

Exploitation of semantic similarity for adaptation of existing terminologies within biomedical area

Marie Dupuch¹, Ludovic Trinquart², Isabelle Colombet^{1,3}, Marie-Christine Jaulent¹, Natalia Grabar^{1,3}

(1) Centre de Recherche des Cordeliers, Université Pierre et Marie Curie - Paris6, UMR_S 872, Paris, F-75006; Université Paris Descartes, UMR_S 872, Paris, F-75006; INSERM, U872, Paris, F-75006 France

(2) AP-HP, Paris France

(3) HEGP AP-HP, 20 rue Leblanc, Paris, F-75015 France

Abstract. We present a novel method for adaptation of existing terminologies. Within biomedical domain and when no textual corpora for building terminologies are available, we exploit UMLS metathesaurus which merges over a hundred existing biomedical terminologies and ontologies. We exploit also algorithms for measuring the semantic similarity in order to limit, within UMLS, a semantically homogeneous space. In order to make possible such measure, we transform symbolic and semantic information associated to UMLS relations into numeric values. Evaluation performed by experts indicates that extracted concepts are relevant and their coverage satisfying. Additional semantic filters allow to improve the precision of the obtained terminology.

1 Introduction

During the last decades, powerful methods for designing and use of ontologies have been proposed. They lead to an increasing number of existing ontologies and terminologies, and several of them are made available for the community. For instance, the OBO (Open Biological and Biomedical Ontologies) initiative is a collaborative experiment in establishing principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The website of this initiative (www.obofoundry.org/) counts currently over 70 ontologies describing various aspects of the biomedical domain, such as model organisms, their anatomy, biological functions, chemical entities, pathologies, imaging methods, and a great variety of other aspects. Going beyond the creation of orthogonal ontologies, biomedical area provides also UMLS [1], which merges currently over 100 biomedical terminologies. Because of the existence and availability of ontologies and terminologies, the research concern moves towards their reuse. We present an experience on reuse and adaptation of the existing ontologies.

The main motivations of our work are: (1) to take advantage of the existing authoritative semantic resources and in this way to facilitate the ontology and

terminology creation task; (2) to propose a methodology and an experience on how the existing semantic resources may be reused; (3) to propose methods for an easy and simple reuse and adaptation of large and general semantic resources. Within a broader context we aim also at (4) contributing to the ontology modularization research [2,3] thanks to identification of semantically homogenous graphs within large monolithical semantic resources.

Our applicational context is related to systematic reviews in biomedical domain. We present this context (sec. 2) and we explain why the adaptation of terminologies is preferred to the building of a new ontology. We present and discuss the related work (sec. 3). We then present our material (sec. 4) and the method we propose for the adaptation of terminologies (sec. 5), which is based on a semantic similarity measure. We next present and discuss the results obtained (sec. 6), and conclude with some perspectives (sec. 7).

2 Applicational context: Systematic reviews

The aim of systematic reviews is to provide a synthesis of scientific literature concerned with a given clinical question. Such questions may be very precise, such as *diagnosis of liver metastasis of a colorectal cancer*, and belong to a given kind of studies: diagnostic (which is the case with this example), therapeutic, prognostic, genetic or observational. Syntheses provided by systematic reviews are often used by health professionals for making clinical decisions. The main concern is then their exhaustivity for which several bibliographical databases are exploited. This usually leads to a huge amount of collected articles (or citations), while most of them are non relevant to the studied clinical question and must be filtered out. Among systematic reviews, review of diagnostic studies is a new area, and this introduces several additional difficulties: (a) classical queries used in systematic reviews [4,5] show to be non optimal; (b) mainly because the indexing of diagnostic studies is not precise [6]; (c) supervised machine learning methods used for the automatic selection of the relevant studies for therapeutic systematic reviews [7–10] are not efficient because of the feeble amount of existing diagnostic reviews. Thus, diagnostic systematic reviews are performed manually which is a real burden to the scientists. We propose to base the process of selection of relevant articles on a semantical information retrieval system.

