

Usage secondaire du dossier médical informatisé à des fins épidémiologiques et d'évaluation de la qualité des soins : le projet SYNODOS

A secondary use of the electronic health record for epidemiological studies and evaluation of health care quality: the SYNODOS project

**Marie-Hélène Metzger^{1,2}, Quentin Gicquel¹, Domoina Rabarijaona³,
Frédérique Segond³, André Bittar⁴, Luca Dini⁴, Lina F. Soualmia⁵, Stéfan J.
Darmoni⁵**

¹Université Lyon I – CNRS-UMR 5558 / LBBE, Villeurbanne, France

²Hôpital de la Croix-Rousse, UHE, Lyon, France

³Objet Direct, Grenoble, France

⁴Holmes Semantic Solutions, Grenoble, France

⁵CISMeF, LITIS EA 4108 - Université de Rouen, France

Résumé

Contexte : L'objectif du projet SYNODOS est de développer une solution générique permettant l'extraction sémantique de données du dossier patient informatisé et d'organiser cette information médicale de manière à ce qu'elle puisse être utilisée à des fins d'études épidémiologiques ou d'évaluation de la qualité des soins.

Méthodes : SYNODOS propose une architecture modulaire qui fait une distinction claire entre les règles linguistiques et les règles du système expert. Les différents modules sont en cours de développement : une interface entre le serveur multi-terminologique et l'analyseur sémantique au cours de la phase d'extraction, règles linguistiques pour extraire les expressions temporelles, une interface entre le moteur linguistique et la base de connaissances.

Résultats: Les résultats du projet seront un système opérationnel intégrant les différents modules technologiques décrits ci-dessus. Le projet évaluera également la qualité des informations extraites dans deux domaines d'application : les infections associées aux soins et le cancer.

Conclusion : Nous avons décrit un projet dont l'originalité réside dans le développement d'une solution unique intégrant différentes technologies nécessaires à la production d'indicateurs épidémiologiques dans le contexte de l'activité hospitalière.

Abstract

Background: The purpose of the SYNODOS project is to develop a generic solution for extracting semantics out of medical data and organize this medical information in such a way that it could be used to support epidemiological studies or evaluate healthcare quality.

Methods: SYNODOS proposes a general modular architecture that makes a clear distinction between linguistic rules and expert system rules. An interface between the semantic analyzer and multi-terminology server upstream during the extraction phase, linguistic rules to extract temporal expressions as well as an interface between linguistic engine and knowledge representation are being developed.

Results: Project outcomes will be an operational system integrating the various technological modules described above. The project will also evaluate the quality of extracted information in two domains: hospital acquired infections and cancer.

Conclusion: We have described a project whose originality lies in the integration into a single solution of the various technologies needed to produce epidemiological indicators in the context of hospital activity.

Mots-clés : *Épidémiologie; Systèmes d'aide à la décision clinique ; Systèmes informatisés de dossiers médicaux ; Évaluation de programme.*

Keywords: *Epidemiology; Decision support systems, clinical; Medical records systems, computerized; Program evaluation.*

1 Introduction

Le dossier patient informatisé (DPI) constitue une source potentielle très importante de données dans des domaines aussi variés que l'aide à la décision médicale, la médecine factuelle, la surveillance épidémiologique ou la fouille de données. Toutefois très peu de données sont structurées et codées dans les DPI pour permettre ce type d'utilisation. À l'heure actuelle, seules les données nécessaires à la tarification à l'activité sont systématiquement structurées et codées dans les dossiers médicaux hospitaliers. Même si d'importants efforts ont été réalisés ces dernières années pour permettre l'utilisation épidémiologique des données du Programme de Médicalisation des Systèmes d'Information (PMSI), notamment en cancérologie, différents travaux montrent les difficultés liées à l'utilisation de ces données médico-économiques à des fins épidémiologiques [1]. Des travaux plus récents tentent de combiner ces données à d'autres sources de données structurées telles que les codes anatomo-pathologiques ADICAP [2]. Toutefois, la plupart des données du DPI sont disponibles dans des documents non structurés en langage naturel, ce qui nécessite la mise au point d'outils de traitement de cette information. Le projet SYNODOS vise à développer une solution de traitement du langage naturel des dossiers médicaux à des fins d'utilisation épidémiologique et d'évaluation de la qualité des soins.

