Medical Imaging Report Indexing: Enrichment of Index through an Algorithm of Spreading over a Lexico-semantic Network

Mathieu Lafourcade
LIRMM
University of Montpellier
France
mathieu.lafourcade@lirmm.fr

Lionel Ramadier
IMAIOS
34090 Montpellier
France
lionel.ramadier@lirmm.fr

Abstract

In medical imaging domain, digitized data is rapidly expanding Therefore it is of major interest for radiologists to be able to do an efficient and accurate extraction of imaging and clinical data (radiology reports) which are essential for a rigorous diagnosis and for a better management of patients. In daily practice, radiology reports are written using a non-standardized language which is often ambiguous and noisy. The queries of radiological images can be greatly facilitated through textual indexing of associated reports. In order to improve the quality of the analysis of such reports, it is desirable to specify an index enlargement algorithm based on spreading activations over a general lexical-semantic network. In this paper, we present such an algorithm along with its qualitative evaluation.

1 Introduction

Widespread digitalization in the health care sector and the implementation of customized electronic medical record result in a rapid increase in the volume of digital medical data. The medical computer systems allow to archive many and varied information (for example medical record, results of medical analyses, X-rays and radiological reports…). Thus, these data are accessible, either to be completed and compared with new results or to adapt the management of patients, or to provide decision support to improve the quality of care. The ability to have easily and efficiently access to these medical data has become a primary objective for health professionals. Thus, a proper indexing of medical reports (surgical, radiological...) optimizes the search for information, not only in a clinical purpose, but also educational. Dinh *and al.*, (2010) then realized a semantic indexing of patients' medical records in order to make it support for some information search procedures. Their indexing method uses MeSH (Medical Subject Headings), involves disambiguation, the extraction of clinical values, and weighting of concepts. Pouliquen, (2002) also performed an automatic indexation through recognition and extraction of medical concepts. He took into account the compound words and word associations to convert a sentence in reference words with the help of a medical thesaurus.

In the field of the medical imaging, the quantity of images and reports increases so much that being able to find quickly and easily the information becomes a major stake. But take full advantage of such a collection of radiological images means being able to quickly identify relevant information and requires that they are properly indexed from their reports. To be effective and useful for practitioners, indexing must consider their requests. Several authors, including Hersh et al., (2001) and Huang et al., (2003) automatically indexed radiological reports using the UMLS metathesaurus. To improve the accuracy of their results, they used a subsection of UMLS terminology, and Hersh *et al.*, (2001) deliberately choose not to include some parts of reports, especially the *indications* section. They thus obtained an index limited to strictly medical terms. However, in practice, in order to efficiently search, practitioners must have the possibility to specify in their requests not only medical specific terms (*digestive perforing, glioblastoma*), but also expressions, compound words, and circumlocutions of general sense (*skiing accident, breast disease in young women, hangman fracture, trauma of the lower limbs*).

Proceedings of Recent Advances in Natural Language Processing, pages 338–345,
Hissar, Bulgaria, Sep 7–9 2015.

Automatic extraction of relevant information from medical corpus is complex for several reasons: firstly, most texts are not structured and contain abbreviations, ellipsis and inaccuracies, on the other hand the amount of information to be analyzed is large and relevance is difficult to determine. The obstacles that hinder relevant indexing are of all kinds: difficulty of automatic semantic analysis (especially the precise analysis of negations, as shown by Huang *et al.,* (2007)), of identifying apocopes *(flu for influenza)* or unfamiliar terms (i.e. absent from the knowledge base), of recognition of medical entities often present in a distorted writing style, of extraction of semantic relations present in the text (Bundschus *et al.,* 2008), etc. To carry out a good indexing, it is crucial to have a knowledge base not only broad-spectrum (i.e. not limited to standardized forms) but also dynamic (i.e. able to evolve and enrich itself by permanent learning).

As far as we know, until now, the automatic indexing of radiological reports has concerned mainly medical terms without considering the general information. However, Xu *et al.,* (2014) were able to identify named entities of anatomical terms using some general resources like Wikipedia and Wordnet besides the usual medical resources i.e. UMLS, RadLex, MeSH et BodyPart3D (http://lifesciencedb.jp/bp3d/). Another type of resource, which had never been used in the medical or biomedical framework allows to consider not only the words and concepts of specialty, but also the common language used in reports (including the *indications* section). This is the lexical-semantic JeuxDeMots network (http://www.jeuxdemots.org) we use as a basis of knowledge and support for automatic indexing of radiological reports.