3 Related work

Approaches usually applied for building domain-specific semantic resources (terminologies or ontologies) exploit textual material [11,12], HTML and XML metadata [13], or databases [14,15]. Only none of these are available at the beginning of a new systematic review: selected citations which could feed the corpus are built as a result of a systematic review. In order to create a terminology relevant to precise clinical questions (*i.e.*, *diagnosis of liver metastasis of a colorectal cancer*), we propose to exploit existing resources and to adapt

them. We aim particularly at UMLS [1] which already gathers over a hundred biomedical terminologies and ontologies.

Up to date, the reusability of ontologies has been discussed [16, 17], although little has been done for their adaptation to new purposes. The addressed research questions are related to creation of a smaller ontology, necessary when ontologies have to be combined between them: this process leads to the problem of coherence and of interdependence between the top concepts [18]; and to revision of description logic-based ontologies in order to incorporate new concepts [19]. Let us also cite *slim GO*¹ subsets of Gene Ontology [20] terms created thanks to functional annotations of genes and proteins. The slims concentrate on a given species or on a set of genes and proteins and group terms exploited for their annotation. In this way, slim sets propose functionally homogeneous sets of terms. As for the ontology modularization research [2, 3], its main notions are related to the logical definition of modularity and of ontological modules, and to their application and combination. Our objective is dissimilar from these experiences, because we do not update or merge existing ontologies, neither make reverse engineering for modularization or reducing the coverage of terminologies. Our aim is to adapt the terminologies to a given medical question: within a terminological graph, we have to define semantically homogeneous and coherent sub-graphs. In this way, we assume we can fit an existing terminology to very precise clinical questions. The proposed approach is designed to be applied to a semantic network or graph containing nodes (terms or concepts) and arcs (relations or links). Such graphs are representative of either terminologies or ontologies, with the difference that in the former case the graph may contain cycles. Our approach must be applicable in both cases. From this point of view, any semantic graph (terminologies, ontologies, semantic networks) are equivalent. For extraction of homogeneous sub-graphs, we propose to exploit semantic similarity measures.

Semantic similarity quantifies similarity between nodes in a given terminology. Its potential applications are information retrieval, data mining, and knowledge discovery in database or decision-support systems, where it may be useful to know that, according to a terminology, some concepts are semantically related. Most of the measures are based on number of arcs between terms (or graph nodes) [21]. Additional notions have been proposed: depth and density of concepts [22]; depth of concepts and types of relationships [23]; informativity of concepts through their occurrences in corpora [24] or density of terminological network around them [25]; path length and change of link direction [26]. Such measures have been also applied within the biomedical area [27–31] through the exploitation of hierarchical relations from four terminologies: MeSH, ICD, Snomed and GO, and evaluated [31, 32]. We propose to exploit these measures in a context related to adaptation of existing semantic resources, such as UMLS: based on semantic similarity we expect to extract from this large semantic network semantically homogenous subsets. For a better respect of semantics of the relations, we exploit all the source terminologies of UMLS and various kinds of

¹ www.geneontology.org/GO.slims.shtml

semantic information associated to UMLS arcs (different types of relationships, redundancy of relations ...).

4 Material

4.1 UMLS: Unified Medical Language System

UMLS [1] is a metathesaurus merging over a hundred biomedical terminologies and ontologies, among which MeSH [33]. In 2009 version, UMLS provide over a million concepts, each of them consists of the preferred term and of its synonyms. UMLS store source information on each term, concept and relation. Each concept has an unique identifier *CUI* and its source terminologies are recorded. When included in UMLS, each concept is assigned to a corresponding hierarchy. For instance, *Colorectal neoplasms* belongs to *Neoplastic Process* hierarchy, and *Magnetic resonance imaging* to *Diagnostic Procedure*. All the concepts are linked between them. UMLS proposes two levels of relations: (1) Over one hundred relations provided by source terminologies: they are heterogeneous and their precision depends on their source; and (2) 16 relations specific to UMLS: they have been established by UMLS and each new included relation is systematically normalised into them. These relationships convey more or less close semantics between concepts:

PAR *has parent*, CHD *has child*, SIB *has sibling*
SY *synonym*, RL *has similar or like relationship*
RQ *related and possibly synonymous*, RN *has narrower relationship*
RO *has relationship other than synonymous, narrower or broader*
RB *has a broader relationship*, RU *related, unspecified*
QB *can be qualifier by*, AQ *allowed qualifier*

4.2 Descriptors for querying bibliographical databases

Any systematic review begins with the querying of bibliographical databases in order to collect bibliographical citations potentially relevant to this review. Typically, the query is done in the Medline database [34], which is the reference within the biomedical area, and in other databases. Queries submitted to Medline have an interest for our work: they contain main clinical concepts expressed with MeSH descriptors. We propose to exploit them because they define precisely the object of the clinical question and because MeSH is one of UMLS terminologies. Here is an excerpt of the boolean Medline query related to *diagnosis of liver metastasis of a colorectal cancer* (clinical MeSH descriptors are in bold characters):

"Colorectal Neoplasms"[MeSH] AND **"Liver neoplasms"**[MeSH] AND **"Laparoscopy"**[MeSH] OR **"Tomography, Emission-Computed"**[MeSH] OR **"Magnetic Resonance Imaging"**[MeSH] OR **"Tomography, X-Ray Computed"**[MeSH] AND ("sensitivity and specificity"[MeSH Terms] OR ("sensitivity and specificity"[TIAB] NOT Medline[SB]) ... OR "diagnosis"[MeSH Terms] OR detection[All Fields] OR accuracy[All Fields]) ...

Because MeSH terms are part of UMLS, these six descriptors (*Colorectal neoplasms*, *Liver neoplasms*, *Laparoscopy*, *Tomography*, *emission-computed*, *Magnetic resonance imaging*, and *Tomography, x-ray computed*) correspond to our entry points in UMLS.

5 Method

We apply semantic similarity measures in order to define within UMLS semantically coherent sub-graphs. Computing of semantic similarity is anchored on MeSH descriptors (sec. 4.2). Our method consists of the following steps: (1) extracting UMLS concepts around the anchoring MeSH descriptors; (2) weighting UMLS arcs with heuristics; (3) computing the unique weight \mathcal{W} ; (4) limiting the semantic space for each concepts. Evaluation of results is performed by two experts working in the area of systematic reviews. They were asked to decide whether a term is relevant to the systematic review question and whether it can be useful for selection of citations or not. Precision is computed for each concept and a mean value is computed for evaluation of the whole experience. Recall was not evaluated because we could not establish an extensive set of relevant terms.

5.1 Extracting concepts around the anchoring MeSH descriptors

The first step consists in extraction of potentially useful concepts from UMLS. For this, MeSH descriptors are used as anchoring points. At this step we have to manage cycles within UMLS graph.

5.2 Weighting UMLS arcs with heuristics

Second step proposes heuristics for assignment of weights to UMLS arcs, which will be later exploited for computing the semantic similarity. Values are positioned on a scale going from 0 to 100, where 0 corresponds to very close concepts, while 100 corresponds to broad concepts. These weights are defined and fitted empirically. We use currently two heuristics:

1. Type of relationships between two concepts T_{rel} . Within UMLS, each relation is labelled as PAR, CHD, SY, SIB, etc. (sec. 4.1). We exploit this information and assume for instance that SY terms $\{\textit{Colorectal Neoplasms}$, $\textit{Tumor of colon}\}$ have a lot of semantic elements in common. Thus, synonymy receives a minimal weight $w_{T_{rel}}(SY)$ set to 0. As for a relationship like RB, it links broad concepts, like $\{\textit{Colorectal neoplasms}$, $\textit{Lower digestive system}\}$, and receives a weight $w_{T_{rel}}(RB)$ set to 90: the involved concepts potentially have a small semantic recovery between them. If a relation is provided by different source terminologies and have different labels, its mean value is considered.
2. Redundancy among sources and number of source terminologies providing a given relation N_{sour} . As UMLS is a metathesaurus and merges several

terminologies and ontologies, some relations can be provided by more than one source. We assume that larger the number of source terminologies for a given relation higher its confidence and smaller the weight assigned to it.

