Procesamiento del Lenguaje Natural, Revista nº 49 septiembre de 2012, pp 129-138 recibido 05-05-12 revisado 31-05-12 aceptado 05-06-12

Labeling Semantically Motivated Clusters of Verbal Relations

Etiquetado de clusters de relaciones verbales motivados semanticamente

Gabriela Ferraro
Dept. of Information and
Communication Technologies,
Universitat Pompeu Fabra
gabriela.ferraro@upf.edu

Leo Wanner
ICREA and Dept. of Information and
Communication Technologies,
Universitat Pompeu Fabra
leo.wanner@upf.edu

Resumen: El clustering de documentos es un campo de investigación popular en los ámbitos del Procesamiento del Lenguaje Natural, la Minería de Datos y la Recuperación de información (RI). El problema de agrupar unidades léxicas mediante clustering ha sido menos estudiado y menos aún, el problema de etiquetar los clusters. Sin embargo, en nuestra aplicación que trata sobre la extracción de tuplas de relaciones para ser usadas como entrada a programas para dibujar diagramas de bloques o mapas conceptuales, este problema es fundamental. La valoración de varias estrategias de etiquetado de clusters de documentos nos revela que algunas de estas técnicas pueden ser también aplicadas para etiquetar nuestros clusters, compuestos por verbos semánticamete similares. Para confirmar esta suposición, llevamos a cabo una serie de experimentos y evaluamos su rendimiento contra *baselines* y un *goldstandard* de clusters etiquetados.
Palabras clave: Etiquetado de clusters, clustering, clasificación de relaciones

Abstract: Document clustering is a popular research field in Natural Language Processing, Data Mining and Information Retrieval. The problem of lexical unit (LU) clustering has been less addressed, and even less so the problem of labeling LU clusters. However, in our application that deals with the distillation of relational tuples from patent claims as input to block diagram or a concept map drawing programs, this problem is central. The assessment of various document cluster labeling techniques lets us assume that despite some significant differences that need to be taken into account some of these techniques may also be applied to verbal relation cluster labeling we are concerned with. To confirm this assumption, we carry out a number of experiments and evaluate their outcome against baselines and gold standard labeled clusters.
Keywords: cluster labeling, clustering, relation classification

1. Introduction

Clustering is a popular field of research in Natural Language Processing, Data Mining and Information Retrieval. Most often, the goal is to group the documents in a given document collection with respect to their semantic similarity (Hearst and Pedersen, 1996; Zhu et al., 2006). Some works also address the problem of grouping lexical units (LUs) according to specific semantic criteria. For instance, (Yang and Powers, 2005) group object nouns with respect to their proximity in a taxonomy. According to their approach, *peach*, *pear*, *apricot*, *strawberry*, *banana*, *melon*, etc. form a single cluster and so do *birch*, *fir*, *oak*, etc. (Sekine, 2005; Schulte im Walde, 2006; Korhonen, Krymolowski, and Collier, 2006; Davidov and Rappoport, 2008) cluster verbal relations into classes such as {*compress*, *reduce*, *minimize*, *trim*, *cut*, etc.}, again in accordance with predefined semantic criteria. However, surprisingly little work has been done so far on labeling the obtained LU clusters; the few proposals made on cluster labeling at all nearly exclusively refer to document clusters. This is despite the fact that an ideal cluster label not only reflects the semantic commonalities shared by all members of a given cluster, but also uniquely differentiates this cluster from other clusters in the collection. It could thus be used in any term generalization task.

In our application, we face such a term ge-

ISSN 1135-5948

neralization task. We aim at distilling relational tuples from functional descriptions such as patent claims in order to provide input to block diagram or a concept map drawing programs. This implies that we directly face the problems of verbal relation clustering and relation cluster labeling.[1] A straightforward use of verbal relation names extracted from a functional description (as, e.g., *comprise* between *automatic focusing device* and *an objective lens* or *include* between *astigmatic optical system* and *optical element*) as done, e.g., by (Cascini and Russo, 2007), is not appropriate: block diagrams and concept maps are conceptual representations. They must achieve a sufficient abstraction over concrete terms. Thus, *comprise*, *include*, *contain*, *have*, etc. are sufficiently similar to be considered the same relation in a block diagram and thus should be captured by the same concept—for instance, 'part-of' and named the same. In the same vein, *cause*, *lead to*, *result in*, etc. should be captured by a single concept—'cause'.

