Inferring Relations and Annotations in Semantic Network: Application to Radiology

Lionel Ramadier[1,2], Manel Zarrouk[1], Mathieu Lafourcade[1], and Antoine Micheau[2]

[1] LIRMM, Montpellier,
France

[2] IMAIOS, MIBI, Montpellier,
France

{lionel.ramadier, manel.zarrouk, mathieu.lafourcade}@lirmm.fr,
antoine.micheau@imaios.com

Abstract. Domain-specific ontologies are invaluable despite many challenges associated with their development. In most cases, domain knowledge bases are built with very limited scope without considering the benefits of plunging domain knowledge to a general ontology. Furthermore, most existing resources lack meta-information about association strength (weights) and annotations (frequency information like *frequent*, *rare,* etc. or relevance information like *pertinent* or *irrelevant*). In this paper, we present a semantic resource for radiology built over an existing general semantic lexical network (JeuxDeMots). This network combines weight and annotations on typed relations between terms and concepts. Some inference mechanisms are applied to the network to improve its quality and coverage. We extend this mechanism to relation annotation. We describe how annotations are handled and how they improve the network by imposing new constraints especially those founded on medical knowledge.

Keywords. Relation inference, lexical semantic network, relation annotation, radiology.

1 Introduction

For more than two decades, medical practice and biomedical research have benefited from the availability of biomedical ontologies [1]. These resources are used for various kinds of semantic analysis such as entity recognition (i.e., identification of biomedical entities in texts as name of genes, diseases, etc.), and relation extraction (i.e., identification of semantic relationships among biomedical entities, for instance, interaction between proteins). In the framework of the UMLS project, which interrelates some 60 controlled vocabularies, upper-level ontology, the UMLS semantic network [2] has been built. In the field of radiology, such a semantic network is used to facilitate or automate the analysis of radiologist reports in order to extract recommended courses of action or to trigger warning systems to improve patient management [3].

There exist reference ontologies in biomedical domain (UMLS), but they might not be suited to a particular domain like radiology because result sets are too large, too complex [4]. To solve this problem, the Radiology Society of North America (RSNA) has created a reference ontology for radiology, RadLex [5]. RadLex and its derivatives rely on English and are not considered medically complete [6]. There is a German version of RadLex [7] but no version exists in French, to our knowledge. More importantly, in the domain of radiology, the relationship between terms is crucial and the ontology model might not capture this information as well as a semantic network.

The ontology indicates generally only the hierarchy between terms and lacks specific relations relevant either to medicine or to how doctors express their knowledge in reports. When making clinical diagnosis based on a radiologist report, it is crucial for the medical practitioner to be presented with information from many different non-hierarchical sources; however, it is not so important for her to know the exact hierarchy of a

Computación y Sistemas Vol. 18, No. 3, 2014 pp. 455-466
ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2024

particular term (as this information is already known beforehand). For example, it is important to give an exhaustive list of symptoms or characteristics of a disease that the medical practitioner should look for regardless of the taxonomic hierarchy associated with each term. These terms can be better linked when modeled by a semantic network and, even better, by a lexical semantic network taking into account facts of medical language. While general-purpose semantic networks will certainly help, they need to be extended to specific domains such as radiology.

The combined method of modeling is important for radiology reports because such reports contain several distinct sections. In the *History* section for example, there are typically descriptive texts written in everyday language, while in other sections, such as *Findings*, the language changes to specialized terms. The goal of the construction of this lexical network is to analyze radiological reports in order to extract terms and relations between them. The aim is to carry out a semantic annotation of medical images to improve their retrieval.

Lexical semantic networks can be manually constructed or generated by algorithmic analysis of texts. For instance, the ConceptNet, a freely available general knowledge base, is generated automatically from the 700 000 sentences of the Open Mind Common Sense Project [8]. However, fully automated generation is generally limited to term co-occurrences since extracting precise semantic relations between terms from corpora remains difficult.

In our combined general purpose-specialist network, we decided to use JeuxDeMots [9] as a basis for the general-purpose network. What we wish to have is a general knowledge base. JeuxDeMots relies on crowdsourcing to construct manually a knowledge base. For this purpose, JeuxDeMots provides a contributive tool called Diko. This tool is important because we can use it to improve the network completeness in specific areas where the game approach is not suitable (relations too complicated, not lexicalized enough). Diko also exploits an inference mechanism [10] to automatically propose relations between terms on the basis of what already exists in the network. This approach of inference is strictly endogenous, as it does not rely on any external resources. JeuxDeMots uses crowdsourcing to attribute incrementally weights to relations between terms. If a large number of users/players associate two given nodes, the weight will be higher than another link mentioned by fewer users. While this user provided weight is adequate for general purposes, it fails in the diagnostic purpose of radiology reports because the overall frequency of a symptom is not a good indication of its relevance. For example, in a clinical situation, many patients complain of a headache and almost none report *arm drift* before suffering a stroke, but *arm drift* is the most important term in this case.

Generally, there is not always a correlation between the associative strength between two terms and its importance. The arc weight indeed implements the associative strength but it correlates neither to the truth nor to the frequency. The medical significance of the relationship should be indicated to generate faithfully this specialist radiology semantic network. The goal of our current work is to develop the cost function that best captures this medical significance and then to train the semantic network through inference mechanisms. We introduce annotation between some relations in the field of radiology in the semantic lexical network. The goal of the relation annotation is to guide the process of inference and semantic analysis.

