



HAL
open science

Décrire les maladies localisées et la physiopathologie pour une ontologie des urgences : un algorithme générique à partir de la FMA

Jean Charlet, Laurent Mazuel, Gunnar Declerck, Patrick Miroux, Pierre Gayet

► To cite this version:

Jean Charlet, Laurent Mazuel, Gunnar Declerck, Patrick Miroux, Pierre Gayet. Décrire les maladies localisées et la physiopathologie pour une ontologie des urgences : un algorithme générique à partir de la FMA. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.15-26. hal-01003849

HAL Id: hal-01003849

<https://hal.science/hal-01003849v1>

Submitted on 10 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Décrire les maladies localisées et la physiopathologie pour une ontologie des urgences : un algorithme générique à partir de la FMA

Jean Charlet^{1,2}, Laurent Mazuel^{2,3}, Gunnar Declerck^{2,4}, Patrick Miroux⁵, Pierre Gayet⁶

¹ Assistance Publique – Hôpitaux de Paris, F-75004, Paris, France ;

² INSERM, U1142, LIMICS, F-75006, Paris, France, Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France, Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France;

³ ANTIDOT, F-13410, Lambesc, France ;

⁴ Costech, Université de Technologie de Compiègne, F-60200, Compiègne, France ;

⁵ Dpt Urgences CHU d'Angers, F-49100, Angers, France ;

⁶ Centre hospitalier de Compiègne, F-60200, Compiègne, France.

Résumé :

ONTOLURGENCES est une représentation termino-ontologique du domaine de la médecine d'urgence initialement développée pour la recherche d'information. Cette ressource est conçue pour assurer *a)* le rôle de modèle du domaine répertoriant tous les concepts pertinents et *b)* le lien entre les concepts et la façon dont ils sont nommés dans les documents du Dossier Patient Informatisé. Cette double fonction permet l'annotation et l'indexation de dossiers patients et assure une recherche d'informations rapide. Le développement d'ONTOLURGENCES nous a amené à nous intéresser à l'articulation des points de vue physiologiques et anatomiques, deux composantes essentielles des modèles en médecine. Pour gérer les difficultés liées à une modélisation croisée de ces deux points de vue, très rapidement complexe et source d'erreur, il nous est apparu nécessaire d'exploiter une ressource de référence, le *Foundational Model of Anatomy*, la FMA, afin de réaliser une modélisation systématique de l'anatomie. Nous avons ainsi développé une méthode qui permet de générer de manière semi-automatique, à partir de la FMA : la branche des maladies localisées – *e.g. maladie du tube digestif* –, la branche de l'anatomie – *e.g. le tube digestif* lui-même – et, enfin, la définition des axiomes de localisation – *e.g. le fait qu'une maladie du tube digestif est localisée sur le tube digestif*. Les premiers résultats permettent de valider la pertinence de notre démarche.

Mots-clés : Ontologie médicale, épistémologie de la modélisation, anatomie médicale, physiopathologie médicale.

1 Introduction

Le projet LERUDI (LEcture Rapide en Urgence du Dossier patient Informatisé) vise à développer un système d'information (SI) offrant aux professionnels de santé une vision synthétique du dossier patient informatisé, et la possibilité d'un parcours rapide de celui-ci, pour permettre des prises de décisions médicales soumises à d'importantes contraintes de temps. Le champ d'expérimentation de ce projet est la lecture d'un dossier hospitalier par un médecin urgentiste. LERUDI est fondé sur une Ressource Termino-ontologique (RTO)¹ nommée ONTOLURGENCES, qui assure *a)* le rôle de modèle du domaine répertoriant tous les concepts pertinents

1. Une RTO se définit comme une ontologie dans laquelle les termes sont rattachés aux concepts de façon systématique et exhaustive. Il y a plusieurs façons de rattacher des termes à des concepts selon ce qu'on veut pouvoir représenter (Reymonet, 2007; Vandenbussche & Charlet, 2009).

et b) le lien entre les concepts et la façon dont ils sont nommés dans les documents du dossier patient. Cette double fonction doit permettre l'annotation et l'indexation de dossiers patients et la recherche d'information (RI) dans les dossiers indexés.

Les différentes étapes du développement de cette RTO ont été décrites dans (Charlet *et al.*, 2012b). Dans le présent article, nous allons nous intéresser à l'articulation, au sein de cette RTO, des points de vue physiopathologiques et anatomiques. Ces deux points de vue jouent un rôle essentiel dans la caractérisation des pathologies en médecine : par définition, une pathologie est un processus biologique anatomiquement localisé, par sa cause ou ses manifestations observables. Le travail de construction d'une ontologie médicale doit donc, d'une façon ou d'une autre, les intégrer et les modéliser explicitement. Or, cette entreprise pose d'importantes difficultés. Si, de prime abord, les grands concepts de l'anatomie (le corps humain et son découpage en éléments plus fins) et ceux de la physiopathologie² sont accessibles, leur modélisation précise devient très rapidement complexe et source d'erreurs. Et la nécessité d'assurer la cohérence (logique aussi bien que sémantique) des modélisations de chacun des deux points de vue rend la tâche plus complexe encore.