One objective of IMAIOS project (that we are conducting in collaboration with radiologists from Montpellier) is to achieve efficient indexing of radiological reports. To this end, we do not use only a description of the terms and concepts of specialty, but we are also working to determine the meaning and usage of terms and abbreviations very common in medicine. McInnes and Stevenson (2014) stressed the difficulty of indexing in the biomedical field, and Ramadier *et al.,* (2014) tries to make the task easier by using annotations and inferences from semantic relations. In this article we show how one can, from the semantic information of reports (in French), set an enlargement of raw built index to improve the recall of information retrieval. Indeed, radiologists may express their queries using generic terms (e.g. *benign brain tumor, brain tumor, benign tumor, tumor*) or consequences, or circumstances, etc. without these terms or expressions are explicitly present in reports. This semantic indexation may be also combined with the content-based image retrieval (CBIR) (Kurtz *et al.* (2014)).

In this article we first present the knowledge base used to achieve this indexing in French langage, i.e. the lexical network JeuxDeMots, then we describe precisely what an enlarged index relative to raw index is, and the index enlargement algorithm based on a spread over the lexical network. Finally we discuss experiments and analyze the results.

2 Index Enlargement and Spreading

The knowledge base on which our radiological reports indexing strategy relies is the lexical network JeuxDeMots (Lafourcade 2007). Although this network is general, it contains many specialty data, including medicine/radiology, which we have added within the framework of IMAIOS project. The network is the basis for a propagation algorithm that aims to increase the raw index obtained through conventional methods of information retrieval.

2.1 The JeuxDeMots Lexical Network

JDM network is a lexical-semantic graph for the French language whose lexical relations are generated both through GWAP (Games With A Purpose, see Lafourcade *et al*., (2015)) and via a contributory tool called Diko (manual insertion and automatic inferences with validations). At the time of this writing, the JDM network contains over 20 million relations between around 500,000 terms. The properties of this network that are important in the context of our work are the following:

- among about 80 lexico-semantic relations of the network, those which are relevant for our indexation project are the relations essentially semantic like hyperonymy, typical features, typical places, typical parts, target, etc.;
- polysemous terms are connected by the relation "refinement" with their various senses. About 9,000 polysemous terms are linked to approximately 25,000 meanings.

Figure 1. Screenshot of the contributory tool Diko showing the entry "tibia fracture". Diko is the online tool of visualization and of contribution of the JeuxDeMots lexical network. Note that the entry *tibia fracture* includes both specific medical relations (such as symptoms, diagnostic ...) and more general associations (such as causes, consequences, etc.).

For example, fracture → fracture (injury), fracture (break), fracture (sociology). The term in brackets is a *gloss* that allows to know or guess the meaning (refinement) of the polysemous word;

- relations are weighted, the weight reflects the strength of association between terms. Approximately 70,000 relations have negative weights, indicating a wrong relation (wrong relations are kept as they may be interesting within the framework of lexical disambiguation). An example is : *fracture du tibia *hypernym (< 0)* fracture (sociology: social dislocation) ;
- when a term *t* is associated with one of the meanings of a polysemous term, there is a relation of inhibition between the other meanings and the term *t*. For example: fracture *inhibition* talus (inclination), talus (printing), talus (embankment), astragale (architecture), astragale (botany), There is at least another meaning of talus or of astragale which are related to the term fracture: talus (os) and astragale (os).

The indexing of keywords in the medical field is often limited to certain aspects of a disease (Andrade, 2000) or to a part of the anatomy. But as the purpose of this indexation is to retrieve documents using also everyday language, we index not only anatomical terms (*knee, anterior wall of the colon, the genu of the corpus callosum,* ...), clinical signs (*plantar reflex*) and the names of diseases (*carcinoma*), but also everyday words (*fall in the bathtub*) likely to be used by the radiologist in his query.

The following table provides an order of size of the amount of information we have at our disposal about the specialty areas that are particularly relevant in the IMAIOS project:

Term	Outgoing links	Incoming links
medicine	21408	22666
anatomy	10477	11453
radiology	382	502
accident	741	956
medical imaging	541	556

Table 1: Number of relations of some key terms within the JDM lexical network.

