
Enrichissement d'ontologie par une base générique minimale de règles associatives

Application aux maladies neurologiques : Les dystonies

Lamia Ben Ghezaiel — Chiraz Latiri * — Mohamed Ben Ahmed —
Neziha Gouider-Khouja **

Laboratoire de recherche RIADI-GDL, ENSI, Campus Universitaire La Manouba,
2010, Tunis

mohamed.benahmed@riadi.rnu.tn, lamiagh@gmail.com

* Faculté des Sciences de Tunis

Unité de recherche URPAH, Campus Universitaire El Manar, 1060 Tunis

chiraz.latiri@gnet.tn

** Institut National de Neurologie, La Rabta 1007 Tunis, Tunisie,

Unité de recherche MOUVEMENTS ANORMAUX ET PATHOLOGIES DES NOYAUX
GRIS CENTRAUX, ngkhouja@gmail.com

RÉSUMÉ. Dans cet article, nous proposons d'utiliser une base générique minimale de règles associatives entre termes (RA), afin d'enrichir automatiquement une ontologie de domaine existante. Initialement, des RA non redondantes entre termes sont extraites à partir d'un corpus du domaine. Ensuite, le rapprochement des termes candidats est effectué grâce à la mise en correspondance entre les concepts de l'ontologie initiale et les prémisses des RA, moyennant trois mesures de distance que nous définissons. Les premiers résultats obtenus sur une ontologie du domaine propre à la maladie des dystonies s'avèrent encourageants.

ABSTRACT. In this paper, we use a minimal generic base of association rules between terms, in order to enrich automatically an existing ontology. Such associations of terms will enable the domain expert to enhance the existing ontology in case those terms are not already defined in the ontology. Three distance measures are defined to move closer these candidate concepts of the initial ones in the ontology. Application on the ontology of dystonia diseases underlines the feasibility and the interest of the proposed approach.

MOTS-CLÉS : Ontologie, enrichissement automatique, règles associatives.

KEYWORDS: Ontology, automatic enrichment, association rules.

1. Introduction

Les ontologies sont un outil de représentation de connaissances et de raisonnement sur ces dernières, permettant d'organiser l'ensemble de concepts dans un domaine spécifique, ainsi que les relations entre ces concepts. Toutefois, les sources d'information, en général textuelles, évoluent en permanence aussi bien par le vocabulaire utilisé que par le sens des éléments de celui-ci. Les ontologies sont ainsi soumises régulièrement à des mises à jour et à des évolutions. Réaliser ces mises à jour manuellement demeure une tâche coûteuse et fastidieuse puisqu'elle mobilise un ou plusieurs experts du domaine pour identifier et classer de nouveaux éléments de vocabulaire dans l'ontologie. Afin d'accélérer ce processus d'évolution et d'adaptation et de lui enlever toute forme de subjectivité, des travaux de recherche assez récents se sont intéressés à la mise en oeuvre de techniques semi-automatiques et automatiques d'enrichissement d'ontologie. La majorité des approches citées dans la littérature, souvent basées sur des outils statistiques ou linguistiques, se sont focalisées sur l'ajout de nouveaux concepts et/ou relations existants entre eux.

Dans cet article, nous proposons une méthodologie de construction d'un réseau conceptuel formé par la combinaison deux types de connaissances, à savoir, des connaissances présentes dans la structure ontologique initiale propre à un domaine et représentées par des liens sémantiques, et des connaissances provenant de la base générique minimale de règles associatives (RA) entre termes (BenGhezaiel *et al.*, 2006), représentant essentiellement des corrélations qui sont appréciées par des mesures statistiques. Afin de valider la méthode d'enrichissement d'ontologie proposée dans un domaine réel, notre réflexion a porté sur la collecte, l'organisation, la représentation, la formalisation et l'enrichissement des connaissances dans le domaine de la neurologie pédiatrique et plus particulièrement, la maladie des dystonies ¹.

La suite de cet article est organisée comme suit : dans la section 2, nous présentons un aperçu des techniques d'enrichissement d'ontologie. La section 3 décrit notre nouvelle approche d'enrichissement d'ontologie par une base générique minimale de règles associatives, suit la section 4 qui décrit les premiers résultats de l'application de l'approche à la maladie des dystonies. Enfin, nous terminons l'article par une conclusion ainsi que les travaux en cours.