In what follows, we focus on the problem of verbal relation cluster labeling. Section 2 describes the problem of cluster labeling in general. Section 3 outlines the verbal relation cluster labeling experiments we carried out to assess how labeling strategies inspired by document cluster labeling perform on the LU cluster labeling task, and Section 4 presents the evaluation of these experiments. In Section 5, we summarize the related work on cluster labeling. In Section 6, we draw some conclusions from our experiences and outline some lines of future work along which the cluster labeling strategies we experimented with can be improved.

2. The problem of cluster labeling

As already mentioned in Section 1, cluster labeling proposals focused so far mainly on document cluster labeling. Document clustering is a key technique in cluster-based search, *scatter-gather*-based document browsing, opinion mining, data mining, etc. (Muresan and Harper, 2004; Pirolli, 2007). Cluster labeling is used in connection with clustering to make the results of clustering more transparent to the user (Osinski and Weiss, 2005; Mika, 2005). As cluster labels, sentences, phrases or simply lists of terms that are assumed to characterize well the clusters in question are taken.

In the clustering literature, two main strategies of document cluster labeling can be identified: (i) internal cluster labeling and (ii) differential cluster labeling. In internal cluster labeling, the label of a given cluster is chosen drawing solely on the content of the cluster itself. For instance, (Chen and Liu, 2004; Cutting, Karger, and Pedersen, 1993) suggest to pick as label a linguistic construction or a sequence thereof (e.g., the title of one of the documents in the cluster, a list of terms, a phrase, etc.) that proves to be closest to the cluster's *centroid* according to measures such as cosine. (Cutting et al., 1992; Osinski and Weiss, 2005) propose frequency-based internal labeling strategies which select as label the term or a list of terms that are most frequent in the given cluster. Internal labeling strategies have the advantage of being simple.

In differential labeling, the label of a cluster is chosen by contrasting this cluster with the other available clusters. Often, statistical measures such as Mutual Information (MI), Information Gain and the χ^2 test are applied, which calculate the statistical dependence of a candidate label on the cluster in question (relatively to the other clusters in the collection): if the candidate label is dependent on the cluster (more than on the other clusters), it is considered a good label for it; see, e.g., (Pellegrini, Maggini, and Sebastiani, 2006; Carmel, Roitman, and Zwerdling, 2009).

Our task is different from the task of document cluster labeling. As already mentioned, we face the problem of labeling clusters of semantically similar verbal LUs such that they can be used as relation labels in a conceptual map-like representation. In such a setting, we need to choose as label a single lexical element (a phrase or a list of terms are not appropriate). Still, the general ideas underlying the internal and differential clustering strategies seem to stay valid: we can choose a label of a cluster either by drawing solely on the members of this cluster or by exploring the influence of the other clusters as well.

When choosing the label, we can either pick one of the lexemes of the cluster in question or choose an abstract label that captu-

[1] Obviously, we also face the problem of relation extraction. However, we cannot delve into this topic here. Interested readers are asked to consult (Ferraro and Wanner, 2011).

res, in a sense, all members of the cluster. To obtain the most suitable abstract term in a cluster, we can look up the members of the cluster in a taxonomy or to look for a common hyperonym among the members using an external resource such as Wordnet (WN). The problem with using WN might be that in contrast to its nominal hierarchies which tend to be rich and deep, the verbal hierarchies in WN are relatively flat and poor.

Another possibility is to enrich clusters by lexemes retrieved from thesauri since thesauri group lexemes according to their similarity of meaning. Our intuition is that enriching clusters by semantically related lexemes retrieved from a thesaurus increases the possibility of finding a common abstract label. This intuition is based on two observations: (i) we can look for the most frequent common thesaurus term in the clusters, avoiding the restriction of assigning as label a lexeme from the cluster itself, (ii) we can further apply statistical tests, such as Mutual Information, which are best suited for clusters that contain overlapping terms.