5.3 Computing the unique weight \mathcal{W}

A unique weight \mathcal{W} is computed for each arc: it corresponds to a balanced mean of the individual weights w provided by the two heuristics (*Trel* and *Nsour*). This mean value is balanced because the heuristics are not equivalent between them. Their balancing follows the same principle as for arcs: they are positioned on a scale going from 0 to 100, and more reliable heuristics receive a lower coefficient. For computing the unique weight \mathcal{W} of arcs, we apply the formula:

$$\mathcal{W}_{[X_1 \dots X_n]} = \frac{\sum_{i=1}^n (W_{[X_i]} * w_{[X_i]})}{\sum_{i=1}^n W_{[X_i]}}$$

where n is the number of heuristics, X a heuristic, $w_{[X_i]}$ weight of a heuristic, and $W_{[X_i]}$ the coefficient assigned to this heuristic, which allows to balance its weight. Normalisation is performed thanks to the division by the sum of coefficients $\sum_{i=1}^n W_{[X_i]}$.

5.4 Limiting the semantic space for each concepts

Computing of homogeneous semantic space within UMLS graph relies on semantic similarity between concepts: terms with bigger similarity will be included in this space. This computing is performed within the previously extracted graphs (sec. 5.1). Among the three existing algorithms (Bellman-Ford, Floyd Warshall and Dijkstra [35]), we used the last one: it allows to exploit weights assigned to arcs and to apply a threshold to paths. Besides, weights and thresholds are the tools we exploit for the limitation of the space within graphs.

Extraction of terminological sub-graph from UMLS leads to the building of a descriptive terminology: it describes, as exhaustively as source terminologies it allows, the area related to the clinical question. Its content can cover, and this is usually the case, terms from various UMLS hierarchies (*i.e.*, pathologies, diagnostic methods, medications, etc). Descriptive terminologies are useful for various purposes, although they can generate noisy results when exploited for diagnostic systematic reviews: presence of certain terms may be annoying. For instance, drug names (*streptozocin* for colorectal cancer) favour articles related to therapeutic studies and thus give less importance to diagnostic studies.

We propose to apply an additional filter in order to better fit the semantic resource to the application. The designed filter exploits the UMLS hierarchies of concepts. Its methodological principle is the following: among the extracted concepts, we select those that belong to the same hierarchies as the query descriptors. Such filtering allows to restrict even more the semantic homogeneity of the extracted sub-graphs.

6 Results and Discussion

Six MeSH descriptors (sec. 4.2) correspond to eight UMLS concepts (with their unique UMLS identifiers): *Colorectal neoplasms* (C0009404), *Liver neoplasms* (C0023903), *Laparoscopy* (C0031150), *Tomography, emission-computed* (C0040405), *Magnetic resonance imaging* (C0024485), *Tomography, x-ray computed* (C0040398), *Tomography, emission-computed* (C1552358) and *Laparoscopy* (C1883297). Two of them (*Laparoscopy* and *Tomography, emission-computed*) have two matching concepts. All of them are used as the entry points in UMLS.