2. Aperçu des techniques d'enrichissement d'ontologies

Le processus d'enrichissement d'ontologie peut être divisé en deux étapes à savoir : une phase *d'apprentissage ontologique* pour rechercher de nouveaux concepts et relations, et une phase de *placement* de ces concepts et relations au sein de l'ontologie. Dans le cadre de notre travail, nous intéressons aux méthodes de découverte de

1. Cette phase expérimentale du travail est effectuée en collaboration avec les médecins spécialistes du Service de Neurologie de l'Enfant et de l'Adolescent de l'Institut de Neurologie de Tunis.

nouveaux termes candidats à partir du texte. Plusieurs travaux dans la littérature ont été proposés afin de couvrir l'une et/ou l'autre de ces étapes.

2.1. Approches existantes pour la découverte de concepts candidats

Nous distinguons deux types de méthodes pour la découverte des concepts candidats, à savoir :

– *Les méthodes statistiques* : elles sélectionnent les termes en fonction de leur distribution dans le corpus (Agirre *et al.*, 2000, Faatz *et al.*, 2002, Parekh *et al.*, 2004), ainsi que sur d'autres mesures plus complexes tels que l'information mutuelle, *tf-idf*, etc, ou encore l'utilisation de lois de distributions statistiques des termes (Neshatian *et al.*, 2004). Ces différentes propositions permettent d'identifier de nouveaux éléments de l'ontologie, mais ne permettent pas de les placer dans l'ontologie, sans une intervention humaine fastidieuse (Jorio *et al.*, 2007).

– *Les méthodes syntaxiques* : elles visent à déterminer la fonction grammaticale d'un mot ou d'un groupe de mots au sein d'une phrase. Elles sont basées sur l'hypothèse suivante : les dépendances grammaticales reflètent des dépendances sémantiques (Bendaoud, 2006). Elle définissent dans une phrase, le verbe (*V*) comme étant la relation qui lie le sujet (*S*) au complément (*C*). Elles présentent ainsi l'inconvénient d'identifier uniquement les relations labélisées par les verbes. Certaines approches utilisent aussi des patrons syntaxiques (Maedche *et al.*, 2002). Les termes extraits illustrent les nouveaux concepts candidats pour l'enrichissement, mais également l'existence de relations entre eux. Toutefois, ces relations ne sont pas étiquetées sémantiquement. De plus aucune mesure évaluant sémantiquement les nouvelles relations ajoutées n'est calculée.

2.2. Approches existantes pour le placement des concepts dans l'ontologie

Après la découverte des termes candidats, il est indispensable de détecter les relations entre ces nouveaux termes ainsi que celles qui le lient à l'ontologie initiale. Dans (Faatz *et al.*, 2002), les auteurs proposent pour cela une approche statistique basée sur la co-occurrence fréquente de termes candidats avec des termes de l'ontologie initiale. L'inconvénient majeur de ces travaux réside dans le fait qu'ils ne permettent pas l'ajout d'une manière précise des nouveaux concepts ni de relations entre eux dans la structure ontologique.

D'autres approches dans la littérature proposent d'utiliser des techniques de fouille de données (Stumme, 2005, Jorio *et al.*, 2007, Hernandez *et al.*, 2007). Les travaux de (Han *et al.*, 2000, Neshatian *et al.*, 2004) se basent sur une méthode de classification afin de rapprocher les termes candidats contenus dans les textes des concepts présents dans l'ontologie. Le principe est similaire que celui explicité dans les approches de (Agirre *et al.*, 2000, Parekh *et al.*, 2004), qui regroupent par une méthode de clustering

des termes en fonction de leur nombre d'occurrences au sein du corpus. Toutefois, ces méthodes ne détectent pas les relations entre les termes candidats, *i.e.*, ces nouveaux termes ne peuvent donc être ajoutés que par une intervention humaine.

Par ailleurs, plusieurs travaux proposent d'utiliser les corrélations fréquentes qui existent entre les termes d'un corpus. Ces approches consistent à extraire des règles d'association (Srikant *et al.*, 1997) entre des termes candidats (Maedche *et al.*, 2000, Bendaoud, 2006, Stumme *et al.*, 2006). A l'issue du processus de fouille, un ensemble règles associatives entre termes est généré. Chaque règle exprime l'existence d'une relation entre deux concepts du domaine. Ce processus d'enrichissement reste semi-automatique car d'une part, le nombre de règles associatives dérivées est très important et d'autre part, une intervention humaine est nécessaire pour définir sémantiquement les relations découvertes et pour les nommer.