In what follows, we carry out a number of experiments in order to assess to what extent the possible approaches sketched above lead to successful verbal relation cluster labeling, and to be able to choose the best one for our applications.

3. Cluster labeling experiments

The input to all cluster labeling strategies described in this section is a set of verb clusters, which have been grouped automatically according to their semantic similarity in a previous step of our application. We have experimented with seven different cluster labeling strategies. Three strategies are internal cluster labeling techniques and four are differential cluster labeling techniques.

3.1. Internal cluster labeling

We experimented with the following internal cluster labeling strategies:

Frequency-oriented labeling (Freq): Choose as cluster label the member of the cluster with the highest frequency in the reference corpus. This strategy is motivated by classic cluster labeling techniques that choose one of the members of the cluster as its label (Osinski and Weiss, 2005; Chen and Liu, 2004). It has the advantage of being simple. Thus, for the cluster:

> C_i = {*bound:63, limit:74, restrain:21, inhibit:101, fasten:49, fix:53, secure:13, lock:28*}

this strategy suggests *inhibit* as cluster label (the suffix ':X' denotes the frequency of the corresponding member in the reference corpus).

Verb hyperonym-oriented labeling (VHyp): Choose as cluster label the most frequent hyperonym of the cluster as it appears in the WN verb hierarchy. To implemente this strategy, first, for each member of a cluster, all its WN hyperonyms are retrieved and the most frequent hyperonym synset is selected. Then, from this synset, the most frequent lexeme in the corpus is chosen as the cluster label. For example, for the above cluster, the most frequent hyperonym synset is:

> C_i(*hyper*) = {*bound3, check4, confine1, limit1, restrain2, restrict3, throttle1, trammel2, decide1, decide upon1, determine4, make-up one's mind1* }

From this hyperonym synset, the most frequent hyperonym found in the reference corpus is *limit:1* (in this case, the suffix ':X' stands for the WN sense). Therefore, *limit* is chosen as the cluster label. This strategy is motivated by the fact that the cluster label should be more abstract to ensure that it captures all members of the cluster.

Thesaurus Freq (ThesFreq): Choose as cluster label the most frequent lexeme found in a cluster populated by LUs from the Open-Office Thesaurus. To populate a cluster by LUs from the thesaurus, for each of the members of the cluster, the verbal lexemes related to it via the different semantic relations are retrieved from the thesaurus. For instance, the following verbal lexemes are associated with the member *lock* of the cluster C_i introduced above:

> *thesaurus(lock)* = {*fasten, fix, secure, lock up, lock up, engage, mesh, operate, move, displace, engage, interlock, interlace, hold, take hold, interlock, embrace, hug, bosom, squeeze, overwhelm, overpower, sweep over, whelm, overcome, overtake, lock in, lock away, put away, shut up, shut away, lock up, confine, pass, go through, go across, construct, build, make*}

The most frequent among them is *fix*. It is thus chosen as cluster label.

Cuadro 1: Examples of the performance of the internal cluster labeling strategies

Gold Standard Clusters	GS	Freq	VHyper	ThesFreq
{comprise, contain, have, include}	contain	comprise	comprise	get
{bound, limit, restrain, inhibit, fasten, fix, secure, lock}	limit	inhibit	limit	fix
{compress, trim, reduce, minimize}	reduce	reduce	cut	lessen
{extract, pull-out}	extract	extract	remove	take-out
{remove, cut, delete, erase, exclude}	remove	remove	remove	take-out
{enter, insert, interpose, introduce, enclose}	insert	insert	connect	introduce
{apply, feed, provide, give, use, supply, render}	produce	provide	provide	give
{hold , maintain, retain, support, prevent}	keep	support	maintain	hold
{accord, allow, let, permit}	let	accord	have	permit

Table 1 displays a sample of the results of the application of the internal cluster labeling strategies to a number of gold standard clusters.

3.2. Differential cluster labeling

The differential cluster labeling strategies have been implemented using the MI and the χ^2 measures.