2.2 Standard Indexing Report

Our corpus includes approximately 40,000 radiology reports (Example 1) concerning the different medical imaging techniques (MRI, scanner, ultrasonography, X-ray radiology, vascular radiology, scintigraphy ...). These reports are written in semi-structured way: they are generally divided into four parts (*indications*, *technique*, *results*, and an optional *conclusion*). Each part is written by the radiologist in a very free style, often with a profusion of acronyms (ATCD for of antecedent, ACR for American College of Radiology, tt for treatment, etc.), of elisions (*the anterior communicating* instead of *the anterior communicating artery*), and all sorts of various improprieties (*influenza* instead of *influenza virus*). Reports contain a lot of implicit information which need to be explicit to realize an indexation meeting the needs of practitioners. For instance it may be very interesting to explicit the expression *middle cerebral artery territory.*

The creation of the index starting from the reports is made by the traditional methods of information retrieval, i.e. term frequency (TF) and document frequency (DF) to calculate the IDF (Inverse Document Frequency). The identification of the compound terms is made upstream compared to the content of JeuxDeMots network. We use the underscore to separate the two parts of a compound word so that it is considered as an entity at the time of the extraction (*tibia_fracture*).

indications : fracture du tibia droit, chute de ski **technique** : une série de coupes axiales transverses sur l'ensemble de la cheville droite sans injection de produit de contraste **étude** : en fenêtres parties molles et osseuses. **résultats** : fractures diaphysaires spiroïdes à trois fragments principaux du 1/3 distal du tibia et de la fibula avec discret déplacement vers l'avant, sans retrait de refend articulaire. Fractures de la base de M2 et de M3 non articulaire et non déplacée. Fracture articulaire de la partie interne de la base de M1 non déplacée. Atrophie avec dégénérescence marquée des corps musculaires de l'ensemble des loges.	atrophie • cheville • chute • corps musculaire • coupe axiale transverse • dégénérescence • déplacement • fibula • fracture • fracture du tibia • loge • non articulaire • non déplacée • ski • spiroïde • tibia

Example 1: typical radiological report (left) and raw index (right). The raw index is the list of extracted terms, ordered alphabetically (the weights are not mentioned and the list is simplified). Compound words are extracted if they are present in the same form in the JDM network.

Despite the frequency filtering, we keep the words in the vicinity of medicine, even for low TF-IDF values. If a word of the report is connected to *medicine* (neighbor at a distance of 1) in the JDM network, then it is added to the index. In the same way, non-medical words (*motorcycle accident*, *influence of drugs*) are captured and added to the index if they are linked to a term itself linked to *medicine* (neighbor at a distance of 2): thus *motorcycle accident* is added because it is linked to *polytrauma* through the consequence relation and *polytraumat* is itself related to *medicine* through the field relation

Moreover, if a term of raw index is polysemous, it is interesting to try to determine the proper refinement: for example, in the above report the words *fracture (fracture), cheville (ankle), chute (fall)* and *loge (compartment)* are polysemous. We will see later that the identification of proper refinement is important. Moreover, the algorithm does not handle negation but in the phrase "*no articular fracture*", the term "*no articular"* will be detected because it belongs to the network JDM.

Thus, the *enlargement is a process intended for adding to the index some terms which are relevant, although they are not in the text.*

accident de ski • **accident de sports d'hiver** • atrophie • cheville • **cheville>anatomie** • chute • **chute>tomber** • corps musculaire • coupe axiale transverse • dégénérescence • **dégénérescence musculaire** • déplacement • fibula • fracture • **fracture articulaire** • **fracture des membres inférieurs** • **fracture multiple** • **fracture diaphysaire** • **fracture du tibia** • **fracture non articulaire** • **fracture non déplacée** • **fracture spiroïde** • **fracture avec déplacement** • **fracture>lésion** • **imagerie médicale** • **jambe** • **lésion** • **lésion osseuse** • loge • **loge>anatomie** • **médecine** • non articulaire • non déplacée • **péroné** • **radiologie** • ski • **spiroïde** • **sports d'hiver** • tibia • **traumatisme des membres inférieurs** • …

Example 2: Enlarged index corresponding to raw index above (the terms are listed alphabetically with the added words in bold). We can see that the general themes of the text are properly identified (*medicine, medical imaging, radiology*), and the polysemous terms were refined with the correct meaning depending on the context.