The rest of the paper is organized as follows. In Section 2, we describe the principles behind the lexical network construction and illustrate them with JeuxDeMots. We also discuss the building of a network specialized in radiology. We present one type of inference: the deduction scheme. In Section 3, we describe the annotation of the relations between medical terms. Section 4 is devoted to our experiments and commenting on their results. Section 5 concludes the paper and points to avenues of future research. This paper is an extended version of [12].

2 Lexical Networks

The type of lexical network we are working with is a *graph with lexical items or concepts as nodes*

Fig.1 Example of network for the word *MRI*

connected through arcs interpreted as relations between items. Those relations are semantically typed and represent (typical) lexical or ontological relationships possible between terms (hypernym, synonym, antonym, part of, cause, consequence, typical location, telic role, semantic role, characteristic, etc.). Besides being typed, relations are weighted and directed (no automated symmetrization is undertaken). The contributive approaches for building such a network are more and more popular because they are both cheap to set up and efficient in quality. In recent years, there is an increasing trend of using on-line GWAPs (games with a purpose) [11] for feeding such resource. The JDM lexical network is constructed through a set of on-line associate games and contributive tools. There exist 10,173,073 relations and 368,604 terms in JDM. We use this GWAP (and also a contributive tool) to build our domain specific ontology. For the word *medicine*, there are ca. 11,000 relations in the database. In the network, about 350 relations exist for the word *MRI* (Figure 1, in French). We briefly describe it in the following section.

2.1 The JDM Game Model

JeuxDeMots is a two player blind game based on agreement on term associations. At the beginning of a game session, the player is given an instruction related to a target term (for example, *give any term that is related to **disease***). The user has a limited time to enter as many propositions as possible. At the end of the allowed time, player's proposals are compared to those of another player in the same game, and points are earned on the basis on the common proposals.

Terms in agreement are added to the lexical network with the relation corresponding to the game instruction. If the relation already exists, its weight is increased, otherwise this relation is added. This game is adequate for general common sense knowledge but may be not very efficient for specialized domain. For building our project, a lexical network for radiology, we use a contributive tool compatible with the JDM lexical network, named Diko, which we explain briefly in the next sub-section.

2.2 The Contributive Model of Diko and Relation Annotations

Diko is a web-based tool for displaying information contained in the JDM lexical network but that can also be used for contribution. The necessity to not be only dependent on the JDM game for the construction of the lexical network comes from the fact that many relation types of JDM are either difficult to grasp for a casual player or not very productive (possibly not many answers). In order to build a specialized knowledge base we use Diko to propose new relations between terms relevant to the domain at hand. The principle of the contribution process is that a proposition made by a user will be voted pro or con by a validator who is expert in radiology. In the field of medicine, we added some relations like *symptom* or *diagnostic*.

However, in the case of most generic concepts (*location*, *has_part*, etc.), it was possible to connect them with the radiological concepts. This contributive work is necessary for building a knowledge substrate for radiology, and eventually, the purpose of the project is to extract in a semi-automatic way words and relations from the radiology reports to enhance the specialized network.

The building of imaging medical network was realized from 40,000 radiology reports representing the different medical imaging techniques (MRI, CT-scan, ultrasound, etc.). The

méningiome Nom, Nom masculin singulier Lemme méningiome [§] Informations diverses wiki

Associations d'idée 35 tumeur - tumeur bénigne - maladie » - maladie (médecine humaine) - IRM » - méninge - maladie neurologique - tumeur bénigne du système nerveux - système nerveux central - tumeur bénigne du cerveau - cerveau » - pathologie - médecine - angiome - opération » - scanner » - tumeur extra-axiale - tumeur cérébrale - scanner (médecine) » - opération (intervention chirurgicale) - bénin (médecine, non cancéreux) - gadolinium - cerveau (anatomie) - imagerie par résonance magnétique - tumeur du système nerveux central - tumeur extra axiale - tumeur du cerveau - tumeur intracrânienne - tumeur des méninges - tumeur bénigne du système nerveux central - tumeur bénigne de la moelle épinière - tumeur maligne du système nerveux central - tumeur maligne - tumeur de la moelle épinière - encéphale 14 maladie neurologique - tumeur extra-axiale - tumeur bénigne du système nerveux - tumeur du système nerveux central - maladie » - tumeur bénigne du cerveau - tumeur bénigne - méninge - tumeur - maladie (médecine humaine) - tumeur bénigne du système nerveux central - tumeur cérébrale - tumeur des méninges - tumeur cérébrale bénigne

Thèmes/domaines médecine - pathologie - neurologie

Génériques H 9 tumeur bénigne du cerveau [fréquent] - tumeur bénigne de la moelle épinière - tumeur - tumeur du cerveau - maladie » - maladie neurologique - tumeur bénigne du système nerveux [abus de langage] - maladie (médecine humaine) - tumeur bénigne - tumeur du système nerveux central - tumeur maligne [rare] - tumeur bénigne du système nerveux central [fréquent] - tumeur des méninges - tumeur maligne du système nerveux central [rare] - tumeur intracrânienne - tumeur cérébrale - tumeur de la moelle épinière - tumeur extra axiale - tumeur extra-axiale