2 Etat de l'art

De nombreux travaux sont en cours pour développer des méthodes de phénotypage à partir d'une utilisation secondaire des données du DPI. Ces méthodes intègrent notamment le traitement du langage naturel (ex : projet eMERGE [3]). Il n'existe pas à l'heure actuelle de langage formel ni d'approche standardisé pour extraire ces phénotypes [3]. Un modèle a été développé aux USA (National Quality Forum's Quality Data Model) qui donne une structure décrivant les concepts

cliniques dans un format standardisé pour la production automatique d'indicateurs de qualité. Ce modèle a été évalué et est utilisé dans le projet eMERGE [4] ou encore par le Consortium SHARPN qui développe une plateforme d'usage secondaire des données du DPI, reposant sur l'utilisation de diverses ressources (terminologiques, modèles de représentation) [5]. Dans le cadre du projet SYNODOS, nous avons développé un modèle conceptuel de données à visée générique et dont l'évaluation est prévue à l'issue du développement de la solution [6]. Par ailleurs une fois le phénotype obtenu par ces techniques, la 2^{ème} étape consiste à exploiter les données extraites pour par exemple, mesurer des associations ou sélectionner des patients éligibles. Il n'existe pas à notre connaissance de solution de langue française intégrant l'ensemble de ces étapes dans une solution unique et conçue pour être utilisée en production hospitalière. Le challenge du projet SYNODOS est de réussir à rassembler toutes les technologies nécessaires à l'exploitation de ces données pour divers usages en milieu hospitalier. Il s'agit donc d'un projet de recherche très appliqué.

L'objectif de cet article est de décrire le processus général de développement de la solution informatique SYNODOS.

3 Matériel et méthodes

Le projet SYNODOS (URL : <http://www.synodos.fr>) réunit 2 laboratoires publics de recherche, l'une experte dans le domaine de la recherche en Informatique Médicale (CISMeF) et l'autre dans le domaine de l'épidémiologie (LBBE), et 2 industriels, l'un spécialisé dans le développement de logiciels et ressources linguistiques (Holmes Semantic Solutions) et l'autre dans l'intégration de solutions de Business Intelligence et technologies web (Objet Direct). Le projet financé par l'Agence Nationale de Recherche (programme TecSan 2012) a été décomposé en 6 tâches scientifiques : 1) définition de l'architecture générale de la solution 2) développement du traitement sémantique des données textuelles médicales 3) interfaçage de l'extracteur de concepts multi-terminologique avec l'analyseur sémantique 4) développement d'un système générateur de règles expertes 5) intégration des différents modules constitutifs de la solution et enfin 6) évaluation des performances de la solution.

Le projet a démarré en octobre 2012 et se déroulera jusqu'à septembre 2015.

3.1 Architecture générale de la solution

La solution SYNODOS est une application web (SYNODOS – Médiateur) qui utilise les services de 2 serveurs distants :

- le serveur terminologique du CISMeF, existant, qui sera enrichi de nouvelles terminologies liées aux besoins spécifiques du projet
- le serveur sémantique, spécifique à SYNODOS, reposant sur les technologies de Holmes Semantic Solutions.

L'accès aux serveurs distants se fera par un « protocole de transfert hypertexte sécurisé » (HTTPS) de telle sorte qu'aucune donnée ne soit transmise en clair. Toutes les fonctionnalités de SYNODOS seront accessibles depuis le médiateur qui fournira l'interface unique de l'application. Différents logiciels seront utilisés pour le développement des modules composant la solution SYNODOS. Le projet s'appuiera sur: un langage de développement (Java), un framework web, un serveur d'applications, un SGBD relationnel, éventuellement une base NoSQL, un moteur BRMS (Business Rules Management System - Drools [7]) et des outils du web sémantique. La liste exacte des composants logiciels n'est pas finalisée, faisant partie du travail de développement en cours.