2.3 Enlargement through Spreading Algorithm

The enlargement strategy is to propagate signals originating from the terms of the raw index over the JDM network. The signal consists in "lighting up" the terms of the raw index within the network and retrieve related terms that light up in their turn.

At each iteration, the terms discharge their current activation to their neighbors. Thus, the total activation is none other than the sum of discharges received by a term during the entire process. For negatively weighted relations (i.e. inhibitory relations), the activation is removed instead of added. A term with negative CA cannot discharge. Iterated sequence is performed synchronously for all terms. Note that the distribution of the signal is proportional to the logarithm of the weight (and not proportional to the weight itself).

Specifically, we can describe the algorithm informally as follows:

Init: terms T of network are associated with a pair of values (CA, TA), *current activation* and *total activation*.

1 for the terms **T** belonging to raw index, we set CA = TA = 1. the terms **T** are sources of activation

2 for all other terms, AC= AT = 0.

3 we set a number of iterations NBI

4 we repeat NBI times the following operation :

5 for each term T of network having neighbors $\{t_1,\dots,t_n\}$ via a relation *r* from T to t_i with a positive weight w_i , we modify CA and TA of t_i :

$$CA(t_i) = CA(t_i) + CA(T) \times \frac{\log(w_i)}{\sum_{k=0}^{n} \log(w_k)}$$

$$TA(t_i) = TA(t_i) + CA(t_i)$$

// activation received by t_i is stored in $TA(t_i)$

6 $AT(\mathbf{T}) = 1$

// all T discharged their activation, we recharge the **T**

7 activated terms are filtered using a percentage of surface S; the remaining activated terms are returned.

Algorithm 1: calculation of an enlarged index starting from a raw index, using a propagation over the JDM lexical network. The two main parameters are NBI (number of iterations) and S (% of retained surface for the filter).

NBI \ S	10 %	20 %	30 %	40 %	50 %
1	22 / 82 %	45 / 80 %	67 / 78 %	93 / 53 %	127 / 38 %
2	31 / 95 %	55 / 92 %	83 / 89 %	211 / 57 %	439 / 41 %
3	48 / 99 %	90 / 97 %	139 / 95 %	356 / 53 %	755 / 34 %
4	111 / 97 %	223 / 92 %	335 / 87 %	747 / 45 %	1259 / 23 %
5	387 / 96 %	774 / 87 %	1161 / 76 %	1671 / 26 %	2089 / 15 %

Table 3 : The *nouv/pert* values depending on NBI and S parameters. NBI is the number of iterations performed in the lexical network. S is the retained part of the area under the curve of the cumulative weights of terms reached by the propagation algorithm.

After the iterations (lines 5-7), we obtain a weighted list of terms that are then ranked in order of decreasing weights. We retain by filtering N terms of the highest weight, such that the sum of their weights is S% of the total weight of the terms of the list.

We chose not to use all relations available in the lexical network JDM; indeed, some of them are too lexical: in the context of our work, they could degrade accuracy. We use the following relations (Table 2) (their relative importance, if different from 1 (default weighting) is indicated in brackets): associated ideas (weight of 1/2), hypernyms (weight of 2), synonyms, typical features, symptoms, diagnostics, parts/whole, typical place, causes, consequences, field, and frequently associated with. In the above algorithm, for simplicity, all relations are equally important (default weighting of 1, otherwise, their relative importance should be placed on both sides of the fraction).

r_associated	free associated terms
r_synonym	synonyms or quasi-synonyms
r_syn_strict	strict synonyms
r_isa	generic term
r_carac	typical characteristics
r_target	target of disease (people, organ etc).of diseases
r_symptoms	symptoms of diseases
r_location	typical location
r_cause	typical causes
r_consequence	typical consequences
r_accomp	what comes often with

Table 2: the relations through which the algorithm spreads to enlarge the index

3 Evaluation of Enlarged Index

We conducted a statistical evaluation of our propagation algorithm by randomly selecting 200 enlarged indexes (from a total of 30,000 calculated). We manually reviewed every term of the enlarged index to determine whether it was relevant or not. A *relevant term* is a term that is considered as appropriate for the description of the report. The presence of irrelevant terms increases the amount of noise when requesting documents. The absence of relevant terms decreases recall.