VHyp MI-oriented labeling (VHyp-MI): Choose as cluster label the hyperonym with the highest MI value. First, for each member of the cluster, its WN hyperonyms are retrieved (as already in the internal VHyp-oriented labeling strategy). Then, for a given cluster, we calculate the MI of each hyperonym and select as label that hyperonym which shows the highest MI value. Consider, for illustration, Table 2, where the MI values of the label candidates for the cluster C_i are displayed.

The hyperonym with the highest MI value turns out to be *moderate*. Therefore, it is chosen as the cluster label.

Cuadro 2: Examples of the MI values of the candidate labels for C_i

Label candidate	MI value
moderate	1391.80
restrict	1394.56
throttle	1394.56
restrain	1389.33
put restrictions on	1388.25
check	1387.05

VHyp χ^2-oriented labeling (VHyp-χ^2): Choose as cluster label the hyperonym with the highest χ^2 value. The procedure is the same as above, only that instead of MI, the χ^2 measure is applied. Table 3 displays the χ^2 values for the different candidate labels for C_i. Since *throttle* shows the highest χ^2 value, it is chosen as label.

Cuadro 3: Examples of the χ^2 values of the candidate labels for C_i

Label candidate	χ^2 value
throttle	90.30
confine	82.29
hold-in	76.31
restrain	65.83
check	57.68
fasten with a lock	34.82

Thesaurus MI-oriented labeling (ThesMI): Choose as cluster label the thesaurus lexeme with the highest MI value. The clusters are populated with the thesaurus matches as in the ThesFreq strategy. An example of the experiment run is shown in Table 4. For the cluster C_i, the term with the highest MI value is *restrict*.

Cuadro 4: Examples of the Thesaurus-MI values for the candidate labels for C_i

Label candidate	MI value
restrict	1392.68
interlock	1388.25
stick	1384.86
tie	1377.48
fix	1374.42
lessen	1374.34

Thesaurus χ^2-oriented labeling

(Thesχ^2): Choose as cluster label the thesaurus lexeme with the highest χ^2 value. Again, the clusters are populated with its thesaurus matches as in the ThesFreq strategy. An example of the results of an experiment run is shown in Table 5. For C_i, the term with the highest χ^2 value is *restrict*.

Cuadro 5: Examples of the Thesaurus χ^2 values for some candidate labels for C_i

Label candidate	**χ^2 value**
restrict	76.91
trammel	76.90
curb	76.31
hold in	65.83
control	34.48
fasten	20.70

Table 6 presents some examples of the performance of the application of the differential cluster labeling strategies.

3.3. Fallback strategies

Sometimes, differential labeling strategies come up with several candidate labels with the same weight. Since we have to decide which of them to choose, we use two different simple *fallback strategies*. The first of them chooses as label the candidate with the highest frequency in the reference corpus. The second picks the label randomly among the candidates with the same weight.

4. Evaluation

We carried out a qualitative evaluation of the implemented cluster labeling strategies, resorting to human judges. For the evaluation, we use a gold standard of 54 verb clusters as the list of clusters to name. The 54 clusters were presented to three judges, together with the labels assigned to each of the clusters by our system and by a human collaborator (the gold standard labels), such that the judges did not know the origin of a label. For each cluster, the judges were asked to qualify all the labels as 'correct' (corr), 'partially correct' (pcorr) or 'incorrect' (incorr). Table 7 shows the evaluation results of the internal labeling strategies and Table 8 the results of differential labeling.

Table 7 reveals that the Freq strategy, which chooses as the label of a cluster its most frequent member, achieves with 78 % of

Cuadro 7: Internal clustering labeling strategies evaluation.