3.2 L'analyseur sémantique

L'analyseur sémantique est une plateforme de traitement du langage naturel développé par Holmes Semantic Solutions qui combine différents types de modules, appliqués de façon incrémentale dans le processus de traitement : modules symboliques (à base de règles), modules statistiques, et modules à base d'apprentissage automatique.

3.3 L'extracteur de concepts multi-terminologique

Le portail multi-terminologique développé par l'équipe CISMef contient 55 terminologies ou ontologies médicales, correspondant à 500,000 concepts médicaux en langue française et 1,5 million en langue anglaise (URL : <http://www.hetop.eu>). A partir de ce portail, un extracteur de concepts multi-terminologique (ECMT V2) a été développé pour indexer les termes médicaux rencontrés dans les documents textuels. Ce traitement est disponible en service web (SOAP ou REST).

Différentes terminologies médicales ont été sélectionnées pour normaliser le langage médical naturel dans le cadre du projet : la CIM-10 (Classification Internationale des Maladies, 10^{ème} revision), le thesaurus MeSH® (Medical Subject Headings) pour l'indexation des concepts diagnostiques ou symptomatologiques, la classification ATC (Anatomical Therapeutic and Chemical Classification) pour le codage des médicaments, la SNOMED International etc. Différentes techniques de mises en correspondance ont été employées dans ce projet : (a) mise en correspondance conceptuelle basée sur le CUI du Metathesaurus® de l'UMLS (Unified Modelling Language System); (b) mise en correspondance basée sur le traitement du langage naturel; (c) mise en correspondance manuelle ou supervisée par les experts terminologiques de l'équipe CISMef.

3.4 Développement d'un système générateur de règles expertes

Pour la finalité de la solution SYNODOS qui consiste à exploiter les données extraites des documents textuels à des fins épidémiologiques ou d'aide à la décision, une simple normalisation des concepts médicaux n'est pas suffisante. En effet, pour une utilisation pertinente des concepts extraits, il est nécessaire d'y associer un étiquetage temporel (s'agit-il d'une donnée reliée dans le parcours du patient à un antécédent, un motif d'entrée, une évolution en cours d'hospitalisation ?). Il est également nécessaire de créer des relations permettant de lier les concepts entre eux : par exemple dans la phrase suivante : « on a assisté à une chute de la gastrinémie », il est indispensable de relier le concept « gastrinémie » à celui de « chute » pour exploiter correctement cette information. Il faudra donc tout d'abord créer une règle qui permet d'affecter un concept du type « gastrinémie » à l'élément « type d'examen biologique » et un concept de type « chute » à l'élément « résultat d'examen biologique » de la base de faits. Ces règles s'appuieront sur l'utilisation des données sémantiques (par exemple négation du type « absence » associé au concept médical) et terminologiques (par exemple le type sémantique UMLS du concept) rattachées à chacun de ces concepts. La relation entre ces 2 concepts sera ensuite établie par l'utilisation des relations construites dans la base de connaissances (relations hiérarchiques ou transversales). Par ailleurs de nouveaux faits seront inférés par combinaison de « faits certains ». Par exemple, la présence d'un écoulement purulent de cicatrice après intervention chirurgicale, permettra d'inférer le fait qu'un patient est porteur d'une infection nosocomiale alors que ce dernier concept n'est pas mentionné dans le DPI. Des relations entre les différents concepts médicaux extraits, entre étiquettes temporelles ou relations sémantiques établies lors du traitement sémantique vont ainsi devoir être construites via un système expert développé par Objet Direct.

Le système expert sera composé d'une base de faits comprenant :

- les faits « certains », provenant du traitement terminologique et sémantique mais également des métadonnées récupérées dans le système hospitalier (ex : âge du patient, spécialité du service, type de compte-rendu traité,...)
- Les faits « inférés », c'est-à-dire générés par le raisonnement effectué par le moteur d'inférence. Pour effectuer ce raisonnement, le moteur d'inférence s'appuiera sur une base de règles. La base de règles sera élaborée par le LBBE et consistera en règles probabilistes, produites par technique de fouille de données et par règles heuristiques.