Pour remédier à ces difficultés, nous avons fait le choix d'utiliser une ressource de référence : le *Foundational Model of Anatomy* (la FMA dans la suite de cet article) pour réaliser le travail de modélisation systématique de l'anatomie dans ONTOLURGENCES. Nous avons ainsi développé une méthode qui permet de réutiliser l'énorme travail de modélisation de la FMA pour générer de manière semi-automatique la branche des maladies localisés – *e.g. maladie du tube digestif* –, la branche de l'anatomie – *e.g. le tube digestif* lui-même – et, enfin, la définition des axiomes de localisation – *e.g. le fait qu'une maladie du tube digestif est localisée sur le tube digestif*.

Dans la section 2, nous présentons la problématique de modélisation à laquelle nous sommes confrontés ; dans la section 3, nous décrivons les deux ressources impliquées dans notre projet, l'ontologie ONTOLURGENCES et la FMA ; dans la section 4 nous présentons la méthodologie proposée et l'algorithme développé ; dans la section 5, nous donnons les premiers résultats de ce travail ; enfin, dans la section 6, nous concluons et proposons quelques perspectives.

2 Modélisation de la physiopathologie vs de l'anatomie

La méthode anatomoclinique a été fondée par l'école française (Leannec, Breteanu, Trousseau) dans la première moitié du XIXe siècle. Elle consiste en la mise en relation de la clinique et de l'anatomie. Cette base de compréhension des maladies servira à leur traitement. Pendant la seconde moitié du XIXe siècle, la découverte de la nature microbiologique des infections (Pasteur) et l'invention de la médecine expérimentale (C. Bernard) impulsent une analyse des mécanismes des maladies : la physiopathologie.

Depuis, l'anatomie, la physiologie et la physiopathologie sont enseignées en médecine et forment le cœur des compétences des médecins. Si l'on cherche à modéliser les connaissances de base d'une spécialité – ici les Urgences – on est rapidement confronté à la nécessité de modéliser l'anatomie et la physiopathologie de la médecine dans les différents champs d'intervention. Par exemple, dans les urgences, on va s'intéresser aux fractures, du point de vue de l'anatomie – l'os – et de la physiopathologie – un traumatisme ou une métastase.

2. On organise habituellement les processus physiopathologiques selon 6 classes : les processus obstructifs, traumatiques, inflammatoire et infectieux, dégénératifs, tumoraux et enfin psychopathologiques.

C'est pourquoi les ontologies en médecine clinique modélisent l'anatomie et la physiopathologie. Comme indiqué en introduction, cette modélisation n'est pas aisée, en particulier en raison : 1) de la taille du domaine en termes de concepts et 2) de la nécessaire interaction de ces 2 points de vues.

Ainsi, un cancer doit être caractérisé par son type de cellules mais doit également pouvoir être localisé. Pour faciliter la conception d'ONTOLURGENCES et en améliorer l'interopérabilité, nous avons cherché à réutiliser différentes ressources du domaine biomédical (Charlet *et al.*, 2012b), notamment la branche des diagnostics de la SNOMED V3.5. La modélisation proposée par la SNOMED V3.5 s'est toutefois rapidement montrée insuffisante, car trop imprécise (granularité insuffisante), mais surtout mal organisée sur le plan de la physiopathologie – ce qui nous a passablement surpris. Des 25000 maladies répertoriées dans la SNOMED V3.5, seules 6500 ont ainsi été conservées (Charlet *et al.*, 2012b).

La cohérence générale d'ONTOLURGENCES à l'issue de ce travail de modélisation nous a toutefois laissé insatisfaits. La physiopathologie était plus ou moins bien modélisée, mais surtout l'anatomie était extrêmement lacunaire, en raison du nombre de concepts impliqués : par exemple, si l'on s'intéresse aux fractures des os, une modélisation satisfaisante sur un plan sémantique et opératoire (utilisable pour la RI avec de bonnes performances) exige : (a) de modéliser l'ensemble des os et leur topographie interne (tête, etc.), et (b) de définir chaque concept de fracture par une relation sémantique référant à l'os associé (ou aux os s'il s'agit d'une fracture multiple ou d'une structure osseuse complexe). Ce travail, extrêmement lourd, doit être répété pour toutes les structures anatomiques impliquées dans les pathologies auxquelles sont susceptible d'avoir affaire les urgentistes. Dans ce contexte, la FMA était bien, par son degré de détail et d'exhaustivité, la référence attendue. Mais inversement, précisément en raison de sa taille, le nombre d'éléments à réutiliser et les difficultés pour s'y orienter rendaient sa consultation et sa réutilisation durant la modélisation problématique, et partiellement aléatoire.

A ces difficultés, s'ajoutent des problèmes de choix de modélisation inhérents à toute conception d'ontologie. Notamment, il y a aujourd'hui un certain consensus sur l'idée qu'une modélisation ontologique bien faite (i) privilégie un axe de différenciation qui sert à organiser les relations de subsomption et (ii) modélise les autres axes via les relations (Bouaud *et al.*, 1995; Bachimont *et al.*, 2002). Ce qui signifie qu'on n'autorise qu'une seule relation de subsomption assertée par concept (pas de polyparentalité). Le premier axe sera constitutif de la hiérarchie de l'ontologie, définissant l'arbre *is-a*. Le second interviendra lors de la classification des concepts définis (Horridge, 2009). Dans ce contexte, nous avons donc l'alternative suivante :

- (1) utiliser le point de vue physiopathologique pour définir l'arbre *is-a* de notre ontologie et reconstruire le point de vue anatomique par classification, ou :
- (2) utiliser le point de vue anatomique pour définir l'arbre *is-a* et reconstruire le point de vue physiopathologique par classification.