	% Corr	**% Pcorr**	**% Incorr**
Gold st.	77 %	17 %	7 %
Freq	78 %	20 %	2 %
VHyp	43 %	25 %	32 %
ThesFreq	58 %	26 %	16 %

correctness the best results. This is somewhat surprising since one would expect that a label that abstracts over the individual members of a cluster would be more appropriate. However, the VHyp strategy shows significantly worse results than Freq, achieving only 43 % of correctness. We assume that this is largely because the hyperonyms in WN tend to be too abstract to serve as a label of their hyponyms—as is, e.g., also the case with *move* for the cluster {*disperse*, *propagate*}. The ThesFreq strategy shows acceptable results, achieving a 58 % of correctness and 26 % of partially correctness. The weakness of the ThesFreq strategy is that it uses all semantic relations in the thesaurus to retrieve candidate labels. The use of synonymy and hyperonymy only appears more promising and will be tested in the future. A baseline strategy that arbitrarily chooses a member of a given cluster as the label of this cluster reaches a 31 % match with the gold standard labels.

Cuadro 8: Differential clustering labeling strategies evaluation.

	%Corr	**%Pcorr**	**%Incorr**
Gold st.	77 %	17 %	7 %
VHyp-MI	50 %	45 %	5 %
VHyp χ^2	60 %	27 %	13 %
ThesMI	70 %	25 %	5 %
Thes χ^2	67 %	22 %	11 %

Table 8 shows the results of the differential cluster labeling strategies. As in internal labeling strategies, the strategies that used the thesaurus perform better than the ones that use verb hyperonyms from WN. The best differential strategy is ThesMI, achieving a 70 % of correctness. The Thesχ^2 strategy has a slightly lower score, achieving a 67 % of correctness.

Cuadro 6: Examples of the performance of the differential cluster labeling strategies

Gold Standard Clusters	GS	VHyp-MI	VHyp χ^2	ThesMI	Thesχ^2
{comprise, contain, have, include}	contain	comprise	incorporate	incorporate	incorporate
{bound, limit, restrain, inhibit, fasten, fix, secure, lock}	limit	moderate	throttle	restrict	restrict
{compress, trim, reduce, minimize}	reduce	trim down	thin-out	find-out	minify
{extract, pull-out}	extract	move forcibly	pull-up	pull-up	press-out
{remove, cut, delete, erase, exclude}	remove	erase	kill	cancel	take-out
{enter, insert, interpose, introduce, enclose}	insert	shut-it	enclose	pull-in	pull-in
{apply, feed, provide, give, use, supply, render}	produce	administer	furnish	furnish	furnish
{hold , maintain, retain, support, prevent}	keep	hold on	hold on to	defend	defend
{accord, allow, let, permit}	let	grant	grant	consent	consent

According to the qualitative evaluation, the performance of one of the internal cluster labeling strategies, namely 'Freq', is the one that is most similar to the performance of our human judge, while the strategies that are based on WN hyperonyms, perform significantly poorer—although in the literature, WN hyperonym hierarchies are most commonly used for lexical labeling. This is partly due to the fact that most of the works on lexical labeling target the labeling of nominal rather than verbal clusters and WN, which is used as reference resource, has, in general, very flat verbal hierarchies.

Differential strategies that use the thesaurus as an external resource show competitive results, as they are close to the human judgements. A weakness of differential labeling is that sometimes labels are low frequency terms and appear somewhat questionable. For instance, in the ThesMI strategy, the cluster {*become*, *convert*, *turn*} is labeled by the term *metamorphose*, which is judged as 'partially correct'. Even if this term reflects the meaning of the cluster it is considered inappropriate to be used as cluster label in the technical domains of our corpus.

5. *Related work*

Although the focus of cluster labeling research has been on document cluster labeling, some proposals exist also for LU cluster labeling. In what follows, we focus on those proposals. Thus, the proposal by (Pantel and Ravichandran, 2004), which addresses the problem of labeling clusters of semantically similar nouns, is an example for internal cluster labeling. The input of their system are semantic classes (clusters of nouns) and the output is a ranked list of label names for each semantic class. First, for each member of a cluster, grammatical signatures that capture its prototypical semantic context in different occurrences are computed. In other words, each word of a cluster is represented by a feature vector where each feature corresponds to a context in which the word occurs. As context, the grammatical functions (such as *subject*, *direct object*, etc.) computed by the Minipar (Klein and Manning, 2003) parser are used. For example, "catch —" represents a verb object context. If the word *wave* would occur in this context, the context would thus include the feature of *wave*. Then, among these signatures, simple hyperonymy patterns, such as "Noun–apposition–Noun" (e.g., H1N1, the disease) are searched. At last, the mutual information scores for each hyperonymy candidate are calculated and the highest scoring hyponym is chosen as the name of the cluster. Further similar proposals of internal labeling include (Carmel, Roitman, and Zwerdling, 2009; Manning, Raghavan, and Schütze, 2008).