3. Enrichissement d'ontologie par une base générique minimale de RA

Notre contribution est fondée sur le déploiement d'une base générique minimale de règles associatives entre termes, notée par \mathcal{MGB} (BenGhezaiel *et al.*, 2006), ne contenant que des associations entre termes non redondantes, afin d'enrichir une ontologie existante de domaine et aboutir à un *réseau conceptuel*. Ce réseau intègre deux types de connaissances à savoir : les connaissances sémantiques provenant de l'ontologie initiale et les connaissances implicites de la base générique \mathcal{MGB} , illustrées par les règles associatives entre termes.

3.1. Cadre formel

Formellement, une ontologie de domaine, notée dans la suite de l'article par \mathcal{O} , est définie comme suit (Cimiano *et al.*, 2004) :

Définition 1 Une ontologie est un quadruplet $\mathcal{O} = \langle C_D, \leq_C, R, \leq_R \rangle$ où C_D est l'ensemble des concepts du domaine, \leq_C est la relation d'ordre partielle définie sur C_D , R est l'ensemble des relations définies sur $C_D \times C_D$ et \leq_R est la relation d'ordre partielle définie sur R .

Nous considérons qu'un **contexte formel d'extraction** est un triplet $\mathfrak{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ où \mathcal{D} représente un ensemble fini de documents du corpus $\mathcal{C}_{\mathcal{MGB}}$, \mathcal{T} est un ensemble fini de termes et \mathcal{R} une relation binaire (*i.e.*, $\mathcal{R} \subseteq \mathcal{D} \times \mathcal{T}$). Chaque couple $(d, t) \in \mathcal{R}$ signifie que le document $d \in \mathcal{D}$ contient le terme $t \in \mathcal{T}$.

Définition 2 Un *termset*² est un ensemble non vide de termes noté par $(t_1, t_2 \dots t_k)$.

2. Terminologie proposée par analogie à celle utilisée en fouille de données, à savoir *itemset*.

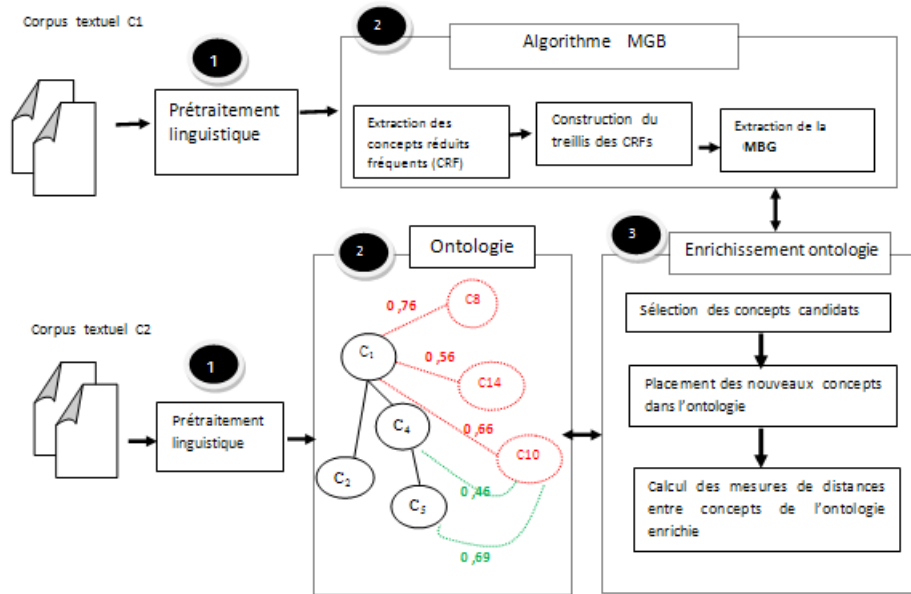


Figure 1. Processus général de l'enrichissement de l'ontologie par la base générique minimale de règles associatives

Définition 3 Une règle associative entre termes non redondante, incluse dans la base *MGB*, est une implication de la forme : $R : T_i \longrightarrow T_j$ tels que T_i et T_j sont deux termsets de tailles respectives n et m termes (BenGhezaiel et al., 2006).