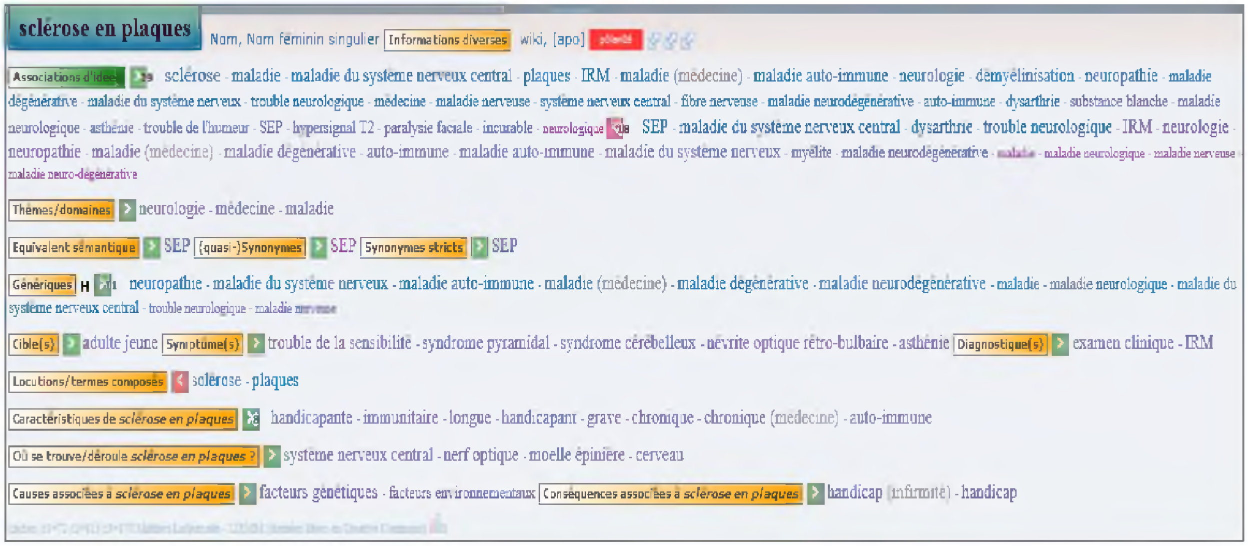
Symptôme(s) crise d'épilepsie Diagnostique(s) scanner » - IRM »

Caractéristiques de *méningiome* bénin » - grosse (gros) - virale - grosse » - cancéreuse - bénin (médecine) » - asymptomatique [possible]

Qu'est-ce qui s'oppose à *méningiome* ? radiothérapie - chirurgie

Où se trouve/déroule *méningiome* ? système nerveux central - encéphale - moelle épinière - cerveau (anatomie) - tête » - système nerveux - corps » - cerveau » - corps humain

Fig. 2. Example of the term *meningioma* with annotations in brackets. Several annotations are possible for a given relation like *frequent*



sclérose en plaques Nom, Nom féminin singulier Informations diverses wiki, [apo]

Associations d'idée 39 sclérose - maladie - maladie du système nerveux central - plaques - IRM - maladie (médecine) - maladie auto-immune - neurologie - démyélinisation - neuropathie - maladie dégénérative - maladie du système nerveux - trouble neurologique - médecine - maladie nerveuse - système nerveux central - fibre nerveuse - maladie neurodégénérative - auto-immune - dysarthrie - substance blanche - maladie neurologique - asthénie - trouble de l'humeur - SEP - hypersignal T2 - paralysie faciale - incurable - neurologique 18 SEP - maladie du système nerveux central - dysarthrie - trouble neurologique - IRM - neurologie - neuropathie - maladie (médecine) - maladie dégénérative - auto-immune - maladie auto-immune - maladie du système nerveux - myélite - maladie neurodégénérative - maladie - maladie neurologique - maladie nerveuse - maladie neuro-dégénérative

Thèmes/domaines neurologie - médecine - maladie

Equivalent sémantique SEP (quasi-)Synonymes SEP Synonymes stricts SEP

Génériques H 11 neuropathie - maladie du système nerveux - maladie auto-immune - maladie (médecine) - maladie dégénérative - maladie neurodégénérative - maladie - maladie neurologique - maladie du système nerveux central - trouble neurologique - maladie nerveuse

Cible(s) adulte jeune Symptôme(s) trouble de la sensibilité - syndrome pyramidal - syndrome cérébelleux - névrite optique rétro-bulbaire - asthénie Diagnostique(s) examen clinique - IRM

Locutions/termes composés sclérose - plaques

Caractéristiques de *sclérose en plaques* handicapante - immunitaire - longue - handicapant - grave - chronique - chronique (médecine) - auto-immune

Où se trouve/déroule *sclérose en plaques* ? système nerveux central - nerf optique - moelle épinière - cerveau

Causes associées à *sclérose en plaques* facteurs génétiques - facteurs environnementaux Conséquences associées à *sclérose en plaques* handicap (infirmité) - handicap

Fig. 3. Example of the term *multiple sclerosis* which has as causes *genetic factors* and *environmental factors* annotated as *possible but uncertain* (Figure reproduced from [12])

first step consisted in accomplishing an inverted index of bigram, trigram and quadrigram. At the second step, the expert submitted a term or concept that she considered relevant for the process of validation/invalidation to the other domain specialist.