The pairs of values in Table 3 are *nouv/pert*, where *nouv* is the average number of terms of enlarged index that are not in the raw index and *pert* the average percentage of the relevant terms in enlarged index.

In practice, the *pert* value is assessed manually just once, regardless of the NBI and S parameters. Indeed we examine for each report all terms obtained in order of decreasing weights, for all parameter values, and then we assess the adequacy of each term. If we find a succession of 5 irrelevant terms, it is considered that the following are also irrelevant. The *nouv* value can be calculated automatically. For the same number of iterations, the greater the retained part is, the larger the number of terms is (low filtering). This means that if the recall is more important, in consideration accuracy tends to decrease (or even to collapse beyond 30%), because the terms added to the raw index are less and less relevant. Conversely, the more the number of iterations increases, the more the relevant terms are likely to be often reached and through multiple paths starting from the terms of the raw index, thus to be reinforced. The lexical network contains loops (direct and indirect) that act as self-reinforcement structures. The computation time dramatically increases with each new iteration, as the number of terms discharging their activation increases very strongly. For NBI = 5, almost the entire network is reached (if we exclude filtering S), its diameter being of about 6 (JDM is small-world

network). Overall the conditions that seem most interesting for a reasonable computation time (a few seconds) are 3-4 iterations and an area of less than 30%.

All ambiguous terms have correctly been disambiguated. This means that the enlarged index systematically included proper refinement when refinement was proposed (this is not necessarily the case for low values of NBI and S). If we recalculate the enlarged index while preventing access to refined terms, the *pert* value decline globally by 10%, regardless of the values of NBI and S. Try to select the correct meaning of ambiguous words can be carried out jointly with the selection of relevant terms and would even tend to favor it. Finally, all the specialty areas identified by the algorithm turned out relevant. Add the relevant specialties in the raw index before the enlargement process does not significantly improve the results (nor degrade them). Note that if we recalculate the raw and enlarged index while giving access, during propagation, only to immediate neighbors of the word *medicine*, whatever the relation, then the *pert* value decline in average by 12%. The use of a wide knowledge base, no limited to the only specialty field would thus improve largely the relevance of the produced index. Thus, the choice not to separate specialized and common vocabulary proves judicious and effective regarding the analysis of radiological reports. Moreover the algorithm is fast and well suited to handle the amount of data generated daily in radiology centers.

One can also notice that the overall process presented above works thematically on the text and semantically on the lexical network. A sharp semantic analysis of the reports would in all likelihood involve a chunk and dependencies analysis. The errors we found (23 terms for 200 indexes, which represent 23 errors for about 10,000 terms) may have different causes:

- lack of information in the knowledge base (20% of error cases);
- lack of semantic role, implying the need for a detailed analysis (55%);
- chimerism - two parts of the report have brought about an irrelevant term (25%).

As mentioned above, the network is characterized by a *never ending learning* approach (adding relations and refinement occurs permanently) in the spirit of Carlson *et al.* (2010). We can therefore reasonably hope that the knowledge base being constantly enriching, errors due to lack of knowledge will rapidly decrease over time. Similarly, the ability to identify the semantic relations and especially semantic roles within the reports would minimize the 2nd and 3rd source of errors.

4 Conclusion and prospects

Our objective is to automatically index radiological reports, using not only the medical terms but also words of common language that may be included in users' queries, especially hospital practitioners. To increase recall without significantly degrading the accuracy, we add in the raw index some implied words or expressions, using the JeuxDeMots lexical-semantic network as a support of knowledge. As far as we know, very few studies take into account the non-medical items in the radiological reports or carry out implicit inference for identify relevant terms. Conventional approaches to improve recall consist primarily of adding some terms more general (hyperonyms) starting from a medical ontology. But it is tangible that when information of general sense is present, the results are improved: the assumption that it would be better not to separate general knowledge and specialized one seems to be confirmed, at least in the context of our indexing works.

