in order to ensure a minimum coverage of the topic related to each property, and the balancing between the polarities. For each adjective and adverb identified, all the possible synsets available on WordNet 3.0 have been considered. A property among the list has been manually associated to a synset by means of the interpretation of the related gloss, assigning a positive, negative or objective value.

The work, valid for the English language, has been performed by two evaluators following some predefined rules in order to define the criteria before associating the synsets to a specific property and to the polarity. A first set of 150 adjectives and adverbs has been selected and together the two evaluators evaluated them in order to align the evaluation criteria. The reason of this approach is due to the similarity of some properties such as *Emotion*, *Moral/Ethic* and *Character* where there is a very slight distinction between their meanings. Then, they have proceeded independently in the evaluation of the remaining adjectives and adverbs. The results obtained independently by the evaluators have been compared in order to emphasize the points of disagreement between them. Every time the categorization by means of the gloss definition of the synsets or the assignment of the polarity generated discrepancy in the interpretation, the results have been compared and discussed by the two people in order to establish a common evaluation. If a convergence of opinions was not possible, the synset was excluded. Only the synsets having a total agreement of both the evaluators in the polarity evaluation and in the association of the properties have been included in FreeWordNet. As said, the level of disagreement was mainly related to the assignment of synsets to the categories *Emotion*, *Moral/Ethic* and *Character*. The cause is that sometimes the glosses of WordNet were not so clear to decide the most correct interpretation. In this case the reviewers have used other dictionaries. The disagreement is widely affected by the classification of the synset in these three categories, as they represent about the 65% of all the synsets of FreeWordNet. Related to the polarity valence of synsets, an agreement near the 100% has been reached because the evaluation of the polarity has been indicated only as positive, negative and objective without any score. At the end about 2.300 pairs of adjectives/synsets and about 480 pairs of adverbs/synsets have been obtained as part of the linguistic resource with a quality and a polarity value associated, as showed in Table 1 and Table 2.

4.1 Measures and Comparisons

In order to evaluate FreeWordNet (FWn), a comparison with SentiWordNet (SWn) and Q-WordNet (QWn) has been performed. But some modifications were required in order to compare the resources. In fact, QWn provides polarity categorization of adjectives and adverbs into one of the two categories, positive and negative, while neutral polarity is seen as the absence of positive or negative polarity. Data about QWn objective terms are not available. So, in the evaluation there was not any consideration about them. SWn assigns three sentiment scores, positivity, negativity and objectivity, which sum is 1. The objectivity is calculated as in (1):

$$\text{obj\_score} = 1 - (\text{pos\_score} + \text{neg\_score}). \qquad (1)$$

Relating to SWn, the biggest score has been selected to determine a unique value of polarity, positive, negative or objective, associated to each synset.

Table 3. The agreement between the three resources.

	FWn-SWn	FWn-QWn	SWn-QWn
Adj.	59,9%	79,8%	54,9%
Adv.	19,3%	100%	27,2%

Table 3 depicts the level of agreement between the three resources, taking into account only the adjectives and adverbs having polarity valence. The evaluation considers the adjectives and adverbs that SWn and QWn have in common with FWn. Table 4 shows the number of adjectives and adverbs in the three resources (tot) and having polarity valence (p/n). Another evaluation of FWn has been made considering a corpus of 100 reviews, composed by 950 sentences, about a hotel of Alghero in Sardinia, as a test set. The syntactic analysis performed on the corpus, produced a list of adjectives and adverbs. Considering their frequency in the corpus the algorithm found 970 adjectives (223 distinct) and 155 adverbs (70 distinct). The corpus has been used in the feature extraction process. Adjectives and adverbs have been disambiguated following the process described in Section 3.1.

Table 4. Adjectives and adverbs having polarity valence.

		FWn	SWn	QWn
Adj	tot	2.268	30.447	
	p/n	1.652	8.909	6.747
Adv	tot	482	5.707	
	p/n	415	409	199

Table 5. Coverage of the resources referred to the synsets identified in the TripAdvisor corpus.

