Text and Language

Structures · Functions · Interrelations
Quantitative Perspectives

Edited by

Peter Grzybek
Emmerich Kelih
Ján Mačutek

prae
sens

Peter Grzybek
Emmerich Kelih
Ján Mačutek
(eds.)

Advisory Editor
Eric S. Wheeler

Text and Language

Structures · Functions · Interrelations.
Quantitative Perspectives

SONDERDRUCK

prae
sens

Measuring semantic relevance of words in synsets

Ivan Obradović, Cvetana Krstev, Duško Vitas

1 Introduction

When delivering a query to an information retrieval (*IR*) system, a user is typically interested in information related to a particular topic, available in texts stored in electronic form. The result of this query is a selection of texts the *IR* system determines as relevant to the query. The information the user is interested in can generally be expressed in terms of *concepts*, abstract ideas or mental symbols that denote objects in a given category or class of entities, interactions, phenomena, or relationships between them. On the other hand, concepts are lexicalized by one or more synonymous words (simple or compound). For example, the concept of a "housing that someone is living in" is lexicalized by the word "house", but also by "dwelling", "home", "domicile", "abode", "habitation" or "dwelling house". Hence, the concept an IR query pertains to is in practice very often formalized by a Boolean OR combination of words, which the user believes best describe the concept in question, e.g. "house OR home OR domicile".

It goes without saying that the choice of words used in a query is of crucial importance for the relevance of the result delivered by the *IR* system. At first glance, the main problem lies in the fact that the user, when composing a query, might omit some words related to the concept, thus reducing system *recall*, the ratio of the number of relevant texts retrieved to the total number of relevant texts available. A simple query expansion by adding the omitted words would seemingly resolve this problem. However, the expansion of the set of words describing a concept in a query, although contributing to the recall in general, has an adverse effect. Namely, due to the fact that many words are homonymous or polysemous, adding new words to the query might reduce *precision*, the ratio of the number of relevant documents retrieved to the total number of (irrelevant and relevant) documents retrieved. Given this trade-off between recall and precision, words used in a query have to be very carefully selected in order to attain an optimal balance between the two.

Lexical resources such as electronic thesauri, ontologies and wordnets offer various possibilities for automatic or semi-automatic refinement of queries by adding new words to the set of words initially specified by the user. However, as we have already pointed out, this query expansion should not be performed blindly, or else it might seriously jeopardize precision. We argue that measures of *semantic relevance* of a word to a concept this word relates to in a particular language can be defined, and that they should be taken into account in query formulation. This semantic relevance is twofold, based on the following assumptions:

1. Synonymous words, which denote a particular concept, are not used with the same frequency to denote this concept. Hence, they bear different semantic relevance to that concept. For instance, the word “home“ is more frequently used to denote the concept defined as “housing that someone is living in” than the word “abode”, and thus has a greater semantic relevance to this concept.
2. A homonymous or polysemous word, which can be used in more than one sense, to denote totally or partly different concepts, is more frequently used to denote one concept than another. Hence, it bears different semantic relevance to each of them. For example, the word “pen” is more frequently used to denote the concept defined as “a writing instrument which applies ink to a surface, usually paper” than it is used for the concept defined as “an adult female swan”, and thus bears greater semantic relevance to the former.
3. In both cases the semantic relevance of a word to a concept can be quantified. It should be noted, however, that measures of semantic relevance we propose here should be distinguished from the mathematical model for computing the importance of a semantic feature in concept identification (Sartori and Lombardi 2004: 440) and the semantic relevance of a word in a given lexical context (Mattys et al. 2005: 486).

We can now conclude that the selection of words in a query with the aim of establishing an optimal balance between recall and precision in an *IR* system is far from a simple task. In this paper our focus is on wordnets as a means for refining queries in *IR* tasks. We propose a set of very simple and natural relevance indices to be used for tuning the query formulation process.

In Section 2 a brief overview of wordnets and the process of development of the Serbian wordnet are described, in Section 3 we describe the construction and possible use of the indices proposed, and in Section 4 some examples are given, followed by a conclusion in Section 5.

