Comparison of clustering approaches through their application to pharmacovigilance terms

Marie Dupuch[1], Christopher Engström[2], Sergei Silvestrov[2], Thierry Hamon[3], Natalia Grabar[1]

(1) CNRS UMR 8163 STL, Université Lille 3, 59653 Villeneuve d'Ascq, France
(2) Division of Applied Mathematics, Mälardalen University, Västerås, Sweden
(3) LIM&BIO (EA3969), Université Paris 13, Sorbonne Paris Cité, France

Abstract. In different applications (*i.e.*, information retrieval, filtering or analysis), it is useful to detect similar terms and to provide the possibility to use them jointly. Clustering of terms is one of the methods which can be exploited for this. In our study, we propose to test three methods dedicated to the clustering of terms (hierarchical ascendant classification, *Radius* and maximum), to combine them with the semantic distance algorithms and to compare them through the results they provide when applied to terms from the pharmacovigilance area. The comparison indicates that the non disjoint clustering (*Radius* and maximum) outperform the disjoint clusters by 10 to up to 20 points in all the experiments.

1 Introduction

In different applications, such as information retrieval, filtering or analysis, it is useful to be able to detect similar terms. For instance, the terms *heart attack*, *myocardial infarction* and *heart disease* are semantically close: when they occur in a document or in a corpus, it may be useful indeed to provide the system with such knowledge, which may allow providing more complete results and also reducing the false negatives. Detection of semantically close words and terms is a very intensive research topics and several studies proposed various methods: paraphrasing [1–3]; term variation detection [4–6]; semantic similarity computing [7–12]; terminology structuring or alignment [13–16], etc. However, once the semantic relatedness between terms is computed, it shows often different degrees of relatedness. Hence, it may be important to distinguish between those terms which are more close and those which have broader and weaker semantic relatedness between them. Typically, the clustering methods are helpful and can be exploited for this. The objective of our work is to compare several clustering methods. The comparison is done with terms from the pharmacovigilance area (usually meaning adverse drug reactions), which have been previously processed with semantic distance and similarity algorithms.

2 Background

We distinguish three types of clustering methods, according to the types of the clusters they generate:

- *Disjoint clusters*, in which a given object may belong to one cluster only. The disjoint clustering is done with algorithms such as *k-means* [17], *k-medoids* and *PAM* [18]. They are adapted to the processing of large data. These algorithms have two specificities: it is necessary to indicate the number of clusters to be generated and the generated clusters are disjoint;
- *Non disjoint clusters*, in which a given object may belong to more than one cluster. The non disjoint clustering is performed with so called fuzzy or soft algorithms. The fuzzy algorithms (*fuzzy c-means* [19], *fuzzy c-medoids* [20] or *axial k-means* [21]) state the degree up to which an object belongs to each concerned cluster. The difficulty with these algorithms is that they require to set up the thresholds, which may be a difficult step. The few existing soft clustering algorithms (*Radius* algorithm [22], PoBOC [23], OKM [24] or *Maximum* algorithm integrated within the *R project*) also allow an intersection between the generated clusters but without specifying the degree of relevance of each entity to a given cluster.
- *Hierarchical clusters* are considered as non disjoint when viewed through the dendrogram (smaller clusters are included into the larger clusters) or disjoint once the dendrogram is cut. Several hierarchical clustering algorithms have been proposed (AGNES [25, 26], BIRCH [27], CURE [28] and DIANA [26]). It is not necessary to set up the classes number, which eases the exploitation.

Our objective is to compare clustering algorithms within context related to the pharmacovigilance (detection and prevention of adverse drug reactions). The specificity of our data is that terms often show several semantic facets: because of their inherent semantics (*i.e.*, *Respiratory failure neonatal* is a malignancy, an abnormality of the respiratory system and a neonatal abnormality) and because of their medical manifestations (*i.e.*, *Respiratory failure neonatal* may appear as sign or symptom of several medical problems: *i.e.*, *Hypovolaemic shock conditions, Anaphylactic/anaphylactoid shock conditions, Hypoglycaemic and neurogenic shock conditions, Torsade de pointes*). For this reason, we put the priority on the non disjoint clustering methods, which allow one term to belong to more than one cluster, and compare them with the disjoint clustering.