Une règle associative R est appréciée par deux métriques statistiques, à savoir le *support* et la *confiance* (Agrawal et al., 1993). Le support de la règle associative $R : T_i \longrightarrow T_j$, noté par $Supp(R)$, exprime la fréquence avec laquelle les deux termsets T_i et T_j co-occurrent ensemble dans le corpus du domaine \mathcal{C}_{MGB} . La confiance de R , notée par $Conf(R)$, exprime la probabilité conditionnelle pour qu'un document contienne termset T_j , sachant qu'il contient le termset T_i . Une règle associative est dite *valide* si sa confiance est supérieure ou égale au seuil minimal de confiance noté par $minconf$.

3.2. Processus d'enrichissement proposé

Le processus général d'enrichissement d'ontologie que nous proposons est illustré par la Figure 1. Nous détaillons dans ce qui suit les trois étapes de notre processus d'enrichissement.

3.2.1. Prétraitement linguistique des deux corpus de domaine

Nous commençons en amont du processus de fouille par une phase de prétraitement linguistique des deux corpus textuels $\mathcal{C}_{\mathcal{O}}$ et $\mathcal{C}_{\mathcal{MGB}}$ du même domaine. Il consiste à détecter les unités linguistiques les plus représentatives du domaine. À partir du corpus $\mathcal{C}_{\mathcal{MGB}}$, nous construisons le contexte formel d'extraction $\mathfrak{R} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ pour dériver ensuite la base générique minimale de règles associatives \mathcal{MGB} .

3.2.2. Extraction de l'ontologie et de la base générique de règles associatives

Nous utilisons l'algorithme GEN-MGB (BenGhezaiel *et al.*, 2006) pour l'extraction de la base générique de RA non redondantes \mathcal{MGB} . Cette base est caractérisée par sa compacité importante, *i.e.*, elle contient un noyau minimal de RA, à partir duquel la totalité des règles redondantes et valides peuvent être déduites moyennant un système axiomatique complet et valide (Latiri *et al.*, 2005, BenGhezaiel *et al.*, 2006).

En considérant le contexte d'extraction textuel \mathfrak{R} , nous adaptons la définition de la base \mathcal{MGB} donnée dans (BenGhezaiel *et al.*, 2006) à la problématique d'enrichissement d'ontologie. Nous proposons de retenir que les RA non redondantes ayant un seul terme du domaine dans la prémisse. Soit :

$$\mathcal{MGB} = \{R : t \longrightarrow T_k \mid Conf(R) \geq minconf \wedge T_k = \{t_1, \dots, t_k\} \subseteq \mathcal{T}\} \quad [1]$$

Nous faisons appel ensuite à un outil semi-automatique tel que PROTÉGÉ2000 (Noy *et al.*, 2000) pour la construction de l'ontologie \mathcal{O} à partir du corpus $\mathcal{C}_{\mathcal{O}}$. Elle est validée en aval par un expert du domaine. L'évaluation du lien sémantique entre les concepts de l'ontologie \mathcal{O} est calculée à partir de la mesure de similarité proposée dans (Wu *et al.*, 1994) qui prend en compte à la fois la profondeur des concepts dans la hiérarchie de concepts et la structure de cette dernière.

Ainsi, la similarité entre deux concepts c_1 et c_2 de l'ontologie \mathcal{O} se calcule comme suit (Wu *et al.*, 1994) :

$$Sim_{Wu}(c_1, c_2) = \frac{2 \times depth(c)}{depth(c_1) + depth(c_2)} \quad [2]$$

Où $depth(c_i)$ correspond au niveau de profondeur du concept c_i et c représente le concept le plus spécifique qui généralise c_1 et c_2 dans \mathcal{O} .

3.2.3. Démarche adoptée pour l'enrichissement

La troisième étape consiste à rapprocher de l'ontologie initiale \mathcal{O} , les termes qui figurent dans les prémisses des règles candidates de la base \mathcal{MGB} . Ces termes sont

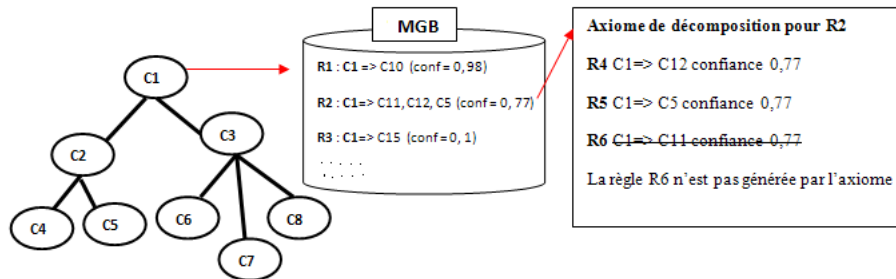


Figure 2. Exemple de calcul de concepts candidats

identifiés comme concepts candidats pour l'enrichissement. Une fois les nouveaux concepts placés dans l'ontologie, nous calculons les différentes mesures de distance qui évaluent les liens sémantiques entre les concepts de l'ontologie enrichie, notée dans la suite par \mathcal{O}_{MGB} .