2 The development of Serbian wordnet

Wordnets were conceived in 1985 by George Miller and his associates at Princeton University who started to develop the Princeton WordNet (*PWN*), or simply WordNet, a linguistic database that maps the way the mind stores and uses language. Its aim was to serve as some sort of a mental lexicon that can be used in the scope of psycholinguistic research projects (Fellbaum 1998: 3). *PWN* was conceived as a semantic network of concepts, where each concept is represented by a set of synonymous English word-sense pairs which, accompanied by a definition of the concept, form the synset for this concept. Concepts are interconnected by semantic relations, such as hypernym/hyponym (kind of, e.g. animal/dog) or holonym/meronym (part of, e.g. hand/finger). This database now contains about 150000 words organized in over 115000 synsetsfor a total of 207000 word-sense pairs.

The EuroWordNet project introduced multilingualism into the semantic network of concepts by building wordnets for seven European languages in a manner similar to *PWN*, and aligning them by interconnecting synsets representing the same concept in different languages by an Inter-Lingual-Index, or *ILI* (Vossen 1998: 75). Along the same lines, the BalkaNet project set as its goal the development of aligned semantic networks for Bulgarian, Greek, Romanian, Serbian and Turkish, while at the same time extending the existing network for Czech, initially developed within EuroWordNet (Tufiş et al. 2004: 11). Thirteen scientific and research institutions from Bulgaria, Greece, Romania, Serbia, Turkey, France, the Netherlands and Czech Republic gathered within the project consortium. Six teams were formed, each responsible for the development of a wordnet in one of the six languages. The core of the Serbian team was the Human Language Technologies (HLT) group at the Faculty of Mathematics, University of Belgrade (Krstev et al. 2004: 147).

The initial development of wordnets for the six BalkaNet languages was planned and realized synchronously. Namely, the core of each monolingual wordnet was built from several commonly agreed sets with a total of 8516 concepts selected from *PWN*. Beyond these sets the network for each language has been developed independently, but always within the framework set by *PWN*. This approach generated some specific problems. Namely, during the work on the development of the network the following questions have often been raised: are concepts linguistically independent or not, are the lexicalization patterns for concepts universal, is the structure of *PWN* valid for other languages as well, is the set of semantic relations built in *PWN* sufficient for all languages (Vossen 2004: 5). Although the work on the development of specific networks for Balkan languages often pointed to a negative answer to these questions, the initially established procedure has not been abandoned. The main reason was to preserve the mapping of Balkanet wordnets to *PWN*, thus making them more applicable in multilingual *IR* tasks. After the termination of the BalkaNet project the development of monolingual networks continued, and at present the Serbian wordnet contains more than 25000 words and about 15000 synsets.

Since wordnets represent concepts by means of synsets, they can be used in various ways for tuning user queries to obtain better recall and precision. The most straightforward is the detection of synonymous words omitted in a query which can improve recall. Through semantic relations wordnets also point to closely related concepts, (e.g. more general or more specific), which could also be candidates for query expansion. However, as we have already pointed out, the addition of words from synsets to a query needs to be scrutinized in some way. The relevance parameters we define in the next section could be used as a straightforward assessment mechanism for candidate words offered by a wordnet within a query refinement task.

3 Relevance indices

In order to assess the relevance of each word in a synset for the lexicalization of the concept it is used for, we will now define a set of very simple and natural indices as numerical measures of this relevance. The semantic relevance of words in the *IR* context is best assessed by observing the way they are used in a corpus of written texts for a particular language. Thus we define our indices in direct relation to the occurrences of words in the corpus. Although the proposed indices were tested using Serbian wordnet synsets and the corpus of Serbian written texts, the methodology can be applied to any other language without modification, provided that both the wordnet and a relevant corpus for that language exist. Let **S** be the finite set of all synsets within a wordnet:

$$\mathbf{S} = \{S_i | S_i \text{ is a synset describing a specific concept, } i = 1, 2, \ldots, n_S\} ,$$

where n_S equals the total number of synsets within a wordnet; we shall also denote by $S_i \geq 1$ the total number of words within a nonempty synset S_i. Let **W** be the finite set of all words used as lexicalizations for one or more concepts:

$$\mathbf{W} = \{W_j | W_j \text{ is a word in at least one synset, } j = 1, 2, \ldots, n_W\}$$

where n_W equals the total number of different words in the wordnet. When a word $W_j \in \mathbf{W}$ is used as a lexicalization of a specific concept, described by synset S_i, it is used in a specific sense (a sense tag is attached to it), thus yielding a word-sense pair. We shall denote by $w_j \geq 1$ the total number of senses the word W_j is used in, or words-sense pairs for that word within the wordnet.