	FWn		SWn		QWn	
Adj-f	76	34%	140	62%	142	63%
Adj+f	617	63%	770	79%	827	85%
Adv-f	12	17%	19	27%	8	11%
Adv+f	22	14%	43	27%	36	23%

These activities produced a set of data useful in order to establish some criterion in the evaluation of FWn. The number of adjectives and adverbs found in the corpus and included in FWn has been compared with the results given by SWn and QWn. Table 5 shows the coverage of the resources referred to adjectives and adverbs disambiguated on the corpus. The values report the number of adjectives and adverbs in term of synsets having polarity valence. In the table, the notation Adj-f denotes the number of distinct adjectives identified in the corpus, without considering their frequency, while Adj+f considers the frequency. The same is valid for the adverbs. In this case the number of adverbs in FWn are less than SWn. The result seems contradict the result about the coverage showed in Table 4, but it is due to a different coverage of the domain in the resources. SWn identifies in the corpus the highest percentage of positive and negative synsets for adverbs (27%) while QWn has the highest value for adjectives (63%). The analysis evidenced that, despite the number of adjectives in FWn is 4-5 times bigger than the other resources, the percentage of adjectives in relation to the corpus is only less than half compared with SWn and QWn. The analysis of the results depicted in Table 5, evidences that FWn and QWn have a different coverage of adjectives and adverbs with a partial overlay. SWn has an almost full coverage of WordNet terms. It is more evident considering the frequency of terms.

5 Conclusions and Future Works

The paper presents FreeWordNet, a linguistic resource of adjectives and adverbs based on WordNet, where each synsets is enriched with a set of properties and polarity values associated. FreeWordNet is mainly involved in the development of an automatic process of feature extraction and especially in the steps of distinction and identification of subjective, objective or factual sentences and contributes in a basic way in the task of contextualization of the features. FreeWordNet is not a finished resource. We are still working on the extension of terms in order to improve the coverage of the resource. The comparison of the measures evidences the need to improve the number of synsets in FreeWordNet, in particular for the adjectives.

Considering that the text analysis is strongly affected by the recognition of adjectives and adverbs having polarity valence, it is evident that the result certainly will benefit by the improvement of the synsets. Anyway, the presence of the set of categories associated to synsets and the polarity values can bring relevant benefit in the analysis of opinions. More in details, the distinction between subjective and factual polarity adjectives and adverbs defined through their categories associated is an implicit capability of FreeWordNet that produce, as a direct result, a relevant element in the recognition and distinction of factual polarity and subjective sentences.

References

1. Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., Subrahmanian, V. S.: Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. In Proceedings of International Conference on Weblogs and Social Media, pp. 203-206 (2007).
2. Rentoumi, V., Giannakopoulos, G.: Sentiment analysis of figurative language using a word sense disambiguation approach. In International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, ACL (2009).
3. Miller, G., A.: WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41 (1995).
4. Lee, D., Jeong, O., Lee, S.: Opinion Mining of customer feedback data on the web. In Proceedings of the 2nd ICUIMC '08 (2008).
5. Wiebe, J., Mihalcea, R.: Word Sense and Subjectivity. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Sydney, Australia (2006).
6. Esuli, A., Sebastiani, F.: PageRanking WordNet synsets: An application to Opinion Mining. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics Vol. 45, Publisher: Association for Computational Linguistics, p. 424-431 (2007).
7. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of LREC-10, 7th Conference on Language Resources and Evaluation, Valletta, MT, pages 2200-2204 (2010).
8. Agerri, R., García-Serrano, A.: Q-WordNet: Extracting polarity from WordNet senses. Seventh Conference on International Language Resources and Evaluation (2010).
9. Valitutti, A., Strapparava, C., Stock, O.: Developing affective lexical ressources. Psychnology: 2 (2004).
10. Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A.: The Role of Domain Information in Word Sense Disambiguation. Natural Language Engineering, special issue on Word Sense Disambiguation, 8(4), pp. 359-373, Cambridge University Press (2002).
11. Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., Gandini, C.: Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. A. Sanso, Language resources and linguistic theory, Franco Angeli, Italy (2007).
12. Angioni, M., Demontis, R., Tuveri, F.: A Semantic Approach for Resource Cataloguing and Query Resolution. Communications of SIWN. Special Issue on Distributed Agent-based Retrieval Tools. (2008).
13. Akkaya, C., Mihalcea, R., Wiebe, J.: Subjectivity Word Sense Disambiguation. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 190–199, Singapore, ACL and AFNLP (2009).
14. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In Proceedings of the International Conference on New Methods in Language Processing, pp. 44-49 (1994).
15. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In Fellbaum, C. (ed.)1998, pp. 265-283 (1998).