The proposal by (Dias et al., 2009) is, in principle, a proposal on document cluster labeling because it addresses the problem of clustering of webpage results and the subsequent labeling of the obtained clusters. However, since it chooses as label of a given cluster a noun or a noun compound it is worth to be mentioned here. It is an example for differential cluster labeling in that the chosen label (i) occurs in most of the URLs of the cluster in question, (ii) discriminates the cluster sufficiently well from the other clusters.

The more complex problem of labeling nodes in a hierarchy (which requires distinguishing more general labels for parents from more specific labels for children) is tackled by (Glover et al., 2002) and (Treeratpituk and Callan, 2006). Some clustering algorithms attempt to find a set of labels first and then build (often overlapping) clusters around the labels; see, e.g., (Osinski and Weiss, 2005; Zamir and Etzioni, 1999; Mika, 2005)—even if, as pointed out by (Manning, Raghavan, and Schütze, 2008), no comprehensive study that compares the quality of such *label-based* clustering with the classic clustering algorithms is known.

As far as labeling clusters of similar verbs is concerned, i.e., the problem addressed in this paper, to the best of our knowledge, no work has been dedicated to this problem as yet.

6. Conclusions and future work

In the context of semantic verb clustering, differential labeling strategies seem more suitable since they take into account the panorama of all clusters. This is coherent with the evaluation results obtained so far: differential labeling strategies outperform nearly all internal labeling strategies; the exception is the internal labeling based on frequency, which performs better.

The results also shows that internal cluster labeling strategies are efficient, but since they do not distinguish terms that are frequent in the collection in general from those that are frequent only in the cluster, they may assign the same label to more then one cluster. With respect to differential labeling, we need to take into account that very low frequency terms should be omitted as label candidates as they would not be the best in representing a whole cluster. So far, we did not apply any frequency filters, such that all terms are considered as label candidates. In the future, we plan to experiment with a hybrid labeling technique that combines internal and differential methods and to take the context of the verbal relations into account. Furthermore, we plan to experiment with other external lexical resources for enriching clusters for the purpose of labeling—among them, synonym dictionaries. Some work has been done in the past on grouping semantically similar nouns and semantically similar adjectives (Rooth et al., 1999; Boleda, Schulte im Walde, and Badia, 2008). Given that verb nominalizations and adjectives are also frequently used in patent claims, both word categories need to be considered in our future work as well.

Bibliografía

Boleda, G., S. Schulte im Walde, and T. Badia. 2008. An analysis of human judgements on semantic classification of catalan adjectives. *Research on Language and Computation*, 6:247–271.

Carmel, D., H. Roitman, and N. Zwerdling. 2009. Enhancing cluster labeling using wikipedia. In *Proceedings of the 32nd international ACM SIGIR Conference*, SIGIR '09, pages 139–146, New York, NY, USA. ACM.

Cascini, G. and D. Russo. 2007. Computer-aided analysis of patents and search for triz contradictions. *International Journal of Product Development*, 4(1):52–67.

Chen, Keke and Ling Liu. 2004. Clustermap: Labeling clusters in large datasets via visualization. In *Proc. of ACM Conf. on Information and Knowledge Mgt. (CIKM*, pages 285–293.

Cutting, D. R., D. R. Karger, and J. O. Pedersen. 1993. Constant interaction-time scatter/gather browsing of very large document collections. In *Proceedings of the 16th annual international ACM SIGIR Conference*, SIGIR '93, pages 126–134. ACM.

Cutting, D. R., J. O. Pedersen, D. Karger, and J. W. Tukey. 1992. Scatter/gather: a cluster-based approach to browsing large document collections. In *Proceedings of the 15th annual international ACM SIGIR Conference*, SIGIR '92, pages 318–329. ACM.