Différentes raisons nous ont amené à adopter la première approche. D'abord par rapport à la physiopathologie : 1) la physiopathologie n'est généralement pas très bien maîtrisée du point de vue de la modélisation ontologique ; 2) corollaire, il n'y a pas de ressource de référence à ce sujet (notre expérience avec la SNOMED V3.5 nous l'a bien montré)³ ; 3) la relation phy-

3. Notre raisonnement à ce sujet est contre-intuitif : puisque la modélisation physiopathologique est difficile à faire et qu'il n'y a donc pas d'ontologie satisfaisante en ce domaine, prenons les moyens de construire cette partie du modèle et, à l'inverse, pour l'anatomie, réutilisons une ressource (la FMA).

siopathologique constitue souvent l'articulation entre les différentes pathologies des patients complexes. Or les pathologies chroniques constituent un enjeu de santé publique actuel, y compris aux urgences⁴, 4) enfin, en termes de concepts à modéliser, la physiopathologie représente une dimension bien plus petite que l'anatomie. 5) l'anatomie est assez bien maîtrisée du point de vue de la modélisation ontologique et cela explique, entre autres, l'existence de la FMA ; 6) parce que la FMA est disponible, on peut imaginer l'utiliser de façon plus ou moins automatique pour constituer la branche anatomique d'ONTOLURGENCES, 7) enfin, pour revenir à la question de l'essence unique de la modélisation, pour beaucoup de pathologies, il existe un positionnement physiopathologique unique alors que celles-ci ont un positionnement anatomique multiple (bronchopneumopathie).

L'essentiel de notre travail de modélisation a donc consisté à analyser aussi précisément que possible le point de vue physiopathologique tel qu'il est représenté dans la médecine d'urgences et à exploiter la FMA pour reconstruire par classification automatique la branche de l'anatomie dans son entier. En termes plus précis, cela revient à :

1. organiser l'ensemble des concepts de maladies modélisés par l'ontologie dans un arbre de subsomption représentant la perspective physiopathologique telle que la comprend la médecine d'urgence – *e.g. AnevrismeCardiaque is-a Anevrisme* – ;
2. développer la branche des maladies localisés avec des concepts définis (pour que puissent être positionnés sous ces derniers les concepts définis en (1) par classification automatique) – *e.g. MaladieCardiaque is-a MaladieDUnOrgane* –, ce qui implique de :
3. définir les axiomes de localisation – *e.g. le fait qu'une maladie du cœur est localisée sur le cœur : MaladieDUnOrgane and (localized some Heart) and (localized only Heart)–*, et enfin,
4. développer la branche de l'anatomie nécessaire à l'expression des axiomes – *e.g. le cœur lui-même : Heart is-a OrganWithCavitatedOrganParts is-a Organ*.

3 Ressources utilisées

Les ressources utilisées sont donc l'ontologie des urgences – ONTOLURGENCES– dans une version en perpétuelle évolution à partir de la 3.0.3 et la version OWL de la FMA.

ONTOLURGENCES v3.03. ONTOLURGENCES sert de test à ce travail depuis sa version de juillet 2013. Elle avait à ce moment 10191 classes, 60 *Object Properties*, 1 *Data property* et 11591 axiomes logiques dont 11339 axiomes de sous-classe et 89 axiomes de classes d'équivalence. L'ontologie est construite sous l'ontologie noyau de OntoMénélas⁵ (Charlet *et al.*, 2012a).

FMA. La FMA (*Foundational Model of Anatomy*) est « une ontologie de référence sur l'anatomie humaine » (Rosse & Mejino, 2003; Rosse & Jr, 2008). Elle est destinée, d'après ses concepteurs, à modéliser l'anatomie humaine canonique, c'est-à-dire « l'anatomie idéale à laquelle chaque individu et ses parties doivent se conformer » (Rosse & Jr, 2008). Elle contient plus de 85 000 classes, 140 relations reliant les classes, et plus de 120 000

4. http://www.ameli.fr/fileadmin/user_upload/documents/cnamts_rapport_charges_produits_2014.pdf

5. <http://purl.oclc.org/NET/spim/ontologies/public/OntoMenelas/>

termes. La plupart des entités sont des structures anatomiques composées de plusieurs parties interconnectées de façon complexe, décrites, par exemple, en termes de topographie, constituants, innervation, vaisseaux sanguins, limites ou frontières. Par exemple, un cœur a deux régions (le côté gauche et le côté droit), des parties le constituant (par exemple, la paroi du cœur, les diverses valves, le réseau d'innervation cardiaque, le septum interventriculaire, le septum auriculo-ventriculaire, les cavités, etc.). Il est innervé par le plexus cardiaque profond et alimenté par les artères coronaires, etc. (Golbreich *et al.*, 2013). De l'avis de tous, et malgré quelques bogues et insuffisances dans le modèle, elle est reconnue comme l'ontologie de référence de l'anatomie humaine.