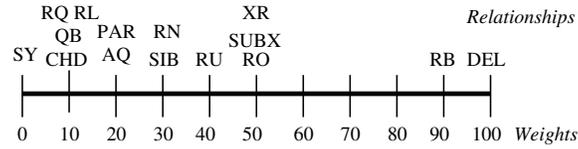
6.1 Extracting UMLS concepts around the MeSH descriptors

The eight UMLS concepts are exploited for performing the first extraction from UMLS. At depth one we obtain 2,590 concepts and 63,911 at depth two. At depth three, this number reaches up to one million concepts, which corresponds nearly to the entire UMLS. In this experience, we observed that terms become not relevant when the depth is superior to two: this is the maximal depth allowed. Notice that this value is a variable of the method and can be adjusted if necessary. More particularly, we may use different depth values according to relationships: more restriction may be required with hierarchical-like relations (CHD, PAR, RB, RN, SIB) than with synonymy-like relations. At this step, density of links within UMLS is very high for concepts *Colorectal Neoplasms* (C0009404), *Liver Neoplasms* (C0023903), *Laparoscopy* (C0031150), *Tomography, X-Ray Computed* (C0040405) and *Magnetic Resonance Imaging* (C0024485). *Colorectal Neoplasms* (C0009404) especially appears to be central to the studied area.

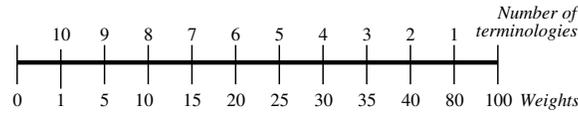
6.2 Weighting UMLS arcs and computing the semantic similarity

This is the main and the most meticulous step of the method. For instance, the weights are defined and adjusted empirically: currently, they are considered as variable values. Present and future evaluations are intended to fit these values and to further improve the performances of the method.

Definition of weights for two heuristics (*Trel* and *Nsour*) are presented on figure 1. On figure 1(a), the 16 UMLS relationships are positioned: synonymy SY receives the minimal possible weight, while with the broadening of the semantics of relationships the weight increases. Presently, all the relationships are considered individually and at the same level. A variation of the method may consist in grouping together those relationships which convey a comparable semantics: CHD/RN among structural relationships, AQ/RQ as functional relationships. Notice that currently relationships from these pairs receive different weights. Such evolution of the method requires a very precise understanding of terminological principles and conventions of the source terminologies. As for figure 1(b), it presents weights defined according to the number of terminologies providing a given relation: it exploits the redundancy of relations in source terminologies. This weight is minimal when a relation is provided by 10 or more source terminologies. It increases as the number of sources decreases. Figure 2 presents



(a) *Trel* heuristic and definition of weights according to type of relationships (see sec. 4.1 for the meaning of relationships)



(b) *Nsour* heuristic and definition of weights according to the number of source terminologies providing a given relation

Fig. 1. Definition of weights for two heuristics: *Trel* and *Nsour*

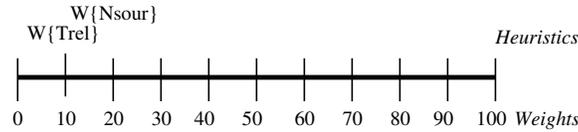


Fig. 2. Weighting of heuristics for computing the unique weight for an arc

weighting of these two heuristics between them: *Trel* is more reliable and receives a lower coefficient (also a variable of the method). With this last operation, all the elementary weights are combined and an unique weight \mathcal{W} is computed for each arc. These weights are next exploited for limiting the terminological graph of UMLS: within each possible path, weights of arcs are summed up and their total value should not exceed the threshold. The threshold is currently set to 65: in this experience, it shows a good compromise for selection of concepts. The method extracts then 2,503 and 2,620 concepts at depth 1 and 2 respectively.

Several limitations appear at this step: bias due to manual and empirical definition of the weights, as well as the fact that UMLS merges several terminologies, each of which is built for a given purpose and may introduce inconsistencies at level of concepts and relations. Nevertheless, we assume that combination of these two heuristics, in respect with the semantics of UMLS relations, allows to block, at least partially, possible biases. Compared to the state of the art, the proposed method presents three original points: (1) it goes beyond the traditional conception of semantic similarity, usually based on exploitation of hierarchical relations only; (2) it combines different heuristics and semantic information, which leads to a better understanding and estimation of the seman-