3 Material

Pharmacovigilance terms: ontoEIM resource. The semantic resource ontoEIM [29] contains terms which describe the adverse drug reactions (*i.e.*, signs and symptoms, diagnostics, therapeutic indications, complementary investigations, medical and surgical procedures, medical and family history). The terms are provided by the MedDRA terminology [30]. The difference with MedDRA is that the ontoEIM terms have been restructured thanks to their projection on the terminology SNOMED CT [31], done through the exploitation of the UMLS [32], where an important number of terminologies are already merged and aligned, among which MedDRA and SNOMED CT. We exploit the preferred MedDRA terms *PT*. Their current rate of alignment with those from SNOMED CT is rather weak: 51.3% (7,629 terms). The restructuring of MedDRA terms makes

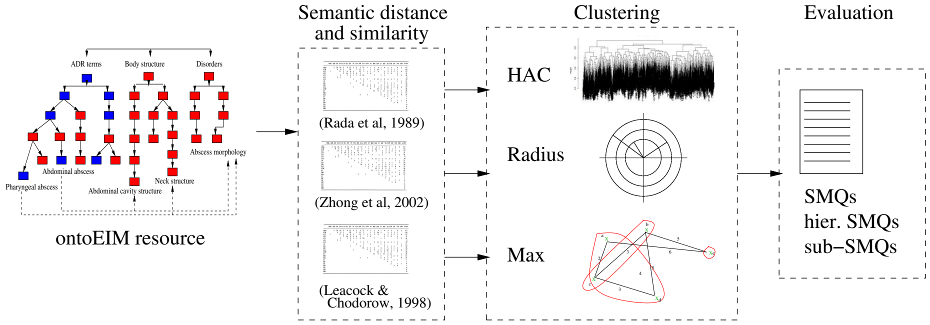


Fig. 1. General schema of the method.

the structure more fine-grained: the SNOMED CT-like hierarchy within ontoEIM contains also terms added to fill in the intermediate levels absent among MedDRA terms. The maximal number of the hierarchical levels within ontoEIM reaches up to 14, while only five hierarchical levels are available in MedDRA.

Reference clusters: Standardized MedDRA Queries (SMQs). Currently, 84 Standardized MedDRA Queries (SMQs) are available, which have been created manually by international boards of experts. The SMQs gather MedDRA terms related to a given safety topic (or medical problem), such as *Cardiac arrhythmias*, *Malignancies* or *Hepatic disorders*. The SMQs are mostly plain lists of terms, but 20 SMQs have the particularity to provide a hierarchical structure. The number of the hierarchical levels they contain vary between 2 and 4. The hierarchically structured SMQs are composed of sub-SMQs (n=92). We exploit these different levels of the reference data: SMQs (n=84), hierarchical SMQs (n=20) and sub-SMQs (n=92).

4 Methods

Figure 1 presents the general schema of the method organized in three main steps: (1) semantic distance and similarity computing between MedDRA terms, (2) MedDRA terms clustering, (3) and evaluation of the obtained clusters against the reference data. For the implementation, we exploit Perl and R^1 languages.

4.1 Computing of the semantic distance and similarity

Semantic distance and similarity algorithms state the semantic relatedness degree between two terms. For instance, *Respiratory failure neonatal* term is closer to term *Respiratory failure* than to *Cardiac failure*. The semantic distance and similarity algorithms may require the use of corpus and/or of terminologies. In our work, we use the terminological resource ontoEIM. Three measures are applied to the 7,629 *PT* MedDRA terms present in the ontoEIM resource in order to compute the distance and similarity values for each pair of terms *t1* and *t2*:

---

[1] http://www.r-project.org

- the *Rada* algorithm [7] computes the semantic distance and relies on the detection and computing of the shortest path *sp*, which corresponds to the sum of the edges of this shortest path: $sp(t1, t2)$;
- the *LCH* Leacock and Chodorow algorithm [9] computes the semantic similarity and relies on the shortest path *sp* and on the maximal depth MAX found within the terminology (MAX=14 within ontoEIM): $-log[\frac{sp(t1,t2)}{2*MAX}]$;
- the *Zhong* approach [10] computes the semantic distance and relies on the absolute depth *depth* of terms and on their closest common parent *ccp*. The milestone value m is computed first for each term: $m(t) = \frac{1}{k^{depth(t)+1}}$, where t is a term, *depth* its absolute depth within a terminology and $k = 2$ (normalization coefficient). Then, the distance between two terms is computed: $2 * m(ccp(t1, t2)) - (m(t1) + m(t2))$, where *ccp* is the closest common parent and m milestone values obtained previously.

Further to the application of these three algorithms, we build three symmetrical matrices 7629*7629 (one for each algorithm). They contain the semantic distance and similarity values between the MedDRA PT terms from ontoEIM.