To improve the quality of the network, we add more medical significance for relationships between terms with the help of annotations. For instance, for the relation *meningioma* (*is-a*) *benign central nervous system neoplasm* we can add the annotation *frequent* regardless of the

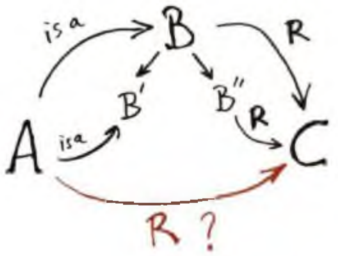


Fig. 4. Informal deductive inference scheme with logical blocking. If A is a B and B has some relation R with C, then it is expected that A has the same relation R with C. However, if B is polysemous, and two different refinements (B' and B'') hold the premises, then the relation A *R* C is most probably wrong (Figure reproduced from [12])

weight of the relation (Figure 2). Another example is given in Figure 3. In Section 3, we present in detail the concept of annotation and its utility.

In order to increase the number of relations in the JDM network, an inference engine has been proposed. The latter functions as if it were a contributor suggesting relations to be validated by human contributors or experts in the case of specialized knowledge. In this paper, we describe one type of inference: the deduction scheme.

The deduction scheme is based on the transitivity of the ontological relation *is-a* (hypernym). If a term A is a kind of B and B has some relation R with C (the premises), then we expect that A holds the same relation with C (the conclusion). The inference engine can be applied on terms having at least one hypernym. If a term has a set of weighted hypernyms, the inference engine deduces a set of inferences. This hypernyms are classified according to the hierarchical order. The weight of a proposed inference is the incremental geometric mean of each occurrence. In fact, this scheme is too simple; in effect, the term B may be polysemous, so a probably wrong inference can be avoided by logical blocking (Figure 4). This mechanism was described in previous work [10].

In case of invalidation of an inferred relation, a reconciliator is invoked to try to assess why the inferred relation is wrong. The reconciliation allows us to identify the cause of a wrong inference: an exception, an error in the premises or transitivity confusion due to polysemy with the identification of the proper word senses at stake.

In what follows, we present the type of inference which we are going to consider (concerning annotation). Nevertheless, there are two other types of inference: induction (from specific to general) and abduction (imitation from examples). However, annotation propagation for these types of inference is quite hazardous.

3 Relation Annotations

In all generality, above all in specialized knowledge, the correlation between the weight of the relation and its importance is not strict. In the case of *hepatocellar carcinoma* the relation with *wash-out* is specific of radiology so the weight of the relation will be low but for the radiologist this relation is important. This is why it appears interesting to introduce annotations for some relations as they can be of a great help in the medical area.

3.1 Relation Types for Radiology

In the lexical network, a relation is formally represented by a 3-tuple:

<$\text{Node}_{\text{start}}$, Relation type/annotation, Node_{end}>

is formally written as

$\text{Node}_{\text{start}}$ (Relation type/ annotation) Node_{end}.

Concerning the field of radiology, the most useful relations which were established by radiologists following their daily clinical practice are shown in Table1.

In radiological ontology like RadLex, there are not so many relations that can useful in the analysis of radiological reports. In information retrieval, this annotation can be helpful to the users. Often, they want to know if a characteristic of a given pathology is rare or frequent. This kind of information is generally absent from a network or ontology. For example, the relation between *measles* and *children* are frequent and this information will be directly available in the network (Figure 5).

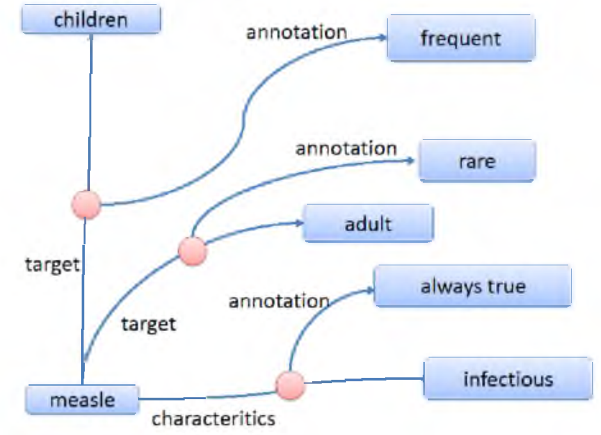


Fig. 5 Relation annotation implementation. A given relation to be annotated is reified (represented by a specific node, here with red circles) and this node is associated to various annotations and any other regular terms. The annotation relation type is a kind of relation among others.

3.2 Annotations Values

These annotations will have a filter function in the inference scheme. The types of annotations are of various nature (mostly frequency and relevance information). Now we present the different main annotation labels:

- frequency annotations: *very rare*, *rare*, *possible*, *frequent*, *always true*;
- usage annotations: *often believed true*, *language misuse*;
- quantifier: any number like 1, 2, 4, etc. or *many*, *few*;
- qualitative: *pertinent*, *irrelevant*, *inferable*, *potential*, *preferred*.

Concerning *language misuse*, a doctor can use the term *flu* (illness) instead of *virus of influenza*: it's a misuse of language as the doctor just makes use of a language shortcut. The annotation *often believed true* is applied to a wrong relation (with a negative weight) which is very often considered as true, for example, *spider* (**is-a/often believed true*) *insect*. This kind of annotations could be used to block the inference scheme.

Qualitative annotation relates to the inferable status of a relation, especially concerning inference. The pertinent annotation refers to a proper ontological level for a given relation. For example: *living being* (*carac/pertinent*) *alive* or *living being* (*can/pertinent*) *die*. Another example refers to synonyms; in this case, it may be relevant to choose a preferred synonym, according to the advice of an expert. For instance, a group of synonyms can be presented as *hepatocellular carcinoma* (*preferred*), *HCC*, *malignant hepatoma*.