Le processus d'enrichissement que nous proposons est itératif et comprend les étapes suivantes :

1) **Calcul des concepts candidats pour l'enrichissement** : Nous calculons pour chaque concept c_i de l'ontologie \mathcal{O} l'ensemble des concepts candidats à être reliés à c_i . Cet ensemble englobe les termes figurant dans les conclusions des règles associatives valides dont la prémisse est c_i ainsi que celles des règles redondantes. Selon l'exemple illustré par la Figure 2, les concepts candidats pour l'enrichissement relatifs au concept c_1 sont $\{c_{10}, c_{12}, c_5, c_{15}\}$.

2) **Placement des nouveaux concepts** : Cette étape consiste à placer les concepts candidats tout en préservant la cohérence des concepts et des relations préétablies dans l'ontologie initiale \mathcal{O} . Ceci permet de ne pas ajouter des redondances relationnelles dans le cas où un concept est candidat à être lié à plusieurs concepts de l'ontologie \mathcal{O} . Autrement dit, le concept candidat relié au concept de l'ontologie est celui qui figure dans la conclusion de la règle associative ayant la confiance la plus élevée.

L'exemple donné par la Figure 3 montre l'ajout des nouveaux concepts c_{10} et c_{11} et le déplacement du concept c_{15} car $Conf(c_1 \Rightarrow c_{15}) > Conf(c_7 \Rightarrow c_{15})$.

3) **Calcul du voisinage de c_i et des mesures de distance** :

Nous définissons la notion de *voisinage* d'un concept c_i de l'ontologie \mathcal{O} comme suit :

Définition 4 Le voisinage d'un concept c_i représente l'ensemble des concepts qui lui sont liés soit par la hiérarchie dans l'ontologie \mathcal{O} , soit par une ou plusieurs règles d'association valides.

Les relations entre c_i et son voisinage, *i.e.*, chaque concept candidat c_k appartenant au voisinage de c_i , sont évaluées à base d'une mesure statistique que nous appelons

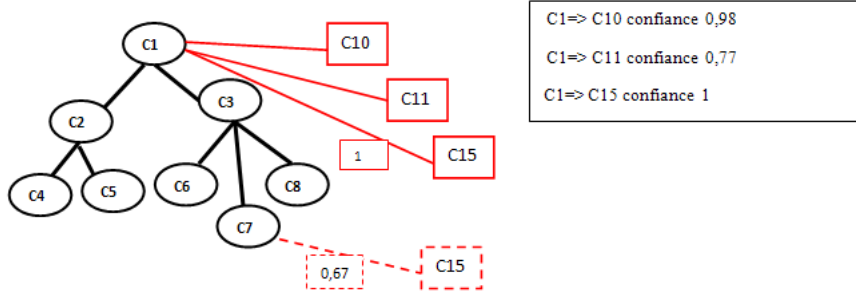


Figure 3. Exemple de placement de concepts candidats

mesure de distance entre c_i et son voisinage, et notée par $Dist_{\mathcal{O}_{MGB}}$. Elle est calculée en fonction de la mesure de confiance des règles associatives intervenant lors de l'enrichissement de l'ontologie \mathcal{O} et les mesures de similarités calculées entre les concepts de la structure ontologique initiale.

La mesure de distance que nous définissons est calculée selon trois cas possibles, que nous énonçons dans les propositions ci-dessous.

Proposition 1 Cas 1 : Si les deux concepts c_i et c_j proviennent de la base générique MGB alors $Dist_{\mathcal{O}_{MGB}}(c_i, c_j) = Conf(R : c_i \Rightarrow c_j)$.

Proposition 2 Cas 2 : Si les deux concepts c_i et c_j appartiennent initialement à l'ontologie \mathcal{O} alors $Dist_{\mathcal{O}_{MGB}}(c_i, c_j) = Sim_{Wu}(c_i, c_j)$.