4.2 Clustering of terms

Once the distances and similarities are computed, we use them for the creation of clusters of terms. We exploit and compare three methods for the clustering of the terms applied to matrices with the semantic distances and similarities:

- *HAC* hierarchical ascendant classification is performed through the *R Project* tools (*hclust* function). This method first chooses the best centers for clusters and then builds the hierarchy of terms by progressively merging smaller clusters to obtain the bigger ones and to build one unique dendrogram. The dendrogram is then segmented into x clusters. We test the following values: {100}, {200}, {300}, {400}, {500}, {1000}, {1500}, {2000} . . . {7000}. After the segmentation, the obtained clusters are exclusive.
- *Radius* method, where every ADR term is considered as a possible center of a cluster and its closest terms are clustered together with it. We test several threshold values x with the three semantic measure approaches, *i.e.*, with *Rada*: $\forall x \in N, such\ as\ x \in [1; 5]$; with *Zhong*: $\forall x \in N, such\ as\ x \in$ [0.001; 0.002; . . . ; 0.021]. The obtained clusters are not exclusive;
- *Max* maximum method is similar to the *Radius* approach but is more permissive. Iteratively, it computes the cost of the clusters union, while the radius approach computes the cost of the inclusion of each term (it does not consider the notion of the cluster). The *Max* thresholds x tested: $\forall x \in N, such\ as\ x \in [2; 5]$. The following steps are repeated for each node:
  1. Assign a node a_1 to a cluster c;
  2. Create the list l containing all the remaining nodes;
  3. For each node a_i from l, compute the cost of the union (c, a_i). The cost corresponds to $Maximum(distance)$ within a cluster, hence the name of the method. If there is more than one node with the same minimal distance value (step 6), then the comparison is done with the next greater value and so on;

		SMQs	Hier. SMQs	sub-SMQs
$Radius_{Rada}$	Threshold	4	4	2
$Radius_{Zhong}$	Threshold	0.006	0.019, 0.013	0.003, 0.004
HAC_{Rada}	Nb of classes	300, 400, 500, 2500	100, 200	3500
HAC_{Zhong}	Nb of classes	500	100	1500, 2500, 4000
Max_{Rada}	Threshold	4	5	4
Max_{Zhong}	Threshold	0.021	0.021	0.009

Table 1. Best clustering parameters (thresholds and number of classes) defined thanks to the cross-validation.

4. Delete all nodes a_i, which belong to l and whose union cost with cluster c is above a given threshold;
5. If l is empty, terminate the algorithm;
6. The node from l which shows the lowest union cost is added to the cluster c and removed from the list l;
7. Restart from the step 3.

4.3 Generated clusters evaluation

Generated clusters evaluation is performed thanks to their comparison with all the SMQs, the hierarchical SMQs and the sub-SMQs. A cluster is associated to the SMQ with which it has the maximal F-measure. Setting of thresholds and of classes number is performed through a ten-fold cross-validation: the data are partitioned into ten subsets, one subset is used for the setting up the methods while the remaining nine subsets are used for the evaluation against the reference data. This process is done ten times with a different training subset each time. Values which show the best performance on training set are applied on the test set. Average performance is then computed with three classical measures (where *relevant* terms are those which are clustered together and which also belong to the corresponding SMQ): precision P (percentage of the relevant terms clustered divided by the total number of the clustered terms), recall R (percentage of the relevant terms clustered divided by the number of terms in the corresponding SMQ) and F-measure F (the harmonic mean of P and R). The final evaluation values are the mean values of those obtained at each cross-validation step. We perform also a detailed analysis of individual clusters and of failures.

5 Results and Discussion

Because the *LCH* and *Rada* algorithms are similar (*LCH* adds the $log[2 * MAX]$ constant), values they provide are also close. Testing these two algorithms allowed to check out the performed computing correctness. In the following, we present and discuss results obtained with *Rada* and *Zhong* similarity distances.

In Table 1, the best settings defined, further to the cross-validation, for the two semantic algorithms and the clustering methods are presented. On the whole,

	SMQs			Hier. SMQs			sub-SMQs		
	P(%)	R(%)	F	P(%)	R(%)	F	P(%)	R(%)	F
$Radius_{Rada}$	45	32	36	40	33	35	56	36	43
$Radius_{Zhong}$	34	24	27	26	36	30	36	27	30
HAC_{Rada}	44	11	17	26	12	16	60	20	29
HAC_{Zhong}	34	16	21	30	16	18	53	22	29
Max_{Rada}	49	30	37	56	26	36	46	38	39
Max_{Zhong}	38	24	29	36	26	30	59	23	33

Table 2. Average performance on 0-100% scale (precision, recall, F-measure), against the three reference data sets, obtained with the best clustering parameters defined.