Puisque la FMA a d'abord été développée dans un langage de *frames*, sa disposition en tant que fichier OWL, et *a fortiori* OWL2, ressort de travaux spécifiques de l'équipe de développement de la FMA comme d'autres chercheurs (Golbreich *et al.*, 2013). La version que nous avons utilisée est la version 3.2.1 en OWL Full de la FMA disponible sur le site en téléchargement⁶ : elle a 84 454 classes, 237 382 instances, 132 *Object Properties*, 167 *Data property* et 1 719 576 axiomes dont 87803 axiomes de sous-classe. Le grand nombre d'axiomes et d'instances sont liés au caractère *full* de cette version de la FMA où chaque classe est à la fois instance et sous-classe de la classe mère.

4 Méthode de génération

4.1 Notations et vocabulaire

Dans la suite de cet article, nous utiliserons les notations suivantes :

- \mathcal{O} l'ontologie de travail initiale et \mathcal{O}_{fma} la modélisation OWL de la FMA.
- Nous noterons l'espace de nom de l'ontologie \mathcal{O} « onto : » et celui de la FMA « fma : ». Ainsi, « onto:Cancer » et « fma:Heart » sont respectivement des URIs valides dans l'ontologie \mathcal{O} et dans la FMA \mathcal{O}_{fma} .
- Nous considérerons qu'une ontologie est définie par l'ensemble des axiomes la constituant. nous utiliserons ainsi la notation ensembliste appliquée à des axiomes (énoncés en écriture fonctionnelle de OWL2⁷) pour parler de l'appartenance d'objet à l'ontologie. Par exemple : $Declaration(Class(fma:Heart)) \in \mathcal{O}_{fma}$.

Nous définissons aussi quelques fonctions utilitaires qui seront utilisés dans les versions plus formelles des algorithmes de l'article⁸ :

- « fma_concept » est une fonction qui permet de récupérer la Classe FMA associée à un identifiant numérique FMA. Pour cela, nous utilisons l'annotation « fma:FMAID » présente dans la FMA :

$$fma_concept(fma_id) = \{cpt | AnnotationAssertion(fma:FMAID\ cpt\ fma_id) \in \mathcal{O}_{fma}\}_1 \quad (1)$$

6. http://sig.biostr.washington.edu/projects/fma/release/v3.2.1/alt_formats.html

7. <http://www.w3.org/TR/owl2-syntax/>

8. Certaines de ces fonctions sont suffixées par un « 1 » car nous prenons le premier élément pour éviter que le type de retour soit un ensemble. En effet, même si la théorie autorise plusieurs annotations, en pratique la FMA ne définit qu'un seul label et un seul identifiant numérique pour chaque concept.

- « `fma_label` » est une fonction qui permet de récupérer le label d'un concept FMA étant donné sa Classe. Pour cela, nous utilisons simplement le `rdfs:label` :

$$fma_label(fma_concept) = \{str | AnnotationAssertion(rdfs:label\ fma_concept\ str) \in \mathcal{O}_{fma}\}_1 \quad (2)$$

- « `fragment` » renvoie le fragment de l'URI, c'est-à-dire l'URI sans espace de nom. Par exemple `fragment(fma:Heart) = "Heart"`

4.2 Hypothèses initiales sur l'ontologie

Puisque la motivation principale est l'enrichissement automatique d'une ontologie médicale ayant besoin de localisation anatomique au moyen de la modélisation de la FMA, l'étape initiale est l'annotation de chaque concept de l'ontologie par un identifiant FMA.

Dans notre ontologie \mathcal{O} , nous avons donc ajouté l'annotation `onto:pourFMA` qui pour chaque concept de \mathcal{O} permet d'associer l'identifiant numérique unique d'un concept de la FMA. Cette étape est faite manuellement par les experts médecins du domaine de l'ontologie. A la demande des médecins, nous avons autorisé la multi-annotation (*i.e.* permettre d'associer plusieurs localisations à un seul concept). Cette multi-localisation se traduira alors par une intersection. Nous ne traitons pas actuellement de situation d'union entre plusieurs localisations.

L'algorithme que nous présentons à la section suivante suppose ainsi que tous les concepts ayant besoin d'une localisation sont annotés par une ou plusieurs annotations `onto:pourFMA`.

4.3 Algorithme de construction

4.3.1 Présentation générale

L'algorithme que nous présentons possède 3 objectifs majeurs :

- recopier la partie de la FMA pertinente dans l'ontologie ;
- construire une arborescence des maladies localisées ;
- ajouter des restrictions `some/only` entre les concepts de l'ontologie initiale.

Ceci nous permet à terme d'effectuer des inférences sur l'ontologie pour classifier automatiquement les concepts originaux sous les nouvelles maladies (e.g. inférer que « Cancer du poumon » est une « maladie du poumon » grâce à la relation de localisation sur « poumon » de la FMA). Les trois sections suivantes présentent les trois parties de l'algorithme.

4.3.2 Recopie de la FMA

C'est la partie la plus simple de l'algorithme. Pour chaque concept possédant une annotation `onto:pourFMA`, nous recopions dans \mathcal{O} l'ensemble des axiomes d'annotations du concept de la FMA ainsi que tous ses concepts parents jusqu'au concept `fma:Physical_anatomical_entity`⁹. Une fois ce concept atteint, la hiérarchie doit être stockée dans l'ontologie d'origine puisque les concepts anatomiques doivent être accessibles pour construire les restrictions de localisation (voir § 4.3.3). Pour ONTOLURGENCES qui représente notre cas d'étude, nous l'avons

9. La hiérarchie de la FMA hors de cette partie est abstraite et ne présente pas d'intérêt de localisation.

stocké sous `onto:StructureAnatomique`, mais le concept exact est bien entendu dépendant de l'ontologie dans laquelle la FMA est fusionnée.