<i>MeSH descriptors</i>	<i>CUI</i>	<i>Experts</i>	<i>Trld=65</i>	<i>P</i>	<i>Filter</i>	<i>P</i>
<i>Colorectal neoplasms</i>	C0009404	16	153	10.5	37	43.2
<i>Liver neoplasms</i>	C0023903	10	224	4.5	131	7.6
<i>Laparoscopy</i>	C0031150	9	249	3.6	44	20.5
<i>Laparoscopy</i>	C1883297	0	44	(0)	0	(100)
<i>Tomography, Emission-Computed</i>	C0040398	0	8	(0)	0	(100)
<i>Magnetic Resonance Imaging</i>	C0024485	19	1,082	1.8	318	6.0
<i>Magnetic Resonance Imaging</i>	C1552358	0	2	(0)	0	(100)
<i>Tomography, X-Ray Computed</i>	C0040405	48	858	5.6	247	19.4
<i>Total number of concepts & Mean precision</i>		102	2,620	3.25	777	49.59

Table 1. Filtering and validation of concepts and the obtained precision \mathcal{P}

tics of the exploited terminologies; and (3) it proposes a method for exploitation of information provided by several existing terminologies.

6.3 Evaluation and additional filters

Evaluation figures are indicated in table 1. For each concept, labels and unique identifiers of concepts are indicated in the first two columns. We then indicate selection performed by experts (column *Experts*), which corresponds to our gold standard. Column *Trld=65* indicates number of concepts selected with semantic similarity approach and threshold set to 65. \mathcal{P} states for precision obtained with this selection. The last two columns *Filter* and \mathcal{P} indicate number of concepts and precision obtained when additional semantic filter is applied. Precision is first computed for each concept, *i.e.* it is 10.5 for *Colorectal neoplasms* and 4.5 for *Liver neoplasms* in the column *Trld=65*. The mean precision is then computed for eight concepts: 3.25 for *Trld=65* and 49.59 for *Filter*. We can observe that precision values by concept vary a lot, which means that currently some concepts are more difficult to process than others. If a global (and not mean) precision is computed, it is 3.89 for *Trld=65* and 13.13 for *Filter*. Specifically with *Trld=65*, precision is low: it varies between 0 and 10.5%. We assume this is due to the fact that at this stage the obtained terminology is descriptive and not specifically oriented on diagnostic systematic reviews. Implementation of additional semantic filter based on UMLS hierarchies increases the precision: values then vary between 6 and 100%, while the mean precision reaches up to 49.59%. We assume that this level of precision becomes satisfying, although it varies greatly according to concepts. Application of other approaches and of semantic similarity measures and study of their performances remain a perspective to this work. First of all, the proposed approach allows to reduce considerably time spent by experts on the terminology building task. It allows also to take advantage of several existing terminologies without prior expert knowledge of these (and of their thousands or even millions concepts). The only expertise required is the knowledge of the medical area related to the clinical question of a systematic review. An additional qualitative analysis of the 102 concepts selected by experts shows that:

- These concepts are mainly provided by three relationships CHD, SIB and PAR. Relationships like RQ and RN occur seldom (2 or 3 times), and RO only once. Other relationships (AQ, QB and RB), although frequent in UMLS, do not occur in the expert selection. These observations will lead to an adjustment of the method and of the proposed weights.
- These concepts are provided by different source terminologies, among which WHO-ART, MEDCIN, NCI, RCD, MSH and SNOMED CT. This fact especially justifies the exploitation of UMLS. If one single medical terminology were used it would reduce significantly semantic coverage of terms proposed to experts. A more detailed analysis of the impact of each terminology exploited must be done yet.

7 Conclusion and Perspectives

We proposed a novel method for adaptation of existing terminologies and ontologies thanks to an original exploitation of semantic similarity and to the extraction of an homogeneous subset of concepts. The experience has been held in biomedical domain where a large number of terminologies are already created and used. This approach is an alternative to the production of ontologies and terminologies when no textual corpora are available.