As we have already mentioned, we build the numerical parameters of a selected word W_j on the occurrences of this word, together with its inflected forms, in the corpus of written texts. We shall denote the total number of these occurrences of W_j as t_j, and the number of times the word W_j is used for lexicalization of a concept described by synset S_i as c_{ij}. In general, the equation

$$\sum_{i=1}^{w_j} c_{ij} = t_j \tag{1}$$

holds. However, given the fact that the wordnet might be incomplete, namely that all senses the word occurs in within the corpus might not be covered by the wordnet, it is also possible that

$$\sum_{i=1}^{w_j} c_{ij} \leq t_j \,. \tag{2}$$

We need to point out that, simple as it may seem at first glance, the establishing of the number of times the word W_j is used for lexicalization of a concept described by synset S_i, that is c_{ij}, can be a tedious task. Namely, unless the corpus has previously been semantically annotated using wordnet word-sense pair codes, the sense in which a word has been used in the corpus must be established manually. In that case, lexicographers have to be involved to determine the sense a word was used in each occurrence, before the corresponding numbers $c_{ij}(i = 1, 2, \ldots, w_j)$ can be established.

We will now proceed to the definition of two types of indices. As one word may appear in different synsets, we will first construct the indices which express the relevance of a particular word W_j to different synsets the word appears in. The most natural way to construct such an index for a particular synset S_i is to compare the number of occurrences of this word in the corpus denoting the concept represented by synset S_i, that is c_{ij}, to the total number of occurrences of this word within the corpus, namely t_j. Thus we define the *wordnet relevance index* of the word W_j to the synset S_i as the ratio of the number of occurrences where this word has been used to denote the concept represented by the synset S_i and the total number of occurrences of this word in the corpus, namely: $WI_{ij} = c_{ij}/t_j$. It is obvious that the index range is $0 < WI_{ij} \leq 1$, where $WI_{ij} = 1$ holds if the word W_j is used in one and only one sense ($w_j = 1$), and that is to lexicalize the concept described by the synset S_i.

It is easy to prove that the sum of all wordnet relevance indices for a given word W_j is:

$$\sum_{i=1}^{w_j} WI_{ij} \leq 1 \ , \tag{3}$$

where the inequality holds only in the case that all senses the word occurs in within the corpus are not covered by the wordnet. On the other hand, as a synset may be composed of several words, we will now construct an index that expresses the relevance of a particular word W_j within synset S_i in comparison to other words in that synset. In order to construct such an index we need to calculate the total number of occurrences of all words within the corpus which denote the concept represented by synset S_i, namely:

$$a_i = \sum_{j=1}^{s_i} c_{ij} \ . \tag{4}$$

We can now define the ratio of the number of occurrences where the word W_j has been used to denote the concept represented by the synset S_i and the total number of occurrences of all words within the corpus denoting the concept represented by the synset: $SI_{ij} = c_{ij}/a_i$ as the synset relevance index of the word W_j to synset S_i. It should be noted that the range of this index is also $0 < SI_{ij} \leq 1$, where $SI_{ij} = 1$ holds when either synset S_i consists of only one word ($s_i = 1$), and that is word W_j, or other words from that synset have not appeared in the corpus. It is obvious that the sum of synset relevance indices for all words in a given synset S_i is

$$\sum_{j=1}^{s_i} SI_{ij} = 1\ . \tag{5}$$

Let us now take a look at a possible interpretation of the two indices. As we have already pointed out, each new word added to a query as a possible lexicalization of a concept generally increases recall and reduces precision. The indices we defined here can point to the possible impact the addition of a word will have on both recall and precision. They also indicate whether a word is synonymous as well as whether it is homonymous or polysemous.

The wordnet relevance index WI_{ij} clearly indicates whether the word W_j is used in the wordnet in one ($WI_{ij} = 1$) or more senses ($WI_{ij} < 1$), namely whether it is a homonymous or polysemous word or not. Further, for homonymous and polysemous words, it indicates the semantic relevance of the word to different concepts it relates to. Given the fact that all wordnet relevance indices of a word sum to a value less or equal to one, the higher the index for one concept, the lower for all the others. For example, a wordnet relevance index $WI_{ij} > 0.5$ indicates that the word W_j is more closely related to the concept denoted by synset S_i, than to all other concepts it also relates to. The higher the wordnet relevance index of a word, the smaller the impact on precision caused by the addition of this word in a query pertaining to the concept denoted by synset S_i. Simply put, the addition of words with high wordnet relevance indices will not considerably decrease precision. However, this index does not give any information as to the possible effect of the addition of the word W_j on recall.