A noter que nous conservons l'URI original de la FMA afin de garder tous les avantages du Linked Data. La recopie en interne ne sert qu'à des fins de visualisation et d'inférence hors ligne sur des logiciels tels que Protégé. Nous ne faisons que des recopies et ne procédons à aucune modification de la sémantique initiale de la FMA.

4.3.3 Création de la hiérarchie des maladies

L'idée générale est de construire un concept « DiseaseOf » pour chaque localisation recopiée de la FMA. Par exemple, si l'on recopie `fma:Heart`, alors nous allons créer un concept `onto:DiseaseOfHeart`. Néanmoins, la hiérarchie anatomique de la FMA ne se transpose pas directement en une hiérarchie des maladies. Il faut utiliser la hiérarchie tout-partie pour être pertinent médicalement parlant. Si par exemple un patient a une douleur thoracique, il a une maladie d'un organe du thorax (cœur/poumon). En suivant la subsomption anatomique, on pourrait inférer que la maladie du poumon est une maladie d'un organe. Si, dans l'absolu, cette subsomption n'est pas fautive, elle n'a aucune pertinence médicale donc aucun intérêt. A l'inverse, par une relation tout-partie, on infèrera que la maladie du poumon est une maladie du système respiratoire inférieur ou une maladie du thorax et on sera médicalement pertinent.

Pour cela, nous allons utiliser les relations tout-partie de la FMA. Elles sont au nombre de 4 (Mejino *et al.*, 2003) :

- `fma:part / fma:part_of`
- `fma:regional_part / fma:regional_part_of`
- `fma:systemic_part / fma:systemic_part_of`
- `fma:constitutional_part / fma:constitutional_part_of`

Ces relations sont symétriques, de sorte que $ObjectPropertyAssertion(X_part\ a\ b) \iff ObjectPropertyAssertion(X_part_of\ b\ a)$. En pratique, la représentation OWL de la FMA liste explicitement tous les axiomes et leur contraire¹⁰.

L'algorithme que nous allons utiliser est alors simple :

- à partir d'un concept donné, si il existe une relation « part_of » partant de ce concept, alors nous créons comme père « DiseaseOf » le concept au bout de ce part_of. Par exemple, si nous avons l'axiome $ObjectPropertyAssertion(fma:part_of\ fma:Lung\ fma:Thorax)$, alors nous créons l'axiome $SubClassOf(onto:DiseaseOfLung\ onto:DiseaseOfThorax)$;
- afin de garantir la connexité finale du graphe, si le concept n'a pas de relation « part_of » nous suivons alors la relation de subsomption classique.

Cette nouvelle hiérarchie est branchée sous le concept `onto:Diagnostic`. L'algorithme est défini plus formellement plus bas (voir algorithme 1).

D'autre part, les concepts de la hiérarchie construite à l'étape précédente peuvent être *définis* tel que les restrictions sont des relations vers la FMA. Par exemple « `onto:DiseaseOfLung` » est sémantiquement défini par « une maladie localisée sur une et uniquement une instance de poumon ». Pour cela, nous utilisons la relation « `onto:localisedOn` » définie entre l'arborescence des diagnostics et l'arborescence des localisations. Soit un concept « `fma:X` » et soit « `fma:Y` » le père de « `fma:X` », alors le concept « `onto:DiseaseOfX` » est un concept défini

10. La parité n'est pas strictement exacte dans le fichier, mais nous n'avons pas trouvé de contre-exemple nous laissant supposer que ce soit pour une autre raison que l'absence de complétude de la FMA.

par l'axiome suivant :

```

EquivalentClass(onto:DiseaseOfX
                 ObjectIntersectionOf(onto:DiseaseOfY
                                     ObjectSomeValuesFrom(onto:localizedOn fma:X)
                                     ObjectAllValuesFrom(onto:localizedOn fma:X)))

```

Si l'on instancie l'exemple précédent avec le couple Poumon/Thorax, cela donne :

```

EquivalentClass(onto:DiseaseOfLung
                 ObjectIntersectionOf(onto:DiseaseOfThorax
                                     ObjectSomeValuesFrom(onto:localizedOn fma:Lung)
                                     ObjectAllValuesFrom(onto:localizedOn fma:Lung)))