Davidov, D. and A. Rappoport. 2008. Unsupervised discovery of generic relationships using pattern clusters and its evaluation by automatically generated sat analogy questions. In *Meeting of the Association* *for Computational Linguistics*, pages 692–700.

Dias, G., S. Pais, F. Cunha, H. Costa, H. Machado, T. Barbosa, and B. Martins. 2009. Hierarchical soft clustering and automatic text summarization for accessing the web on mobile devices for visually impaired people. In *Proceedings of the FLAIRS Conference*, pages 231–236.

Ferraro, G. and L. Wanner. 2011. Towards the derivation of verbal content relations from patent claims using deep syntactic structures. *Knowledge-Based Systems*, 24:1233 – 1244.

Glover, E. J., K. Tsioutsiouliklis, S. Lawrence, D. M. Pennock, and G. W. Flake. 2002. Using web structure for classifying and describing web pages. In *Proceedings of the 11th international conference on World Wide Web*, pages 562–569. ACM.

Hearst, M. A. and J. O. Pedersen. 1996. Reexamining the cluster hypothesis: scatter/gather on retrieval results. In *Proceedings of the 19th annual international ACM SIGIR Conference*, SIGIR '96, pages 76–84, New York, NY, USA. ACM.

Klein, D. and C. Manning. 2003. Accurate unlexicalized parsing. *Proceedings of the 41st meeting of the Association for Computational Linguistics.*

Korhonen, A., Y. Krymolowski, and N. Collier. 2006. Automatic classification of verbs in biomedical texts. In *Proceedings of the 21st ACL*, ACL-44, pages 345–352. Association for Computational Linguistics.

Manning, C, P Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval.* Cambridge University Press, Cambridge.

Mika, K. 2005. Findex: Search Result Categories Help Users when Document Ranking Fails. In *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 131–140, New York, NY, USA. ACM.

Muresan, G. and D. J. Harper. 2004. Topic modeling for mediated access to very large document collections. *Journal of the American Society for Information Science and Technology*, 55:892–910.

Osinski, S. and D. Weiss. 2005. A concept-driven algorithm for clustering search results. *IEEE Intelligent Systems*, 20:48–54, May.

Pantel, P. and D. Ravichandran. 2004. Automatically labeling semantic classes. In *HLT - NAACL*, pages 321–328.

Pellegrini, M., M. Maggini, and F. Sebastiani. 2006. M.: Cluster generation and cluster labelling for web snippets: A fast and accurate hierarchical solution. Technical report, In Proceedings of the 13th SPIRE 2006.

Pirolli, P. 2007. *Information Foraging Theory: Adaptive Interaction with Information.* Oxford University Press.

Rooth, M., S. Riezler, D. Prescher, G. Carroll, and F. Beil. 1999. Inducing a semantically annotated lexicon via em-based clustering. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 104–111. Association for Computational Linguistics.

Schulte im Walde, S. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, 32(2):159–194.

Sekine, S. 2005. Automatic paraphrase discovery based on context and keywords between ne pairs. In *International Workshop on Paraphrase.*

Treeratpituk, P. and J. Callan. 2006. Automatically labeling hierarchical clusters. In *Proceedings of the 2006 international conference on Digital government research*, dg.o '06, pages 167–176, New York, NY, USA. ACM.

Yang, D. and D. Powers. 2005. Measuring semantic similarity in the taxonomy of wordnet. In *ACSC '05: Proceedings of the Twenty-eighth Australasian conference on Computer Science*, pages 315–322. Australian Computer Society, Inc.

Zamir, O. and O. Etzioni. 1999. Grouper: A dynamic clustering interface to web search results. pages 1361–1374.

Zhu, Y. H., G. Z. Dai, B. C. M. Fung, and D. J. Mu. 2006. Document clustering method based on frequent co-occurring words. In *Proc. of the 20th Pacific Asia Conference on Language, Information and Computation (PACLIC)*, pages 442–445, Wuhan, China, November.