3.5 Apprentissage et évaluation des performances de la solution SYNODOS

Deux corpus de données textuelles sont utilisés dans le cadre du développement de la solution informatique SYNODOS. Le 1^{er} corpus a été constitué dans le cadre d'un précédent projet de recherche ALADIN-DTH [8] visant à développer un outil de repérage sémantique des infections nosocomiales dans les documents textuels médicaux. Dans le cadre de ce projet, 1,600 documents médicaux textuels composés de comptes rendus d'hospitalisation, lettres de sortie (et comptes rendus opératoires en chirurgie) ont été annotés par des médecins enquêteurs. Cette étude a été réalisée dans 4 centres hospitaliers universitaires français (Lille, Lyon, Nice, Rouen) sur la période allant de 2009 à 2011. Les dossiers médicaux des services de spécialités suivantes avaient été inclus : chirurgie digestive, neurochirurgie, chirurgie orthopédique et réanimation adulte.

La méthode de sélection et d'annotation de ces comptes rendus a été décrite en détail dans une autre publication [9]. Ce corpus de données textuelles a été divisé par tirage aléatoire des dossiers en 2 jeux : un jeu d'apprentissage et un jeu test. L'annotation consistait à fournir une référence pour la détection des infections nosocomiales.

Un 2^{ème} corpus de données textuelles sera constitué dans le cadre de ce nouveau projet et concerne un domaine médical très différent : la prise en charge diagnostique du cancer du côlon. Le choix de ce domaine médical vise à évaluer le caractère générique de la solution. Le nouveau corpus sera composé de 300 documents textuels médicaux, sélectionnés dans le centre anti-cancer régional de Rhône-Alpes chez des patients pris en charge pour un nouveau cancer du côlon. Le corpus sera constitué des documents textuels traçant la prise en charge du cancer du côlon : le compte rendu hospitalier, le compte rendu d'anatomo-pathologie, comptes rendus d'imagerie, comptes rendus de consultation, comptes rendus opératoires, compte rendu de concertation pluridisciplinaire). Un tirage aléatoire de 100 dossiers permettra de développer les algorithmes de repérage du cancer du côlon ainsi que les éléments diagnostiques nécessaires au calcul de différents délais de prise en charge.

Une annotation complémentaire d'un sous-groupe de dossiers est en cours afin de fournir la référence pour l'apprentissage et l'évaluation du peuplement de la base de faits. L'annotation est semi-automatique et réalisée à l'aide d'une application développée par le LBBE sur les logiciels R et MS Access. Cette annotation permet de reconstituer l'histoire médicale du patient, d'identifier le motif d'entrée, la prise en charge initiale et l'évolution. Il s'agit d'une restructuration a posteriori de la trajectoire de soins du patient effectuée dans le cadre de la prise en charge de l'événement de santé d'intérêt pour l'étude (par exemple prise en charge pour pose d'une prothèse de hanche et suivi dans les 12 mois). Cette annotation permet donc d'étiqueter les événements de santé temporellement dans cette trajectoire (s'agit-il d'un antécédent, d'un motif d'entrée,...), de distinguer les répétitions textuelles se rapportant au même événement de santé de celles correspondant à de réelles répétitions du même événement de santé dans la trajectoire du patient. La base d'annotation de l'application constitue le modèle conceptuel de la base de faits [6].

Les résultats aboutissant au repérage de l'infection nosocomiale par la solution SYNODOS seront validés par les médecins investigateurs sur un échantillon de 400 comptes rendus d'infections nosocomiales (soit 100 rapports par spécialité) et sur 400 comptes rendus de patients sans infection nosocomiale (soit 100 rapports par spécialité). Les résultats obtenus en termes de sensibilité / spécificité seront considérés comme les performances définitives de l'outil. Pour l'estimation du délai de prise en charge du cancer du côlon, le délai estimé par méthode automatique sera comparé au délai estimé par méthode manuelle sur 200 documents textuels. L'indicateur d'évaluation retenu sera la différence moyenne de délai de prise en charge du cancer du côlon obtenu entre la solution SYNODOS et la méthode manuelle de référence.