```

Algorithme 1 Construction d'un concept DiseaseOf

```

fonction build_parent_part(cpt) :
  // Avec  $X \in \{regional, systemic, constitutional, \}$ 
  retourne  $\{r \mid ObjectPropertyAssertion(X\_part\ r\ cpt)\} \cup$ 
           $\{r \mid ObjectPropertyAssertion(X\_part\_of\ cpt\ r)\}$ 

fonction build_disease_of(concept_fma) :
  label_fma = fma_label(concept_fma)
  new_uri = "onto: DiseaseOf" + fragment(concept_fma)
  insert( $\mathcal{O}$ , Declaration(Class(new_uri)))
  insert( $\mathcal{O}$ , AnnotationAssertion(rdfs: label new_uri "Disease of " + label_fma)))
  parent_part = build_parent_part(concept_fma)
  si parent_part  $\neq \emptyset$ 
    pour chaque  $c \in parent\_part$ 
      si  $c = fma : Physical\_anatomical\_entity$ 
        insert( $\mathcal{O}$ , SubClassOf(Class(new_uri), onto: Diagnostic))
      sinon
        parent_new_concept = build_disease_of(c)
        insert( $\mathcal{O}$ , SubClassOf(Class(new_uri), parent_new_concept))
  retourner Class(new_uri)
pour chaque  $c \in \{p \mid SubClassOf(concept\_fma, c) \in \mathcal{O}\}$ 
  si  $c = fma : Physical\_anatomical\_entity$ 
    insert( $\mathcal{O}$ , SubClassOf(Class(new_uri), onto: Diagnostic))
  sinon
    parent_new_concept = build_disease_of(c)
    insert( $\mathcal{O}$ , SubClassOf(Class(new_uri), parent_new_concept))

```

4.3.4 Ajout de restrictions dans l'ontologie initiale

Afin que les inférences puissent localiser correctement les concepts de l'ontologie initiale, il est nécessaire d'ajouter des restrictions aux concepts de la hiérarchie construite en 4.3.3.

Ces restrictions seront équivalentes aux restrictions utilisées pour les concepts « DiseaseOf », mais les concepts ne seront pas définis (*cf. infra*). D’autre part, nous devons tenir compte du fait que les annotations `onto:pourFMA` peuvent être multiples pour un concept donné (nous considérons cette multiplicité comme une intersection).

Ainsi, pour chaque concept `onto:X` de l’ontologie initiale contenant une annotation `onto:pourFMA` vers les identifiants « id_1 », « id_2 », ..., « id_n », les axiomes suivants seront ajoutés :

SubClassOf(onto : XObjectSomeValuesFrom(onto:localizedOn fma_concept(id_1))

...

SubClassOf(onto : XObjectSomeValuesFrom(onto:localizedOn fma_concept(id_n))

Et si $n = 1$:

SubClassOf(onto : XObjectAllValuesFrom(onto:localizedOn fma_concept(id_1))

Sinon :

*SubClassOf(onto:X ObjectAllValuesFrom(onto:localizedOn
ObjectIntersectionOf(fma_concept(id_1)*

...

fma_concept(id_n)))

4.3.5 Conclusion

L’algorithme final consiste en la juxtaposition des 3 parties précédentes comme indiqué dans l’algorithme 2.

Algorithme 2 Enrichissement global de l’ontologie cible

Pour chaque $(c, id) \in \{(c, id) | AnnotationAssertion(onto:pourFMA c id) \in \mathcal{O}\}$

$c_{fma} = fma_concept(id)$

`copy(c_{fma} , \mathcal{O})`

`build_disease_of(c_{fma})`

`add_restrictions(c , c_{fma})`

5 Résultats et discussion

L’ontologie fabriquée a 12396 classes, 60 *Object Properties*, 1 *Data property* et 17072 axiomes dont 13332 axiomes de sous-classe et 3559 axiomes de classes d’équivalence. Ces chiffres sont intéressants à comparer à ceux de l’ontologie dans sa version précédente, en particulier au niveau des axiomes de classes d’équivalence, 89 vs 3559 qui traduisent la création de nouveaux concepts définis (voir figure 1).

Un certain nombre de remarques méthodologiques peuvent être faites :

Restriction sur les localisations. Quand on localise une maladie avec l’annotation `pourFMA`, on a un effet de bord liée à la représentation formelle des classes dans une ontologie. Ainsi, si on annote une maladie du poumon avec l’identifiant du poumon de la FMA puis une maladie plus précise – *e.g.* emphysème pulmonaire – avec le même identifiant,