The annotation **inferable** is supposedly to be put when a relation is inferable (or has been inferred) from an already existing relation, for example: *dog* (*carac/inferable*) *alive* because *dog* (*is-a*) *living being*.

The annotation **potential** may be put for terms above the pertinent ones in the ontological hierarchy, for example: *bird* (*has-part/always true*) *wings* and *animal* (*has-part/potential*) *wings*. Finally, the annotation **irrelevant** is put for a true relation that is considered as too far below the pertinent level, for example, *animal* (*has-part/irrelevant*) *atoms*.

The **quantifier** represents the number of parts of an object. Each human has two lungs so the quantifier is 2. This kind of annotation is not necessarily a numeral, but can be of more or less subjective value, like *few*, *many*, etc.

The annotation **frequency** is of five types (*always true*, *frequent*, *possible*, *rare* and *exceptional*), and there are two qualitative types (*pertinent* and *irrelevant*). We attributed empirical values to each annotation's label like 4 to *always true*, 3 to *frequent*, 2 to *possible*, 1 to *rare* and 0 to the rest of the annotations. These allow us to select some annotations to facilitate or block the inference scheme.

The first annotations have been made by hand, but with the help of the inference scheme, they will spread through the network. To improve the quality of the network and to prevent some incoherent inferences, some kind of annotation should block the potentially absolute relations. For instance, the annotation *language misuse* or *irrelevant* will block the inference scheme.

The annotation **possible** is a special case. Depending of the configuration of the system, it may block (stricter approach) or not block (lenient

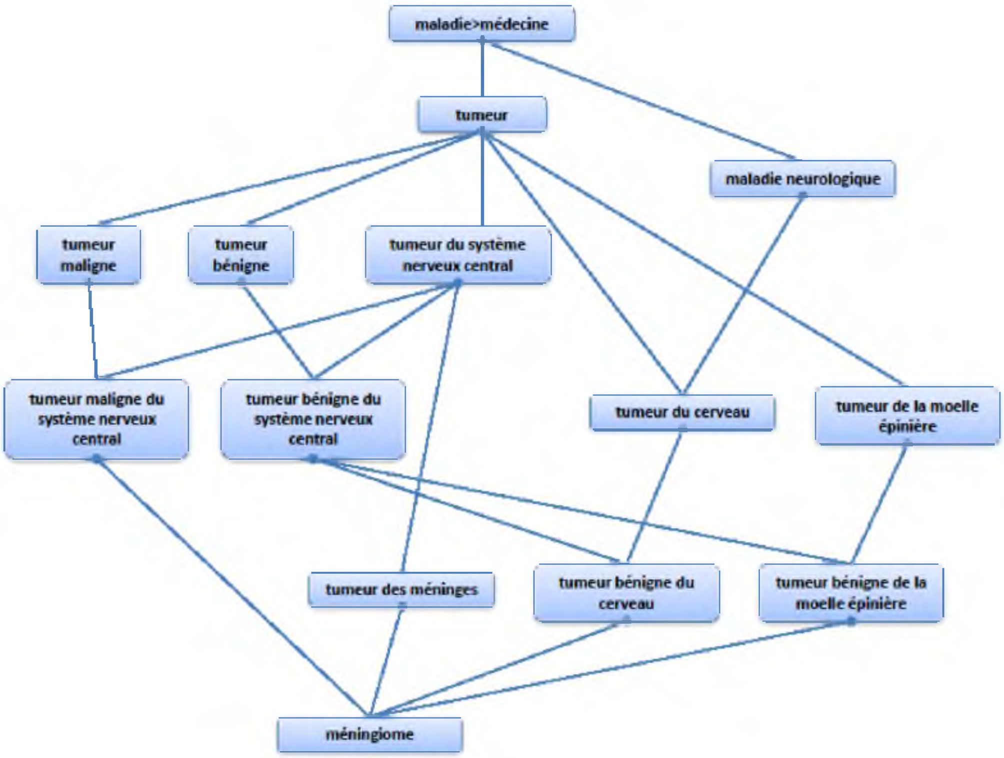


Fig. 6. The hierarchy implicitly described for the term *méningiome*. Each term of the hierarchy is included in at least one path (see path lists in the text of Section 3.3). The hierarchy can be partially deduced from the paths

approach) the inference mechanism. If a system is lenient, we may obtain many inference proposals that might be wrong (high recall, low precision). On the other hand, if a system is strict, we reduce the risk of wrong proposals, but at the cost of missing adequate ones (low recall, high precision).

3.3 Using Generic Partial Orders

Moreover, to have the most accurate annotation, we need to order the central terms from the most specific terms to the less specific ones as we did in [12]. That is, we try to reconstitute the taxonomic order related to the hypernym relation (*is-a*). Here we explain the algorithm in detail. The Generic Paths Algorithm has been designed for this purpose.