4 Résultats

4.1 Prétraitement des données hospitalières pour l'exploitation SYNODOS

La solution SYNODOS sera installée dans la zone démilitarisée du réseau sécurisé de l'établissement de santé ou chez un hébergeur agréé de données de santé afin de bénéficier de toute la sécurité liée au traitement de données médicales (cf. Figure 1).

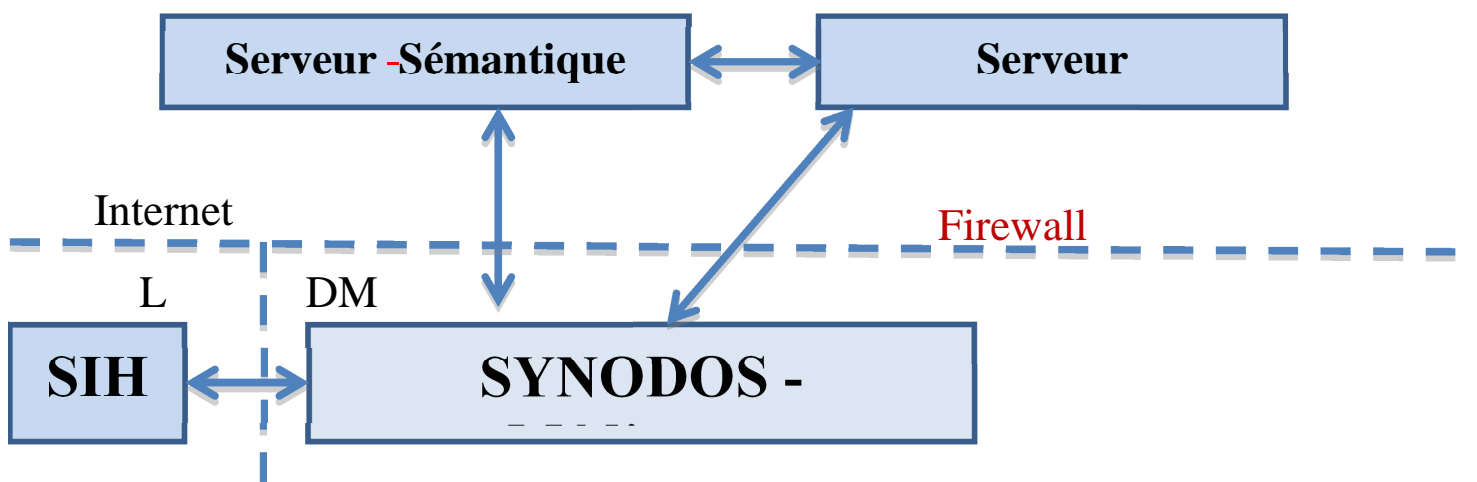


Figure 1 : Interface système d'information hospitalier, médiateur SYNODOS et serveurs Web distants

Objet Direct fournira une machine virtuelle prête à l'emploi, qu'il suffira de déployer sur le superviseur de l'établissement. Un module d'importation des données provenant du système d'information hospitalier récupèrera les métadonnées et les données textuelles « poussées » par le SIH de l'établissement. Le module procède ensuite à une anonymisation automatisée des données nominatives ou indirectement nominatives (ex : noms de personnes, noms de lieux, numéros de téléphone, adresses, adresse mail, etc.) afin de permettre les échanges avec les services web distants et afin également de permettre un usage épidémiologique des données. Ce module permettra également de désanonymiser le document après traitement pour certains types d'usage (par exemple surveillance intra-hospitalière des infections nosocomiales, où le dossier patient une fois repérée par le système doit être validé par l'équipe d'hygiène).