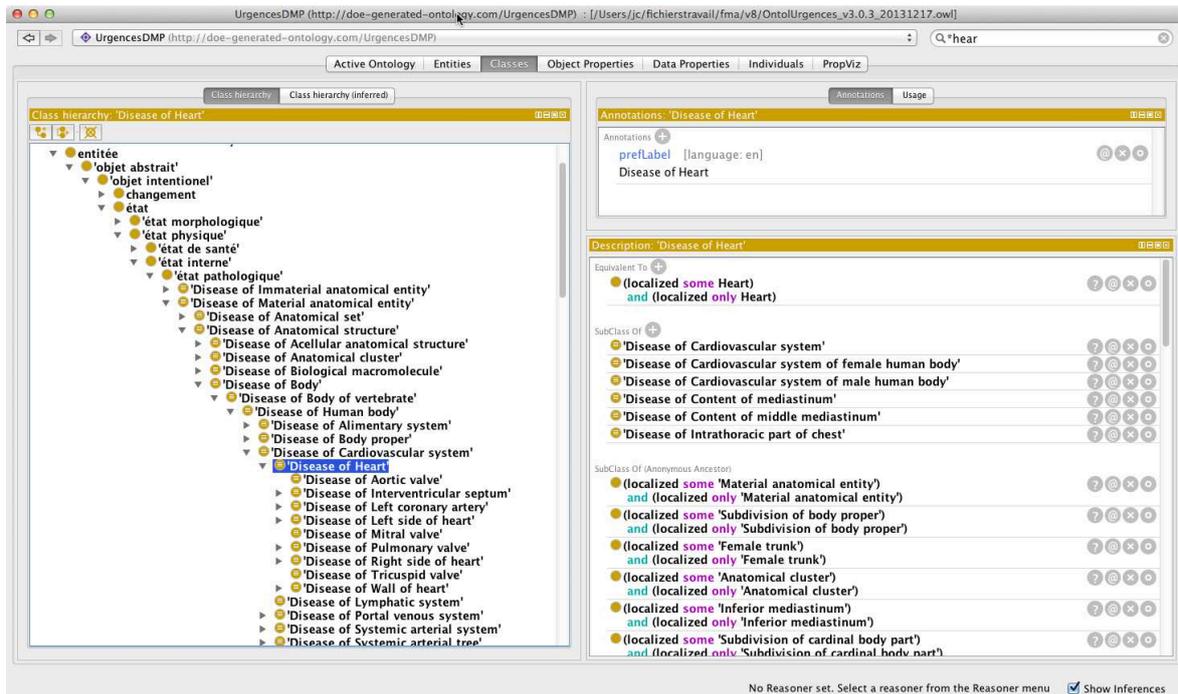


FIGURE 1 – Copie d’écran de l’ontologie vue dans Protégé au niveau du concept « Disease of Heart ». On voit au-dessus l’arborescence des maladies remontant aux concepts de l’ontologie noyau de OntoMénélas. Au niveau de la fenêtre des descriptions, on voit la description du concept défini et une grande partie des localisations héritées des concepts parents.

une application simpliste de l’algorithme qui génère les localisations, générerait deux concepts définis équivalents. Pour éviter cela mais pour que cette localisation soit directement attachée à un concept, il faut donc localiser le concept mais seulement en tant que restriction sur un concept primitif. Ainsi ce concept est localisé sur le poumon mais ce n’est pas définitoire, d’autres précisions (restrictions) sont possibles.

Sur la nature et la qualité de la FMA. On a précisé au 5, la version de la FMA que nous utilisons. En fait, ce n’est pas un problème car même s’il y a de nombreuses discussions sur la façon de générer la FMA dans un des différents dialectes (Golbreich *et al.*, 2013) d’OWL – OWL 1 Light, OWL 1 Full, OWL 2 DL, ... –, les liens de tout-partie ne sont pas impactés. Notre seul intérêt en l’espèce est que le dialecte OWL dans lequel la FMA est livrée soit stable pour éviter d’avoir à remettre en place plusieurs API de chargement. Par ailleurs, la FMA n’est pas exempte de bogues – plutôt rares – dont on ne discutera pas ici mais elle a surtout été construite avec des choix de modélisation discutables (au sens propre). Notre but n’est pas de revenir sur les choix de modélisation de la FMA, voire de les corriger car, à ce moment, notre processus de construction de l’ontologie ne pourrait pas être automatisée sur la partie anatomie. Comme, en pratique, la structure des part_of de la FMA est très complexe, nous cherchons plutôt à l’analyser et l’utiliser le plus finement possible.

Utiliser les relations « is-a ». Quand on utilise les part_of pour construire nos maladies localisées, les concepts de plus haut niveau générés ne remontent pas à un concept unique

en raison de la structure de la FMA : les relations « is-a » doivent former un treillis, les autres relations n'apportent pas cette contrainte donc notre algorithme non plus. Or pour des raison de clarté, il serait logique que cela soit le cas. La solution est d'utiliser les relations « is-a » quand on n'a plus rien d'autre. La pertinence médicale n'est alors pas évidente à cet endroit du treillis mais on est de toute façon à des niveaux de généralités pas très médicaux. Par analogie avec les ontologies de haut niveau, on pourrait considérer le haut des maladies localisées, comme un point de vue un peu « philosophique » sur celles-ci. Mais la justification de ce haut est encore à construire. Par ailleurs, on note que le treillis des « Diseaseof » est très dense en arcs par endroit. En analysant rapidement le treillis, on n'a rien vu qui n'ait pas de pertinence médicale mais on peut imaginer de nouvelles analyses pour diminuer le nombre d'arcs.

Un algorithme additif. L'un des avantages de notre algorithme est qu'il est incrémental : il est possible de l'appliquer sur une ontologie résultat elle-même de l'algorithme. Autrement dit, il n'est pas nécessaire de garder une version « hors-algorithme » pour pouvoir l'appliquer plusieurs fois, ni de nettoyer manuellement les nouvelles données pour s'en servir. Chaque concept est re-parcouru et les axiomes et restrictions sont créés si nécessaires. De la même manière, les arborescences « Disease » et FMA sont réduites afin de ne garder que les concepts utiles (lors de la suppression d'une annotation, il est en effet possible qu'un concept de la FMA deviennent inutilisé et par conséquent que le concept « Disease » associé devienne lui-même inutilisé). Il existe néanmoins une contrainte : si l'utilisateur change (ou supprime) manuellement le littéral d'une annotation « pourFMA » sur un concept donné, alors il doit aussi retirer manuellement les axiomes et restrictions que l'algorithme a créés. En effet, les axiomes ajoutés n'étant pas étiquetés, il est difficile (voir impossible) de faire la différence algorithmiquement entre un axiome ou une restriction ajoutée manuellement est un axiome ou une restriction calculée automatiquement. Pour finir sur ce sujet, on notera que le fait d'avoir des arborescences réduites à ce qui est nécessaire est en accord avec l'idée – défendue par les auteurs – que la FMA est une ontologie de référence dont on prend juste ce qui nous est nécessaire.