Precision obtained reaches up to 49.59% and reduces considerably the manual validation by experts. Evaluation of recall could not be done as no reliable gold standard exists, but experts found the current coverage satisfying. This method should be tested and evaluated in other contexts and applications of the biomedical area. Besides, the difference between general and precise ontologies should be studied, and especially the criteria which allow to differentiate them.

Compared to the state of the art, the proposed method presents three original points: (1) it goes beyond the traditional conception of semantic similarity, usually based on exploitation of hierarchical relations only; (2) it combines different heuristics and semantic information, which leads to a better understanding and estimation of the semantics of the exploited terminologies; and (3) it proposes a method for exploitation of information provided by several existing terminologies. Exploitation of these semantic data is an exciting aspect of the method. First, it makes possible application of various operations on terminological graphs (such as computing of semantic similarity). Second, it leads to a better consideration of the semantics of terminological relations.

This is an ongoing work and it has several perspectives: creation of classes of semantically close relationships instead of considering them individually; taking into account the relative depth of nodes, according to relationships; fitting the weights and threshold; study of the individual impact of each heuristic; comparison with previous work [29, 30, 32]; definition of other criteria and heuristics on the basis of other semantic information of UMLS network. As we mentioned, use of UMLS can lead to a bias because it merges several terminologies which ontological involvement and commitment are different and this can introduce inconsistencies. However, the precise nature of these possible inconsistencies has

to be studied yet, as well as their impact on results. Besides, we plan to study whether the use of several heuristics allows to inhibit them.

Finally, terminological sub-graphs generated by this approach will be used by an automatic tool for selection of articles in the context of systematic reviews. This will provide yet another evaluation of the results provided by the proposed method.

References

1. NLM: UMLS Knowledge Sources Manual. National Library of Medicine, Bethesda, Maryland. (2008) www.nlm.nih.gov/research/umls/.
2. D'aquin, M., Schlicht, A., Stuckenschmidt, H., Sabou, M.: Ontology modularization for knowledge selection: Experiments and evaluations. In: 18th International Conference on Database and Expert Systems Applications - DEXA'07, Regensburg, Germany (2007) 874883
3. Stuckenschmidt, H., Parent, C., Spaccapietra, S.: Modular Ontologies. Concepts, Theories and Techniques for Knowledge Modularization. Springer (2009)
4. Meade, M., Richardson, W.: Selecting and appraising studies for a systematic review. *Ann Intern Med* **127**(7) (1997) 531–7
5. Haynes, R.B., Wilczynski, N.L.: Optimal search strategies for retrieving scientifically strong studies of diagnosis from medline: analytical survey. *BMJ* **330**(7501) (2005) 1162–3
6. Leeftang, M., Scholten, R., Rutjes, A., Reitsma, J., Bossuyt, P.: Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *Clin Epidemiol* **59**(3) (2006) 234–40
7. Aphinyanaphongs, Y., Tsamardinou, I., Statnikov, A., Hardin, D., Aliferis, C.: Text categorization models for high-quality article retrieval in internal medicine. *J Am Med Inform.* **12**(2) (2005) 207–16
8. Cohen, A., Hersh, W., Peterson, K., Yen, P.: Reducing workload in systematic review preparation using automated citation classification. *JAMIA* **13**(2) (2006) 206–19
9. Demner-Fushman, D., Few, B., Hauser, S., Thoma, G.: Automatically identifying health outcome information in medline records. *JAMIA* **13**(1) (2006) 52–60
10. Kilicoglu, H., Demner-Fushman, D., Rindfleisch, T., Wilczynski, N., Haynes, R.: Towards automatic recognition of scientifically rigorous clinical research evidence. *J Am Med Inform Assoc* **16**(1) (2009) 25–31
11. Condamines, A., Rebeyrolle, J.: CTKB : A corpus-based approach to a terminological knowledge base. In: Proceedings of Computerm'98 (First Workshop on Computational Terminology), Coling-ACL'98, Université de Montréal, Montréal, Quebec, Canada (1998) 29–35
12. Maedche, A., Staab, S.: Mining ontologies from text. In Dieng, R., Corby, O., eds.: EKAW 2000, Juan les Pins (2000)
13. Giraldo, G., Reynaud, C.: Construction semi-automatique d'ontologies à partir de DTDs relatives à un même domaine. In: Actes de la conférence Journées Francophones de l'Ingénierie des Connaissances (IC), Rouen (28-30 mai 2002)
14. Krivine, S., Nobécourt, J., Soualmia, L., Cerbah, F., Duclos, C.: Construction automatique d'ontologies à partir d'une base de données relationnelles : application au médicament dans le domaine de la pharmacovigilance. In: IC 2009. (2009)