On the other hand, the synset relevance index SI_{ij} indicates whether the word W_j is synonymous when it relates to the concept denoted by synset S_i. Namely, $SI_{ij} = 1$ means that only the word W_j is used to lexicalize the concept denoted by S_i, whereas $SI_{ij} < 1$ means that the synset S_i contains at least two words. As synset relevance indices for all words in a synset sum to 1, a relevance index $SI_{ij} > 0.5$, indicates that the word W_j is more related to the concept denoted by synset S_i, than all other words within the synset. Adding a word with such a relevance index in a query pertaining to the concept denoted by S_i should considerably raise the recall. On the other hand, the index does not give any information as to the possible effect of the addition of the word W_j on precision.

Hence, the assessment of the effects the addition of a word will have should be made by observing both indices. The “ideal candidate” to be added to a query pertaining to the concept lexicalized by words in synset S_i would be a word W_j from this synset with both a high wordnet and a high synset relevance index. Conversely, a word that has a low value for both indices is a poor candidate and should be omitted in query expansion. If the user him/herself has already inserted the word in the query he/she should be advised to eliminate it.

The two indices can be combined in several different ways. We propose here a *global relevance index* GI_{ij} of the word W_j to the concept denoted by S_i the word belongs to, as a weighted arithmetic mean of the two indices:

$$GI_{ij} = \alpha WI_{ij} + \beta SI_{ij}\,, \tag{6}$$

where $\alpha + \beta = 1$. In case the user cannot decide which is more important, precision or recall, the values of α and β should be both equal to 0.5; if, however, s/he gives priority to recall, the value of β should be raised at the expense of α, whereas if the user is more concerned with precision, then a greater value should be given to α than to β.

We believe that the simple measures of relevance proposed in this section could be of value to the user when deciding which words offered by the wordnet should be considered for query expansion.

Finally, since we have based our approach on the idea of extending a query using a wordnet, we should point out that another index exists that measures the extent to which the wordnet covers all possible senses of a word as indicated by the corpus (Obradović et al. 2004: 183). Namely, due to the fact that all senses of a word that appear in the corpus are not necessarily covered by the wordnet, which we have already mentioned, a *wordnet coverage index* for the word W_j can be defined as the ratio

$$CI_j = \frac{\sum_{i=1}^{w_j} c_{ij}}{t_j}\,. \tag{7}$$

This index does not give any information pertaining to recall or precision but rather the "quality" of the wordnet with respect to word W_j. The index ranges between 0 and 1, and in the case of full coverage is equal to 1.

4 The validation procedure

The proposed approach was validated using the Serbian wordnet and different corpora of Serbian written texts. For validation purposes a set of words that we called pivotal words was chosen among the nouns and verbs that generate the largest number of word-sense pairs in Serbian wordnet. In the next step all synsets in which the pivotal words appeared were analyzed, and the words that appear in these synsets with the pivotal words were identified, and named *supporting words*. The pivotal and supporting words formed the "lexical sample" as defined by the SENSEVAL project (Kilgarriff and Rosenzweig 2000). The main objective of the validation procedure was to assess whether the initial presumptions on the twofold semantic relevance of the words to corresponding concepts, and the relevance indices defined, are supported by experimental data.

The first corpus of approximately 1.7 million words used in the validation procedure consisted of contemporary newspaper texts. Using the available lexical tools concordances were produced for all inflectional forms of both pivotal and supporting words. Since the corpus was not semantically tagged using wordnet word-sense pair codes, the concordances of around 10000 items had to be manually analyzed by lexicographers. The senses of pivotal and supporting words were identified and marked using word-sense pair codes from the Serbian wordnet. Cases where senses of the word were not covered by the wordnet were marked as “other”. On basis of the results obtained indices introduced in Section 2 were calculated. Before proceeding to an analysis of a few examples of relevance indices it should be noted that the wordnet coverage indices pointed out that the coverage of the corpus by the wordnet still varies considerably. Namely, for the words analyzed the wordnet coverage index ranged from 0.246 to 1. Only 3 out of 12 pivotal words that have been chosen had the value of the wordnet coverage index equal to 1, which means that only for these words have all the senses identified in the corpus been included in the Serbian wordnet.