Proposition 3 Cas 3 : Si c_i est un concept ajouté à l'ontologie \mathcal{O}_{MGB} et il est lié au concept c_k de l'ontologie initiale \mathcal{O} tel que $Dist_{\mathcal{O}_{MGB}}(c_k, c_i) = Conf(R : c_k \Rightarrow c_i) = \beta$ alors tout concept c_x de l'ontologie \mathcal{O} en relation avec c_k tel que $Sim_{Wu}(c_k, c_x) = \alpha$, est aussi en relation avec c_i . Dans ce cas, la mesure de distance est mixte, i.e., $Dist_{\mathcal{O}_{MGB}}(c_i, c_x) = \alpha \times \beta$.

Les trois cas sont illustrés dans la Figure 4.

4. Application aux maladies neurologiques : Les dystonies

Afin de valider notre approche d'enrichissement d'ontologie, nous proposons de l'appliquer sur des données réelles relatives à la maladie des dystonies. Les corpus des données nous ont été fournis par le service de Neurologie de l'Enfant et de l'Adolescent de l'Institut National de Neurologie de Tunis. Les dystonies correspondent à un groupe de maladies. Le symptôme ou la manifestation physique, correspond à des contractions prolongées, involontaires des muscles d'une ou de plusieurs parties du corps, entraînant souvent une torsion ou une distorsion de cette partie du corps.

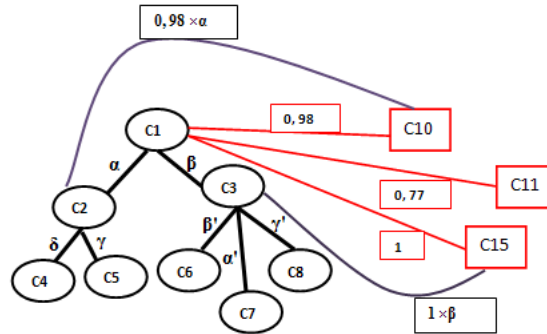


Figure 4. Exemple de calcul de la mesure de distance $Dist_{O_{MGB}}$

L'application proposée comprend les trois étapes mentionnées ci-dessus, comme suit :

1) **Prétraitement et génération du contexte formel d'extraction** : Notre évaluation expérimentale est conduite sur une population de 200 patients du service de Neurologie de l'Enfant et de l'Adolescent. Ils ont été classés en deux groupes à savoir : le groupe des patients atteints d'une *dystonie primaire* et ceux qui ont une *dystonie secondaire*. Chaque patient dispose d'un ensemble de données : des données personnelles et des données cliniques. Nous avons opté pour un choix méthodologique qui consiste en une étude progressive des dystonies. Nous nous intéressons dans cette première phase de l'approche uniquement aux maladies des dystonies primaires. Le contexte formel d'extraction correspondant au domaine des dystonies primaires est un triplet $\mathfrak{K}_{\mathcal{D}} = (\mathcal{D}, \mathcal{P}, \mathcal{R})$ où \mathcal{D} représente un ensemble fini de dossiers de patients, \mathcal{T} est un ensemble fini de propriétés décrivant les patients et \mathcal{R} une relation binaire (i.e., $\mathcal{R} \subseteq \mathcal{D} \times \mathcal{P}$). Chaque couple $(d, p) \in \mathcal{R}$ signifie que le dossier du patient $d \in \mathcal{D}$ contient la propriété $p \in \mathcal{P}$. Nous avons défini également un dictionnaire de correspondance qui contient un ensemble de 512 labels représentant l'ensemble des données cliniques, personnelles, symptômes, traitements, etc. avec leurs identifiants.

2) **Extraction de la base générique relative aux maladies des dystonies primaires**

Nous appelons l'algorithme GEN-MGB (BenGhezaïel *et al.*, 2006) qui prend en entrée le contexte formel d'extraction $\mathfrak{K}_{\mathcal{D}}$ ainsi que le dictionnaire de correspondance pour extraire la base générique de RA.