La question de la langue des termes. ONTOLURGENCES est une RTO qui sert entre autres à faire de la RI. La version française des termes doit donc être fournie par l'ontologie. Le travail fait ici est en anglais. Pour le faire en français, il faudrait que chacun des éléments anatomiques de la FMA ait un terme français associé. Ce n'est pour l'instant pas le cas. On envisage pour cela de reprendre des traductions faites par le CISMef sur le portail EHTOP¹¹ et de les proposer aux auteurs de la FMA pour l'enrichir en termes français.

6 Conclusion

Une première étape de l'évaluation de ce travail a été faite en interne et a d'abord permis de mettre au point l'algorithme exposé ici. La complexité du modèle de la FMA et la compréhension de ce que ça pouvait donner a amené de très nombreux cycles de mise au point. Le premier résultat est que nous avons réellement à notre disposition un algorithme qui nous affranchit d'une tâche longue, fastidieuse et source d'erreurs. Mais cet algorithme nécessite en-

11. <http://www.ehtop.eu/>

core des mises au point et un travail manuel à la fin. La question est aussi de savoir si l'on pourra s'affranchir de ce travail manuel ou le contenir à quelques actions précises terminales.

L'étape suivante de l'évaluation pourrait être de reprendre la nouvelle version de ONTOLURGENCES dans le projet LERUDI et faire une comparaison de l'indexation de quelques dossiers patients informatisés avec l'ancienne version *versus* la nouvelle. Une perspective à plus long terme est la mise en œuvre de cet algorithme sur une ontologie d'un autre domaine médical, qu'elle ait déjà été développée – et là, on suit le même processus que pour ONTOLURGENCES avec des tâches liées au « re-engineering » d'une RTO existante mais que, pour raisons de place, nous n'avons pas développé ici – ou qu'elle soit en cours de développement et la partie anatomique est alors développée par la méthodologie exposée ici. Pour terminer, le travail présenté ici n'est valable que pour la médecine et la FMA mais c'est un choix assumé, en raison du caractère unique de cette ontologie de référence et de l'importance du résultat poursuivi.

Références

- BACHIMONT B., ISAAC A. & TRONCY R. (2002). Semantic Commitment for Designing Ontologies : A Proposal. In A. GOMEZ-PÉREZ & V. BENJAMINS, Eds., *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, volume 2473 of *Lecture Notes in Artificial Intelligence*, p. 114–121, Sigüenza, Espagne : Springer Verlag.
- BOUAUD J., BACHIMONT B., CHARLET J. & ZWEIGENBAUM P. (1995). Methodological principles for structuring an “ontology”. In *Proceedings of the IJCAI'95 Workshop on “Basic Ontological Issues in Knowledge Sharing”*, Montréal, Canada.
- CHARLET J., BACHIMONT B., MAZUEL L., DHOMBRES F., JAULENT M. & BOUAUD J. (2012a). OntoMenelas : motivation et retour d'expérience sur l'élaboration d'une ontologie noyau de la médecine. *Technique et Science Informatiques*, **31**(1).
- CHARLET J., DECLERCK G., DHOMBRES F., GAYET P., MIROUX P. & VANDENBUSSCHE P.-Y. (2012b). Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et de modélisation. In S. SZULMAN, Ed., *Actes des 23^{es} Journées Ingénierie des Connaissances*, p. 33–48, Paris, France.
- GOLBREICH C., GROSJEAN J. & DARMONI S. J. (2013). The Foundational Model of Anatomy in OWL 2 and its use. *Artificial intelligence in medicine*, **57**(2), 119–32.
- HORRIDGE M. (2009). *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools*. The University Of Manchester, 1.2 edition. 109 pages.
- MEJINO J. V., AGONCILLO A. V., RICKARD K. L. & ROSSE C. (2003). Representing complexity in part-whole relationships within the Foundational Model of Anatomy. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, p. 450–4.
- REYMONET A. (2007). Modélisation de ressources termino-ontologiques en OWL. In F. TRICHET, Ed., *Actes des 18^{es} Journées Ingénierie des Connaissances*, p. 169–180, Grenoble, France : Cépaduès. ISBN 978.2.85428.790.5.
- ROSSE C. & JR J. L. V. M. (2008). The Foundational Model of Anatomy Ontology. In A. BURGER, D. DAVIDSON & R. BALDOCK, Eds., *Anatomy Ontologies for Bioinformatics : Principles and Practice*, chapter 4, p. 59–118. New York : Springer.
- ROSSE C. & MEJINO J. L. V. (2003). A reference ontology for biomedical informatics : the Foundational Model of Anatomy. *Journal of biomedical informatics*, **36**(6), 478–500.
- VANDENBUSSCHE P.-Y. & CHARLET J. (2009). Méta-modèle général de description de ressources terminologiques et ontologiques. In F. GANDON, Ed., *Actes des 20^{es} Journées Ingénierie des Connaissances*, p. 193–204, Hammamet, Tunisie.