Data for the Serbian noun *lice* and verb *proizvesti* obtained from the newspaper corpus are given in Table 1 and Table 2. The first column is the concept number, the second its definition, and the third the sense in which the pivotal word is used to describe the concept. Column four gives the frequencies of the appearance of the pivotal word in different senses within the corpus and the following columns give the frequencies for supporting words. In the last three columns the total number of occurrences of all words within the corpus which denote the concept is given, followed by the wordnet and synset relevance indices. In the bottom row of the table the total number of occurrences of both the pivotal and supporting words within the corpus is given.

The pivotal word *lice* has eight possible senses, and thus belongs to eight different synsets. In six of them, it is the only synset word, whereas in two of them supporting words *uloga*, *lik* and *strana* also appear. However, in the newspaper corpus this word was identified in only three out of eight possible senses (concepts 1, 2 and 3). Concept 4 was added to the table because of the appearance of the supporting word *strana* in the corpus. Cases when the synset relevance index of a word is 1 are not of great interest for query expansion, since this is the only word denoting the concept and it has to be used in any case. We will thus only point out that data from Table 1 show that *lice* has the greatest ordnet relevance index for concept 2. However, it is interesting to observe the effect of this word to queries pertaining to concepts 3 and 4. Both of its indices for concept 4 are 0, which means that adding this word to a query pertaining to this concept is not advisable, since it would not improve

Table 1: Relevance indices for the word *lice* obtained from newspaper corpus

	Concept	Sense	c_{ij}	uloga	lik	strana	a_i	WI_{ij}	SI_{ij}
1	The front of the human head	1a	33	*	*	*	33	0.063	1.000
2	A part of a person that is used to refer to a person	2a	353	*	*	*	353	0.675	1.00
3	An actor's portrayal of someone in a play	2b	1	34	3	*	38	0.002	0.026
4	A surface forming part of the outside of an object	5a	0	*	*	5	5	0.000	0.000
	Other		136						
		t_j	523	208	20	861			

recall and would have a detrimental effect on precision. The same is basically true for concept 3, since both indices are also very low. Finally, the wordnet coverage index for *lice* is $CI_j = 0.740$, which indicates that around 26% of the meanings of this word are not yet covered by the wordnet.

As for the pivotal word *proizvesti*, its wordnet coverage index $CI_j = 0.985$, which means that less than 2% of the meanings of this word are not covered by the wordnet. Table 2 indicates that this word has the greatest wordnet relevance index to concept 3, with the corresponding synset relevance index being moderately low. However, expanding the query pertaining to concept 3 with this word could be recommended: recall should be moderately raised, but precision should not be significantly affected.

In order to test the impact of the nature of the corpus to the values of relevance indices an additional validation was performed on a small literary corpus of 0.5 million words for a selected set of words. As indicated by Table 3, showing data for the word *lice* obtained from the literary corpus, index values can be largely affected by the nature of the corpus. Thus, for example, the wordnet relevance index of the noun *lice* has dramatically changed for senses 1a and 2a. This does not come as too much of a surprise since the concept that the meaning 2a refers to is more used in newspaper texts, whereas the concept that the meaning 1a refers to is more a literary concept. The changes seem to be far less dramatic for the synset relevance indices, but in order to draw some final conclusions, the impact of the nature of the corpus on relevance indices should be more systematically tested on larger corpora.

In general, the order of words within a synset is arbitrary. However, once the indices are calculated, they provide for an ordering of words in the synset.

Table 2: Relevance indices for the word *proizvesti* obtained from newspaper corpus

	Concept	Sense	c_{ij}	prouzrokovati	potaknuti	iznedriti	proizvoditi	napraviti	a_i	WI_{ij}	SI_{ij}
1	Cause to occur or exist	1a	6	31	1	*	*	*	38	0.09	0.16
2	Be the cause or source of	1b	1	*	*	0	*	*	1	0.02	1
3	Create or manufacture a man-made product	3	59	*	*	*	106	21	186	0.88	0.32
	Other		1								
		t_j	67	31	1	99	114	159			

Several possibilities exist, but a natural ordering would be in decreasing order of the global relevance index with parameters α and β chosen according to the preferences of the user. In order to optimize query expansion, the candidate words for expansion could then be offered to the user in this order.