For example, for the term *hepatocellular carcinoma* the (several) partial orders for its hypernyms will be

- *hepatocellular carcinoma → malignant tumors of liver → tumor of liver → liver pathology → pathology*
- *hepatocellular carcinoma → malignant tumors of liver → tumor of liver → tumor → pathology*

Another example, with the term *méningiome* (*meningioma*, in French):

- *méningiome → tumeur bénigne du cerveau → tumeur bénigne du système nerveux central → tumeur bénigne → tumeur → maladie>médecine*
- *méningiome → tumeur bénigne du cerveau → tumeur du cerveau → maladie neurologique → maladie>médecine*

- *méningiome → tumeur bénigne de la moelle épinière → tumeur de la moelle épinière → tumeur du système nerveux central → tumeur → maladie>médecine*
- *méningiome → tumeur des méninges → tumeur du système nerveux central → tumeur → maladie>médecine*
- *méningiome → tumeur maligne du système nerveux central → tumeur maligne → tumeur → maladie>médecine*

The annotation will be different depending on the term at hand. To choose a proper annotation (or several annotations) for a newly inferred relation, this order plays an important role. The annotation, if any, of the most specific term is more crucial than that of less general terms (at higher levels of the hierarchy).

We must take into account this fact and exploit it when dealing with the inference mechanism with annotations.

The idea of the Generic Paths Algorithm is to produce ordered lists of generic words from a given term. The lists are ordered from the most specific term (in fact, the target term) to the most general one. When produced, these lists are useful for various kinds of propagation algorithms, amongst which is the one we use for propagating annotations.

First, an unordered list of generic terms is extracted from the lexical network.

For the previous example (*mkéningiome*), the raw list of generic terms (RT) is the following: *tumeur*, *maladie>médecine*, *tumeur bénigne*, *maladie neurologique*, *tumeur bénigne du cerveau*, *tumeur du cerveau*, *tumeur cérébrale*, *tumeur intracrânienne*, *tumeur bénigne du système nerveux central*, *tumeur bénigne de la moelle épinière*, *tumeur du système nerveux central*, *tumeur des méninges*, *tumeur maligne du système nerveux central*, *tumeur maligne*, *tumeur de la moelle épinière*.

This list has first to be filtered in order to keep only one term in case of orthographic variants and strict synonyms. In the previous example, *tumeur du cerveau* and *tumeur cérébrale* are strict synonyms, and only one of this is kept as representative for producing paths. We note this list of filtered terms as *FT*.

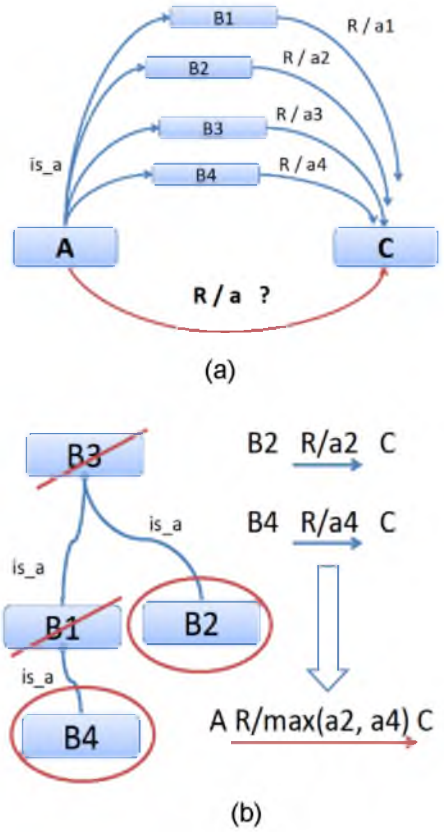


Fig. 7. Use of the hierarchy to select the most accurate annotation to link to an inferred relation via several central terms. The most specific terms are selected and the max rule applied

The second step is to order terms as if they all follow an order relation. Note that the produced result is a set of paths $P = \{p_1 ... p_n\}$ where p_i stands for a path. A path is an ordered set of terms that belong to *FT*. We initialize *P* to the empty set.

For each $t \in FT$, we insert it at its proper place in each *p* of *P* (like in any sorting algorithm by insertion). If *t* cannot be included, we add a new list to *P* ($P = P \cup \{t\}$). In a given path *p*, a term *t* can be inserted between two consecutive terms t_a and t_b if and only if (test function) $t_a < t$ and $t < t_b$, where $x < y$ means that *y* is a generic term of *x*.

When all terms t have been added to one or several paths of *P*, the algorithm ends, and the result is *P*. Again, *P* is not the definition of the hierarchy graph, but a set of paths (ordered sequences of terms) that completely cover the graph (no term is left out).

The process has a cubic complexity in the number of terms, considering the test function as basic operation. This polynomial complexity is definitively manageable in practice as the actual number of generic terms is rarely over 100. Hence, the hierarchy paths can be computed on the fly, and barely need to be stored in the database.

Even for monosemic terms, most of the time, multiple views related to the term lead to a hierarchy, which takes the form of a direct acyclic graph and is always not a linear chain of generic terms. The term *méningione*, which is monosemic, is a typical example.

In the inference mechanism, the term B (central term in Figure 4) plays a crucial role. We look at the hierarchy of the terms B according to which a specific relation was inferred many times and we keep the most specific one. If we end up with two or more terms, we apply the max rule to the values corresponding to each annotation. The result will be the value of the annotation we will assign to the inferred relation (Figure 7a, b).

4 Experimentation

In some previous experiments [10], the deduction engine was applied to the whole lexical network in order to assess the approach efficiency. In this paper, we unleashed the experiment on a subset of the lexical network JDM that contains all the hypernym relations (*generic*/*is-a*) and is based on the deduction scheme and all manually annotated relations. This reduction allowed us to diminish drastically the size of the search space.

4.1 Unleashing Relation Inference

To increase the accuracy of the result and to avoid inference of noisy relations, we blocked inferences on relations that are annotated as irrelevant, or *exception*.