The results presented here are preliminary and require a substantive assessment of indexes on the whole corpus. These first results look promising, but we need to be able to automate the evaluation in order to do it on a larger scale. We could then further analyze the reports by seeking to extract relations between words using analogy/comparison with the relations of the JeuxDeMots lexical-semantic network. Another improvement would be to develop automated means for recognizing negation. So, the indexation would concern not only the terms, but also the semantic relations between them. One objective of the IMAIOS project is also to extract from medical reports some new knowledge to enrich the lexical network. Finally, we also plan to deduct from the corpus some rules of inference and thus make an authentic reasoning, i.e. to propose by deduction and by induction new medical information or even diagnosis.

References

Andrade M. A. and Bork, P. 2000. *Automated extraction of information in molecular biology.* FEBS letters, Elsevier, 476/1, pp. 12–17.

Bundschus M., Dejori M., Stetter M., Tresp V. and Kriegel H.-P. 2008. *Extraction of semantic biomedical relations from text using conditional random fields.* BMC bioinformatics, **9**:207, 14 p.

Carlson A., Betteridge J., Kisiel B., Settles B., Hruschka E. R., and Mitchell T. M. 2010. *Toward an architecture for never-ending language learning*. In AAAI, 2010, 8 p.

Dinh D., Tamine L. *et al.* 2010. *Vers un modèle d'indexation sémantique adapté aux dossiers médicaux de patients.* In Conférence francophone en Recherche d'Information et Applications, CORIA 2010, pp. 325–336.

Hersh W., Mailhot M., Arnott-Smith C. and Lowe H. 2001. *Selective automated indexing of findings and diagnoses in radiology reports.* Journal of biomedical informatics, 34(4), pp. 262–273.

Huang Y. and Lowe H. J. 2007. A *novel hybrid approach to automated negation detection in clinical radiology reports.* Journal of the American Medical Informatics Association, 14(3), pp. 304–311.

Huang Y., Lowe H. J. and Hersh W. R. (2003. A pilot study of contextual UMLS indexing to improve the precision of concept-based representation in xml-structured clinical radiology reports. Journal of the American Medical Informatics Association, 10(6), pp. 580–587.

Kurtz, C., Beaulieu, C. F., Napel, S., & Rubin, D. L. (2014). A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. *Journal of biomedical informatics*, *49*, pp. 227-244.

Lafourcade M. (2007). *Making people play for lexical acquisition with the JeuxDeMots prototype.* In SNLP'07 : 7th international symposium on natural language processing.

Lafourcade M., Le Brun N., Joubert A. (2015) *Games with a Purpose (GWAPS),* John Wiley & Sons, ISBN: 978-1-84821-803-1, 158 p.

Langlotz C. P. 2006. Radlex : A new method for indexing online educational material. Radiographics, 26(6), pp. 1595–1597.

Mcinnes B. T. and Stevenson M. 2014. *Determining the difficulty of word sense disambiguation.* Journal of biomedical informatics, 47, pp. 83–90.

Pouliquen B. 2002. *Indexation de textes médicaux par extraction de concepts, et ses utilisations.* Thèse de doctorat, Faculté de Médecine, Université Rennes 1, juin 2002, 163 p.

Ramadier L., Zarrouk M., Lafourcade M. and Micheau A. 2014. *Annotations et inférences de relations dans un réseau lexico-sémantique : application à la radiologie.* TALN 2014, Marseille, juillet 2014, pp. 103-112.

Ramos J. 2003. *Using TF-IDF to Determine Word Relevance in Document Queries.* In Proceedings of the first instructional conference on machine learning., 4 p.

Robertson S. E. and Jones K. S. 1976. *Relevance weighting of search terms.* Journal of the American Society for Information science, 27(3), pp. 129–146.

Robertson, S. 2004. "Understanding inverse document frequency: On theoretical arguments for IDF". *Journal of Documentation* **60** (5): pp. 503–520.

Xu Y., Hua J., Ni Z., Chen Q., Fan Y., Ananiadou S., Eric I., Chang C. and Tsujii J. 2014. *Anatomical entity recognition with a hierarchical framework augmented by external resources.* PloS one, 9(10), e108396.

Zarrouk M., Lafourcade M. and Joubert A. 2013. *Inference and reconciliation in a crowdsourced lexical semantic network.* Computación y Sistemas, 17(2), pp. 147–159.