Table 3: Relevance indices for the word *lice* obtained from newspaper corpus

	Concept	Sense	c_{ij}	uloga	lik	strana	a_i	WI_{ij}	SI_{ij}
1	The front of the human head	1a	380	*	*	*	380	0.936	1
2	A part of a person that is used to refer to a person	2a	3	*	*	*	3	0.007	1
3	An actor's portrayal of someone in a play	2b	3	6	1	*	10	0.007	0.300
4	A surface forming part of the outside of an object	5a	2	*	*	4	6	0.005	0.333
	Other		18						
		t_j	406	22	25	287			

Besides query expansion, the indices defined in this paper can also be used for wordnet refinement. Namely, if the value of the synset relevance index SI_{ij}

for the word W_j is close to 0, it can indicate that the word has been misplaced in synset S_i, especially in the case when at the same time both its total occurrence in the corpus t_j and the total number of occurrences of all words within the corpus which denote the concept represented by synset S_i, namely a_i, are considerably greater than 0. For instance, that could be the case for the word *napraviti* in the synset denoting concept 3 in Table 2. The total number of occurrences of the word *napraviti* is relatively big ($t_j = 159$) and the total number of occurrences of all words within the corpus in the synset denoting concept 3 is also considerably high ($a_i = 186$). However, if the synset relevance index for *napraviti* is calculated for the synset denoting concept 3, a relatively low value ($SI_{ij} = 0.113$) is obtained. Thus, the synonymy of the word *napraviti* with the pivotal word *proizvesti* should be reconsidered.

5 Conclusion

The wordnet and synset relevance indices proposed in this paper as a measure for semantic relevance of a word to a concept the word denotes have been applied on a small sample of chosen words and corpora for validation purposes. The results obtained indicate that the rationale for their definition rests on solid grounds. However, further analysis and testing on larger and balanced corpora are needed for their proper assessment. The problem within the validation procedure is the determination of senses a word is used in the corpus. Namely, a prerequisite for this validation is the tagging of the words in the corpus with senses used in the wordnet. To that end, automatic or semi-automatic procedures are needed in order to alleviate the time-consuming task of manual sense assignment. The indices can be useful in query expansion for determining the impact of the addition of a word on the precision and recall of the query. The calculation and assignment of indices should be focused on the most frequently used words in the corpus in the initial phase. The sensitivity of indices to the type of texts they are drawn from has been noted, but it also needs further investigation. Relevance indices can be used for wordnet refinement as well, since the determination of synsets for a given concept is not always a simple task.

References

Fellbaum, C.
1998 “Introduction”. In: Fellbaum, C. (ed.), *WordNet: An Electronic Lexical Database.* Cambridge, Mass.: MIT Press, 1–19.

Kilgarriff, A.; Rosenzweig, J.
2000 “English SENSEVAL: Report and Results”. In: *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC-2000.* Athens, 1239–1244.

Krstev, C.; Pavlović-Lažetić, G.; Obradović, I.; Vitas, D.
2004 “Using Textual and Lexical Resources in Developing Serbian Wordnet”, in: *Romanian Journal of Information Science and Technology*, 7/1-2; 147–161.

Mattys, S.L.; White, L.; Melhorn, J.F.
2005 “Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework”, in: *Journal of Experimental Psychology: General*, 134/4; 477–500.

Sartori, G.; Lombardi, L.
2004 “Semantic relevance and semantic disorders”, in: *Journal of Cognitive Neuroscience*, 16/3; 439–452.

Obradović, I.; Krstev, C.; Pavlović-Lažetić, G.; Vitas, D.
2004 “Corpus Based Validation of Wordnet Using Frequency Parameters”. In: Sojka, P.; Pala, K.; Smrz P.; Fellbaum C.; Vossen P. (eds.), *Proceedings of the Second International WordNet Conference, GWC 2004.* Brno: Masaryk University, 181–186.

Tufiş, D.; Cristea, D.; Stamou, S.
2004 “BalkaNet: Aims, Methods, Results and Perspectives. A General Overview”, in: *Romanian Journal of Information Science and Technology*, 7/1-2; 9–43.

Vossen, P.
1998 “Introduction to EuroWordNet”, in: *Computers and the Humanities*, 32/2-3; 73–89.

Vossen, P.
2004 “Introduction to the Special Issue on the BalkaNet Project”, in: *Romanian Journal of Information Science and Technology*, 7/1-2; 5–6.

Contents