Moreover, more detailed results and experiments with the deduction engine are provided in [10]. The deduction inference engine was applied on around **150,000** relations and produced over **2 million** relations, **700,000** of them were distinct, which makes an average of 3 occurrences per relation (see Table 1).

Table 1. Number of inferred relations from those already existing ones

Existing relations	153,765
Inferred relations	2,123,533
Distinct inferred relations	729,510

4.2 Relation Annotation Propagation

The annotation inference engine is applied as the second part of the system. Therefore, it will be unleashed over the base of relations previously enriched with the deduction engine. The relation annotation system runs only on the inferred relations, it takes in consideration the annotations of the premises used to infer a certain relation as mentioned. If there is just one premise, the annotation of this premise, if any, is applied to the inferred relation. If there are many premises, the system will rebuild the hierarchy between these ones and will keep the annotation of the nearest premise as the most accurate one. In case of having a number of premises with the same level in the hierarchy, a maximum rule is applied on them and the annotation having the strongest number (*always true*: 4, *frequent*: 3, *possible*: 2, etc.) will be applied to the inference. This system guarantees a good accuracy of the annotation spreading.

As noticed, contrary to the original deduction engine, here we allowed redundancy because it increases the accuracy of the results of the relation annotation spreading system.

To clarify, we give an example as follows.

Premises: *stroke* (*is-a*) *cerebral infraction* **&** *cerebral infraction* (*diagnosis*/*frequent*) *MRI*.

Inferred relation: *stroke* (*diagnosis*/*frequent*) *MRI* (1).

Premises: *stroke* (*is-a*) *cerebrovascular disease* **&** *cerebrovascular disease* (*diagnosis*/*possible*) *MRI*.

Inferred relation: *stroke* (*diagnosis*/*possible*) *MRI* (2).

The annotation system having two occurrences (1) and (2) of the same relation *stroke* (*diagnosis*) *MRI*, annotated differently (*possible*, *frequent*) will decide to keep the strongest one (*frequent*). It is informed about the annotation strength by empirical values we have attributed to each annotation label according to their frequency: 4 to "*always true*", 3 to "*frequent*", 2 to "*possible*", 1 to "*rare*" and 0 to the rest of the annotations.

The annotation inference system applied on the relation base stemmed from the deduction engine run annotated more than 10,000 relations starting from only 100 ones (Table 3). The ratio between existing relations and produced ones at this stage is over 1 to 100.

In this experiment, we did not consider potential and inferable annotations (more than 43,000 distinct annotations for one unique run, 97% are correct and 3% false) because they are more utility annotations than semantically relevant ones in the context of radiology. Instead, we focused here on the annotations illustrating frequency since it is very important information in the radiological area.

The number of annotated relations per annotation label does not depend on the initially existing number, as it can be noticed in Table 1, but simply on the number of the ongoing hypernym relations of the central term of the scheme as in the simplified example we are giving now.

The basic inference scheme is the following:

A (is-a) B & B (R/annot) C → A (R/annot) C.

Example:

non-small-cell-lung carcinoma
hepatocellular carcinoma } (*is-a*) *malignant tumor*
glioblastoma
& *malignant tumor* (*carac*/*frequent*)
poor pronastic

Three relations annotated as *frequent* (*non-small-cell-lung carcinoma* / *hepatocellular carcinoma* / *glioblastoma* (*carac*/*frequent*) *poor pronastic*).

The bigger the number of hypernym relations toward the term B (*malignant tumor*) which has an outgoing relation annotated (*malignant tumor* (*carac*/*frequent*) *poor pronastic*), the bigger is the number of annotated relations.

Table 2. Number of annotations inferred after the application of the relation annotation system on the existing ones

Annotation's Label	Existing annotation	Inferred annotation
Frequency: ***frequent* & *always true***	38	8,709
Frequency: ***possible***	16	172
Frequency: ***rare* & *very rare***	7	42
Qualifier: ***often believed true***	1	9
Qualifier: ***irrelevant***	27	2,387
Quantifier	12	178
Total	103	11,497

However, for the existing annotated relations that do not contribute a lot in the inferring process, as the annotation *frequent* (Table 2), they are attributed to relations which are ineligible to the annotation scheme which is deductive, for example:

- *Hepatocellular carcinoma* (*carac*/*frequent*) *hypervascular*

The term *Hepatocellular carcinoma* does not have any ongoing hypernym relation (x (*is-a*) *hepatocellular carcinoma*), so in this case the annotation *frequent* will not generate other annotations.

We statistically evaluated the produced annotations, and it appears than 87% of them were evaluated as "correct", 5 % as "incorrect" and the rest (8 %) as "debatable" (it means that experts might discuss its validity but rather if the frequency value should be modified).

In this experiment, we applied the relation/annotation system through a single run. However, naturally, the system, which is actually running iteratively along the contributions and games, uses the new terms and annotations added and the previously inferred ones to continue annotating more relations.