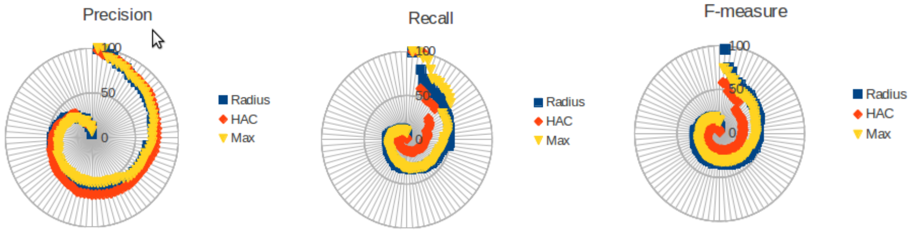
we observe that settings, which lead to smaller clusters (lower *Rada* and *Zhong* thresholds and higher number of classes), prove to suit best the generation of sub-SMQs. On the contrary, the hierarchical SMQs require the largest clusters (high *Rada* and *Zhong* thresholds and lower number of classes). The threshold values required for the generation of the whole set of the SMQs (hierarchical and non hierarchical, but excluding separate sub-SMQs) are intermediate. Such an observation was expected. It is closely related to the size of the generated clusters and it follows the logics of the reference data: the sub-SMQs provide the smallest clusters, while the hierarchical SMQs the largest.

In Table 2, we indicate average evaluation values (precision, recall and F-measure) obtained with the three reference data sets (SMQs, hierarchical SMQs and sub-SMQs) and with the best parameters (Table 1). During cross-validation, we give advantage to F-measure. Compared to previous experiments without cross-validation, which gave advantage to precision [22], we obtain currently a best balance between precision and recall, and the whole performance is improved by 5 to 10%. Still, precision values remain higher than recall values. This is due to the fact that the generated clusters, whatever the methods and reference data, are smaller than the reference data. The generated clusters typically capture a given aspect of the reference SMQs: their recall is low, while precision may reach up to 60%. On the whole, the task related to the automatic creation of the SMQs remains difficult. With the currently exploited resources and methods we can capture but partially the terms relevant to a given medical condition.

With values indicated in Table 2, we can also propose a comparison between the three clustering methods tested in this work. Whatever the reference data, the non disjoint clusters outperform the disjoint HAC clusters by 10 to 20 points of F-measure. The only situation in which the HAC method is better is observed with precision obtained with the *Rada* algorithm and sub-SMQs. Besides, this is the best precision we obtain with the presented experience: 60%. The next best precision is also obtained with the sub-SMQs but with *Max* method and *Zhong* algorithm: 59%. Our results seem to indicate that non disjoint clustering is more suitable for the aimed task. MedDRA terms may indeed be specific to more than one medical condition and belong to several clusters. Logically, this aspect is better captured when the non disjoint clusters are generated.

	Precision			Recall			F-measure		
	average	min	max	average	min	max	average	min	max
$Radius_{Rada}$	56	4	100	36	7	100	43	8	96
HAC_{Rada}	60	12	98	20	1	100	29	2	58
Max_{Rada}	46	7	100	38	8	100	39	14	74

Table 3. Average, minimal and maximal values for precision, recall and F-measure obtained against the reference sub-SMQs with the best clustering parameters defined with the *Rada* algorithm.



(a) Precision (b) Recall (c) F-measure

Fig. 2. Evaluation results against the 92 sub-SMQs (Precision, Recall and F-measure) and the three clustering methods applied to the *Rada* semantic distance matrix.

In Table 2, we presented the average values of the evaluation measures, while there is an important variability according to the clusters and the reference data. Hence, in Table 3, we present also, along with the average values, the minimal and maximal performance obtained with the reference sub-SMQs. We can see that the interval is very important and that there is indeed a very important variability across the sub-SMQs, as presented in Table 3, but the situation is similar with two other sets of the reference data. In relation to this observation, some of the SMQs are better reproduced than others. Among the best clusters, we have *Gastrointestinal obstruction*, *Liver-related coagulation and bleeding disturbances*, *Ischaemic cerebrovascular conditions*. Among the less competitive clusters, let's cite *Reproductive premalignant disorders* and *Pregnancy complications*.