15. Kamel, M., Aussenac-Gilles, N.: Construction automatique d'ontologies à partir de spécifications de bases de données. In: IC 2009. (2009)
16. van Heijst, G., Schreiber, A., Wielinga, B.: Using explicit ontologies in KBS development. *IJHCS* **46**(2-3) (1997) 183–292
17. Guarino, N.: Understanding, building, and using ontologies. *IJHCS* **46**(2-3) (1997) 293–310
18. Wang, Z., Wang, K., Topor, R., Pan, J.Z.: Forgetting concepts in DL-Lite. In Verlag, S., ed.: WWTP. (2008) 245–57
19. Qi, G., Haase, P., Huang, Z., Ji, Q., Pan, J.Z., Volker, J.: A kernel revision operator for terminologies - algorithms and evaluation. In: International Conference on The Semantic Web. (2008) 419–34
20. Gene Ontology Consortium: Gene Ontology: tool for the unification of biology. *Nature genetics* **25** (2000) 25–29
21. 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
22. Zhong, J., Zhu, H., Li, J., Yu, Y.: Conceptual graph matching for semantic search. In: In Proc. of 10th International Conference on Conceptual Structures, ICCS2002, LNCS 2393, Springer Verlag. (2002) 92–106
23. Sussna, M.: Word sense disambiguation for free-text indexing using a massive semantic network. In: Conference on Information and Knowledge Management. (1993)
24. Resnik, P.: Disambiguating noun groupings with respect to wordnet senses. (1995)
25. 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)
26. Hirst, G., Onge, D.S.: Lexical chains as representations of context for the detection and correction of malapropisms. The MIT Press (1) (1998)
27. Bousquet, C., Jaulent, M.C., Chatellier, G., Degoulet, P.: Using semantic distance for the efficient coding of medical concepts. In: Annual Symposium of the American Medical Informatics Association (AMIA), Los Angeles, CA (2000) 96–100
28. Lord, P., Stevens, R., Brass, A., Goble, C.: Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* **19**(10) (2003) 1275–1283
29. Caviedes, J.E., Cimino, J.J.: Towards the development of a conceptual distance metric for the umls. *Journal of Biomedical Informatics* **37** (2004) 77–85
30. Pedersen, T., Pakhomov, S.V., Patwardhan, S., Chute, C.G.: Measures of semantic similarity and relatedness in the biomedical domain. *J Biomed Inform* **40**(3) (2007) 288–99
31. Lee, W.N., Shah, N., Sundlass, K., Musen, M.: Comparison of ontology-based semantic-similarity measures. In: AMIA Annu Symp Proc. (2008) 3848
32. McInnes, B., Pedersen, T., Pakhomov, S.: UMLS-interface and UMLS-similarity : Open source software for measuring paths and semantic similarity. In: AMIA. (2009) 431–5
33. National Library of Medicine Bethesda, Maryland: Medical Subject Headings. (2001) www.nlm.nih.gov/mesh/meshhome.html.
34. NLM: Medline: medical literature on-line. National Library of Medicine, Bethesda, Maryland. (2009) www.ncbi.nlm.nih.gov/sites/entrez.
35. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. Seconde edn. MIT Press and McGraw-Hill (2001)