3) **Extraction de l'ontologie initiale des dystonies primaires**

L'ontologie des dystonies primaires initiale, notée par $\mathcal{O}_{\mathcal{D}\mathcal{P}}$, représente une hiérarchie qui relie des entités médicales en mettant l'accent sur le caractère étiologique de la maladie, i.e., dystonie primaire, dystonie plus, dystonie paroxystique, etc. Les

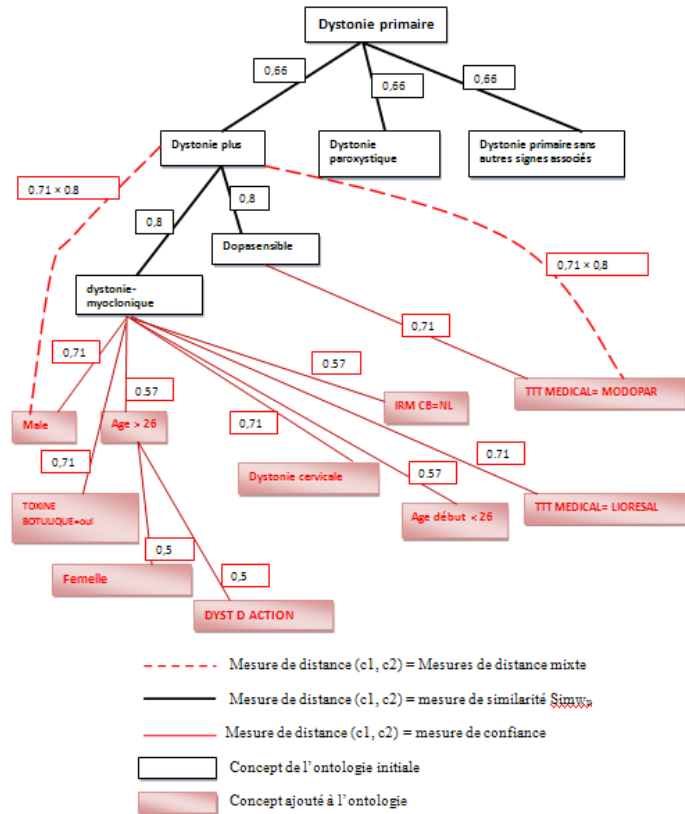


Figure 5. Ontologie des dystonies primaires enrichie par la base générique MGB

concepts de l'ontologie sont reliés entre eux par des liens *est-un*. Cette ontologie initiale est illustrée dans la Figure 5.

4) Enrichissement de l'ontologie et construction du réseau conceptuel des maladies des dystonies primaires

La Figure 5 représente un extrait de l'ontologie enrichie par les concepts de la base générique de règles associatives.

En interprétant le résultat du processus d'enrichissement avec les médecins du service Neurologie de l'Enfant et de l'Adolescent de l'institut National de Neurologie de Tunis, qui ont en amont validé l'ontologie initiale des dystonies primaires, nous avons déduit que l'enrichissement de l'ontologie leur a permis de dégager des connaissances additionnelles qu'ils ont approuvées, telles que :

- Relier les diagnostics étiologiques des maladies aux caractéristiques topographiques (dystonie membre supérieur, membre inférieur, focale,...), aux diagnostics génétiques (DYT11, DYT5,...), aux symptômes, aux traitements, etc.

- Déduire une prédominance masculine pour la maladie de dystonie myoclonique avec une mesure de distance entre le concept sexe masculin et dystonie myoclonique égale à 0.74.

- Déduire que l'âge de début de la maladie de dystonie myoclonie est généralement inférieur à 26 ans avec une mesure de distance entre (Age de début < 26) et (dystonie myoclonie) égale à 0.67. Nous pouvons conclure alors que c'est une maladie précoce.

- Déduire que le traitement médical souvent prescrit pour la dystonie myoclonique est LIORESAL, avec une mesure de distance égale à 0.71.

5. Conclusion

Différentes techniques d'enrichissement d'ontologies ont été proposées dans la littérature. Leurs limites proviennent du fait qu'elles ne permettent de traiter l'ensemble du processus d'enrichissement sans l'intervention de l'expert du domaine. Dans cet article, nous avons présenté un processus d'enrichissement d'ontologie automatique par une base générique de règles associatives. L'originalité de notre approche est qu'elle exploite le maximum de concepts pour l'enrichissement sans recourir à des connaissances a priori. Son avantage est de permettre l'apprentissage de la distance représentée par n'importe quelle relation de l'ontologie enrichie. Nos expérimentations sur un jeu de données réel lié aux maladies des dystonies primaires montrent la faisabilité et l'intérêt d'une telle approche. Nous projetons à court terme d'utiliser l'ontologie enrichie dans les systèmes de recherche d'informations médicales lors de la phase d'indexation.