In Figure 2, we present the evaluation values for individual sub-SMQs. In this circular lay-out, the results are not projected on the x and y axes. The 360 degrees correspond to the 92 reference sub-SMQs, while the radius 0-100 scale allows to position the evaluation measure values. For a given evaluation measure, the values are first sorted in a decreasing order and then projected. While reading the figures, it is necessary to notice that more a given line is closer to the outer border, the better the results for the corresponding method and measure. For instance, we can see that precision values start with 100% performance, that more than 1/3 of the precision values are higher than 50%, and that less than 1/4 of the precision values are lesser than 50%. We can see also that the highest F-measure values do not start with 100%, but with lower values (see Table 3).

	HAC_{Rada}			$Radius_{Rada}$			Max_{Rada}		
	P	R	F	P	R	F	P	R	F
Anaphylactic anaphylactoid shock conditions	100	8	14	60	23	33	43	23	30
Infectious biliary disorders	40	18	25	86	54	66	69	81	75

Table 4. Detailed performance for two randomly chosen clusters.

Figure 2 provides several observations: (1) once again we can observe that there is an important variability between the SMQs; (2) very often, the precision is high while the recall is low (the generated clusters are smaller than the SMQs and show their different aspects); (3) the *HAC* clusters provide the lesser F-measure and recall performance (the red line, which correspond to this clustering is the closest to the center of the circles); (4) the *Radius* and *Max* clustering methods are comparable (the corresponding lines are often superimposed). We observe also that recall and F-measure are lower with *HAC* than with the other two clustering methods, which is due to the disjoint clusters generated by *HAC*.

We did a detailed analysis of two randomly chosen clusters (corresponding to two sub-SMQs), presented in Table 4: *Anaphylactic anaphylactoid shock conditions* and *Infectious biliary disorders*. *Anaphylactic anaphylactoid shock conditions* sub-SMQ contains 13 terms. The *HAC* method proposes only one relevant term (*Renal failure acute*). While *Radius* and *Max* method provide with respectively 5 and 7 relevant terms. Situation is very similar with the *Infectious biliary disorders* cluster: it contains 11 terms, among which 5 are also provided by *HAC*, 7 by *Radius* and 13 by *Max*. In both cases, the non disjoint clustering (*Radius* and *Max*) is more suitable for the aimed task than the disjoint clustering (*HAC*). Terms which are not collected with our methods are too distant in the exploited resource. Other methods and approaches should be used to capture them.

6 Conclusion and Perspectives

We presented an experiment on the comparison between three clustering methods (hierarchical ascendant classification, *Radius* and *Max*) applied to pharmacovigilance terms, which have been previously processed with the semantic distance and similarity algorithms. Our objective is to compare between disjoint and non disjoint clustering methods. The exploited reference data are composed of manually created sets of pharmacovigilance terms related to various medical conditions. The cross-validation allowed to define the best clustering parameters (thresholds and number of classes), which favor the global performance (F-measure) and which reach the best balance between precision and recall. As for the comparison between the clustering methods, the non disjoint clustering (*Radius* and *Max*) outperform the disjoint clusters by 10 to up to 20 points for nearly all the experiments and evaluation measures. Hence, the non disjoint clustering captures better the multi-facet characteristics of the pharmacovigilance terms. We assume these results are also relevant to other tasks and applications dealing with semantics and language data.

In the future, we plan to apply the proposed methods to other terminological resources, such as UMLS [32] or its subsets. In the current experience, only individual clusters have been considered, while we showed in past work that their merging is helpful as it allows increasing the recall almost without decreasing precision: merging of the clusters issued from the current study is a perspective. We also observed that there is a great variability across the clusters and the reference data. Currently, we apply the same setting to all the reference data, while we can distinguish several settings suitable for subsets of the medical conditions. We assume, this may capture better the inherent semantics of these subsets and to improve the overall results. Moreover, we plan to combine this method with other methods (exploitation of corpora and of Natural Language Processing methods...). Finally, we would like to test the proposed methods for the creation of novel SMQs describing not yet covered medical conditions.

Acknowledgement. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) and the DGA, under the Tecsan grant ANR-11-TECS-012. The authors also would like to thank anonymous reviewers for their constructive remarks.