Table 3. Relevant relations in the radiology field with explanation, examples, and their annotations

Relation type	Explanation, examples and typical annotations
is-a	Hypernym, *MRI* is-a *medical imaging* (*possible*)
has-part	Element of the term, *liver* has-part *segment I (always true)*
characteristic	*Hepatocellular carcinoma* carac *hypervascular* (*frequent*)
typical location	Typical place where the term/object in question can be located, *multiple sclerosis* typ location *central nervous system (always true)*
target	Population affected by the term, *measles* target *children (frequent)*
diagnosis	Examen, *multiple sclerosis* diag *MRI* (*frequent*, *crucial*)
symptom	Symptom, *measles* symptom *fever* (*frequent*)
against	What the start term opposes/fight/prevents*, malignant tumor* against *chemotherapy (frequent)*
cause	B(that you have to give) is a cause of A, *cirrhosis* cause *alcoholism (frequent)*
consequence	The end term is a possible consequence of the start term*, stroke* consequence *hémiplegia (possible)*

5 Conclusions

In this article, we presented some issues concerning relation annotation in the context of 1) building a lexical semantic network with games and user contributions and 2) inferring new relations from existing ones. To be able to enhance the network quality and coverage, we proposed a consolidation approach based on a relation inference engine taking into account relation **annotations**. The annotation system we presented in this paper is complementary to the lexical network consolidation system presented in [10]. This enhanced lexical network consolidation approach can provide (with the help of the annotation system) some important information which can be used for analyzing reports not only in the radiology domain as shown previously, but also in other specific domains and, certainly, for common sense reasoning applications.

It seems interesting and valuable to us to develop a knowledge representation for a specialized domain like radiology or parasitology to be included into a general lexical network. Indeed, specialized knowledge comes along common sense (but, obviously, not always the other way around).

To understand automatically medical reports of a given specialty, common sense inference is at least as important as specific knowledge of the domain.

Further research must improve the spreading relation annotation schema and the specialized lexical inference (in radiology as well as any other domain of interest) with the help of both experts and non-expert contributors.

The annotation schema as presented here is a first step toward representation of beliefs, uncertain knowledge, and points of view. Being able to make inferences both about facts and annotations in such a context is a major step toward automatic intelligent semantic analysis.

References

1. **Bodenreider, O. (2008).** Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform*, *47*, 67–79.
2. **Lomax, J. & McCray, A.T. (2004).** Mapping the gene ontology into the unified medical language system. *Comparative and functional genomics*, *5*(4), 354–361.
3. **Yetisgen-Yildiz, M., Gunn, M.L., Xia, F., & Payne, T.H. (2013).** A text processing pipeline to extract recommendations from radiology reports. *Journal of biomedical informatics,* (46), 354–362.
4. **Mejino Jr, J.L., Rubin, D.L., & Brinkley, J.F. (2008).** FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. *AMIA Annual Symposium Proceedings*, 465, American Medical Informatics Association.

5. **Rubin, D.L. (2008).** Creating and curating a terminology for radiology: ontology modeling and analysis. *Journal of digital imaging*, 21(4), 355–362.
6. **Hong, Y., Zhang, J., Heilbrun, M.E., & Kahn Jr, C.E. (2012).** Analysis of RadLex coverage and term co-occurrence in radiology reporting templates. *Journal of Digital Imaging*, *25*(1), 56–62.
7. **Gerstmair, A., Daumke, P., Simon, K., Langer, M., & Kotter, E. (2012).** Intelligent image retrieval based on radiology reports. *European radiology*, *22*(12), 2750–2758.
8. **Liu, H. & Singh, P. (2004).** ConceptNet—a practical commonsense reasoning tool-kit. *BT technology journal*, *22*(4), 211–226.
9. ***Lafourcade, M (2007).*** *Making people play for Lexical Acquisition. 7th Symposium on Natural Language Processing (SNLP 2007)*, Pattaya, Thailande, 8 p.
10. **Zarrouk, M., Lafourcade, M., & Joubert, A. (2013).** Inference and Reconciliation in a Crowdsourced Lexical-Semantic Network. *Computación y Sistemas*, Vol. 17, No. 2, 2013 147–159.
11. **Thaler, S., Siorpaes, K., Simperl, E., & Hofer, C. (2011).** A survey on games for knowledge acquisition. *Rapport technique, STI*, 26 p.
12. **Ramadier, L.M., Zarrouk, Z, Lafourcade, M., & Micheau, A. (2014*).*** Spreading Relation Annotations in a lexical semantic network: Applied to radiology. *International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2014. Lectures Notes in Computer Science*, Vol. 8403, 40–51.

Lionel Ramadier is a Ph.D. student at University Montpellier 2 (France) conducting research in Natural Language Processing applied to the medical domain as a member of the TEXTE team of LIRMM (Laboratoire d'Informatique, de Robotique, et de Microélectronique de Montpellier). He also works at IMAIOS, a website company specialized in medical e-learning and anatomy.

Manel Zarrouk is a Ph.D. student at University Montpellier 2 (France) conducting research in Natural Language Processing as a member of the TEXTE team of LIRMM (Laboratoire d'Informatique, de Robotique, et de Microélectronique de Montpellier). Her research interests are inferences of schemas and rules in large lexical semantic networks, consolidation, and evaluation of lexical resources for semantic analysis.

Mathieu Lafourcade is Associate Professor at University Montpellier 2 (France) and the Head of the TEXTE team of LIRMM. His research interests are semantic analysis with bio-inspired approaches (Ant Colony Algorithm), lexical semantics with large knowledge networks and vector models, lexical acquisition through contributive systems like games with a purpose. M. Lafourcade is the inventor of the JeuxDeMots project.

Antoine Micheau received his M.D. degree in 2007 from the University of Montpellier. He is a cofounder of IMAIOS, a website company specialized in medical e-learning and anatomy.

Article received on 10/01/2014, accepted on 01/02/2014.