Remerciements

Nous adressons nos vifs remerciements à tous les membres de l'équipe des médecins chercheurs du service de neurologie pédiatrique de Tunis. En particulier, nous remercions docteur Hanene Ben Rhouma et docteur Narjess Ben Fradj qui ont consentis beaucoup d'efforts pour valider l'approche sur des données réelles.

6. Bibliographie

- Agirre E., Ansa O., Hovy E., Martinez D., « Enriching very large ontologies using the WWW », *Proceedings of ECAI 2000 workshop on Ontology Learning*, 2000.
- Agrawal R., Imielinski T., Swami A., « Mining Association Rules between sets of items in large Databases », *Proceedings of ACM SIGMOD Conference, Washington, DC, USA*, p. 207-216, May, 1993.
- Bendaoud R., « Construction et enrichissement d'une ontologie à partir d'un corpus de textes », *Actes des Rencontres des Jeunes Chercheurs en Recherche d'Information (RJCRI'06)*, Lyon, p. 353-358, mars, 2006.

L. Ben Ghezaiel et al.

- BenGhezaiel L., Latiri C., BenAhmed M., « Fast-MGB : Nouvelle approche d'extraction de base générique minimale de règles associatives », *Actes des 6 èmes journée : Extraction et Gestion de connaissances, EGC'2006, Lille, France*, p. 217-222, Janvier, 2006.
- Cimiano P., Hotho A., Stumme G., Tane J., « Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies », *Proceedings of ICFCA'04*, p. 189-207, 2004.
- Faatz A., Steinmetz R., « Ontology enrichment with texts from the WWW », *Proceedings of the 2nd Semantic Web Mining Workshop at ECML/PKDD, WS'02*, 2002.
- Han E.-H., Karypis G., « Centroid based document classification : Analysis and experimental results », *Proceedings of The 4th European Conference of Principles of Data Mining and Knowledge Discovery*, p. 424-431, 2000.
- Hernandez N., Chrisment J. M. C., Egret D., « Modeling context through domain ontologies », *Journal of Information Retrieval, Contextual Information Retrieval Systems*, vol. 10, n° 2, p. 143-172, avril, 2007.
- Jorio L. D., Abrouk L., Fiot C., Hérin D., Teisseire M., « Enrichissement d'ontologie basé sur les motifs séquentiels », *Actes de la Plateforme AFIA 2007, Atelier Ontologies et gestion de l'hétérogénéité sémantique*, 2007.
- Latiri C., Bellagha W., BenYahia S., « VIE-MGB : A Visual Interactive Exploration of Minimal Generic Basis of Association Rules », *Proceedings of the third International Conference on Concept Lattices and their Applications (CLA'05), Olomouc, Czech Republic*, p. 179-196, September, 2005.
- Maedche A., Pekar V., Staab S., *Ontology Learning Part One - On Discovering Taxonomic Relations from the Web*, Springer Verlag, p. 301-322, 2002.
- Maedche A., Staab S., « Mining ontologies from text », *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. Springer-Verlag*, vol. 1937, 2000.
- Neshatian K., Hejazi M. R., « Text categorization and classification in terms of multi-attribute concepts for enriching existing ontologies », *Proceedings of the 2nd Workshop on Information Technology and its Disciplines*, p. 43-48, 2004.
- Noy N., Ferguson R. W., Musen M. A., « The Knowledge Model of Protégé-2000 : Combining Interoperability and Flexibility », *Proceedings of the 12th European Knowledge Acquisition Workshop (EKAW'00)*, 2000.
- Parekh V., Gwo J. P., Finin T., « Mining Domain Specific Texts and Glossaries to Evaluate and Enrich Domain Ontologies », *Proceedings of the International Conference of Information and Knowledge Engineering*, 2004.
- Srikant R., Agrawal R., « Mining generalized association rules », *Future Generation Computer Systems*, vol. 13, n° 23, p. 161-180, 1997.
- Stumme G., « Ontology Merging with Formal Concept Analysis », *Proceedings of Semantic Interoperability and Integration*, 2005.
- Stumme G., Hotho A., Berendt B., « Semantic web mining : State of the art and future directions », *Web Semantics : Science, Services and Agents on the World Wide Web*, vol. 4, n° 2, p. 124-143, June, 2006.
- Wu Z., Palmer M., « Verb semantics and lexical selection », *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics*, p. 133-138, 1994.