References

1. Barzilay, R., Elhadad, N.: Sentence alignment for monolingual comparable corpora. In: EMNLP. (2003) 25–32
2. Paca, M.: Mining paraphrases from self-anchored web sentence fragments. In: Knowledge Discovery in Databases: PKDD 2005. (2005) 193–204
3. Max, A., Bouamor, H., Vilnat, A.: Generalizing sub-sentential paraphrase acquisition across original signal type of text pairs. In: EMNLP. (2012) 721–31
4. Jacquemin, C.: A symbolic and surgical acquisition of terms through variation. In Wermter, S., Riloff, E., Scheler, G., eds.: Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing, Springer (1996) 425–438
5. Daille, B., Habert, B., Jacquemin, C., Royauté, J.: Empirical observation of term variations and principles for their description. Terminology **3**(2) (1996) 197–257
6. Hahn, U., Honeck, M., Piotrowsky, M., Schulz, S.: Subword segmentation - leveling out morphological variations for medical document retrieval. In: Annual Symposium of the American Medical Informatics Association (AMIA), Washington (2001)
7. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on systems, man and cybernetics **19**(1) (1989) 17–30
8. Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: Proceedings of Associations for Computational Linguistics. (1994) 133–138
9. Leacock, C., Chodorow, M.: 4. In: Combining local context and WordNet similarity for word sense identification. (1998) 305–332
10. Zhong, J., Zhu, H., Li, J., Yu, Y.: Conceptual graph matching for semantic search. In: 10th International Conference on Conceptual Structures, ICCS2002, LNCS 2393, Springer Verlag. (2002) 92–106

11. Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in wordnet. In: Proceedings of ECAI 2004, the 16th European Conference on Artificial Intelligence. (2004) 1089–1090
12. Nguyen, H., Al-Mubaid, H.: New ontology-based semantic similarity measure for the biomedical domain. In: IEEE Eng Med Biol Proc. (2006) 623–8
13. Maedche, A., Staab, S.: Mining ontologies from text. In: EKAW. (2000)
14. Bodenreider, O., Pakhomov, S.: Exploring adjectival modification in biomedical discourse across two genres. In: Workshop Natural Language Processing in Biomedical Applications of ACL. (2003) 105–112
15. Grabar, N., Zweigenbaum, P.: Lexically-based terminology structuring. In: Terminology. Volume 10. (2004) 23–54
16. D'aquin, M., Euzenat, J., Le Duc, C., Lewen, H.: Sharing and reusing aligned ontologies with cupboard. In: K-CAP 2009. (2009) 179–180
17. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. (1967) 281–297
18. Kaufman, L., Rousseeuw, P. In: Clustering by means of medoids. (1987) 405–416
19. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York (1981)
20. Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L.: Low complexity fuzzy relational clustering algorithms for web mining. In: IEEE Trans. Fuzzy System. (2001) 595–607
21. Lelu, A.: Modles neuronaux pour lanalyse de donnes documentaires et textuelles. Phd thesis, Universite de Paris VI, Paris, France (1993)
22. Dupuch, M., Bousquet, C., Grabar, N.: Automatic creation and refinement of the clusters of pharmacovigilance terms. In: ACM IHI. (2012) 181–190
23. Cleuziou, G., Martin, L., Vrain, C.: PoBOC: an overlapping clustering algorithm. application to rule-based classification and textual data. In: ECAI. (2004) 440–444
24. Cleuziou, G.: OKM: une extension des k-moyennes pour la recherche de classes recouvrantes. In: EGC. (2007) 691–702
25. Johnson, S.: Hierarchical clustering schemes. Psychometrika **32** (1967) 241–254
26. Kaufman, L., Rousseeuw, P.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York (1990)
27. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: an efficient data clustering method for very large databases. In: ACM SIGMOD. (1996) 103–114
28. Guha, S., Rastogi, R., Shim, K.: Cure: an efficient clustering algorithm for large databases. In: ACM SIGMOD. (1998) 73–84
29. Alecu, I., Bousquet, C., Jaulent, M.: A case report: using snomed ct for grouping adverse drug reactions terms. BMC Med Inform Decis Mak **8**(1) (2008) 4–4
30. Brown, E., Wood, L., Wood, S.: The medical dictionary for regulatory activities (MedDRA). Drug Saf. **20**(2) (1999) 109–17
31. Stearns, M., Price, C., Spackman, K., Wang, A.: SNOMED clinical terms: overview of the development process and project status. In: AMIA. (2001) 662–666
32. NLM: UMLS Knowledge Sources Manual. National Library of Medicine, Bethesda, Maryland. (2008) www.nlm.nih.gov/research/umls/.