4.2 Indexation terminologique et analyse sémantique des documents textuels

Une fois les documents textuels formatés avec leurs métadonnées hospitalières dans un fichier XML, un 1^{er} traitement linguistique va être réalisé : détection et découpage des phrases par l'analyseur sémantique. Puis l'analyseur envoie une requête en mode SOAP au serveur terminologique afin que chaque concept de chaque phrase puisse être indexé. L'ECMT renvoie alors dans un fichier XML pour chaque concept médical reconnu, plusieurs métadonnées qui seront nécessaires à la suite du traitement (cf. Figure 2) : a) l'offset du 1^{er} caractère du concept médical b) le « concept unique identifier (CUI) » et le « type sémantique » si le concept médical est inclus dans le Metathesaurus de l'UMLS, c) sa spécialité médicale identifiée par des liens sémantiques établis manuellement par l'équipe CISMef ou par l'utilisation de terminologies contenant ce niveau de granularité. Le serveur renvoie la liste des termes indexés à l'analyseur sémantique selon un protocole HTTPS. L'analyse est effectuée phrase par phrase afin d'obtenir des délais de réponse compatibles avec la suite du traitement par la solution SYNODOS.

Après l'indexation, la suite du traitement linguistique consiste en la tokénisation, l'étiquetage morpho-syntaxique, la lemmatisation, l'analyse morphologique et l'analyse syntaxique en dépendances. L'analyseur propose une représentation graphique des dépendances syntaxiques, en permettant une visualisation et donc un contrôle de la qualité (cf. Figure 3).

```
- <return>
- <cis:sentence idpatient="0" idsentence="0" iddoc="0">
  <cis:text>le patient est arrivé aux urgences sous oxygène après avoir fait une crise d'asthme.</cis:text>
- <cis:indexations>
  - <cis:indexation end="84" idterm="9114" idcismef="ADI_PA_9114" offset="78" start="78" ter="ADI">
    - <cis:term>
      <cis:label lang="fr">ASTHME</cis:label>
    </cis:term>
    <cis:categorization />
  </cis:indexation>
  - <cis:indexation end="48" idterm="V03AN01" idcismef="ATC_CD_V03AN01" offset="41" start="41" ter="ATC">
    - <cis:term>
      <cis:label lang="fr">V03AN01 - oxygène</cis:label>
    </cis:term>
    <cis:categorization />
```

Figure 2 : extrait du fichier XML produit en sortie de l'extracteur multi-terminologique CISMef

L'analyse sémantique proprement dite se base ensuite sur le résultat de cette analyse syntaxique, ainsi que sur des connaissances ontologiques ou terminologiques. Cette analyse consiste à représenter les éléments de sens, permettant d'utiliser l'information extraite de façon adéquate : qui a fait quoi ? quand ? comment ? où ?

Pour représenter le résultat de l'analyse sémantique, l'analyseur permet de définir des relations (« prédicats logiques ou faits »). Par exemple pour la phrase « le patient est arrivé aux urgences sous oxygène après avoir fait une crise d'asthme », la Figure 3 représente l'analyse sémantique produite. Les nœuds correspondent aux entités (objets, endroits, personnes, etc.) et les arcs aux relations sémantiques définies entre ces entités.

L'analyseur sémantique envoie ensuite selon le protocole HTTPS les données issues du traitement sémantique sur un répertoire de stockage de la solution SYNODOS. Ces données sont stockées sous forme de fichiers XML avec un document textuel stocké par fichier.

4.3 Le système expert

Deux niveaux de règles expertes vont devoir être élaborés : 1) des règles expertes, permettant de traiter les données issues des 2 modules (terminologique et sémantique) et des métadonnées hospitalières afin de les ordonner dans la base de faits. Ce type de règles a été nommé « règles de transition ». Il s'agit par exemple de créer une relation entre un diagnostic médical tel que « pneumopathie à pneumocoques » et le statut « antécédent » de cette pathologie notifiée dans le compte-rendu; 2) des règles expertes permettant d'enrichir la base de faits avec de nouveaux faits, tels que par exemple une détection d'infection nosocomiale.

Les règles de transition font appel à des règles expertes de type « linguistique » ou à des règles expertes de type « médical ». Certaines règles nécessitent de combiner les 2 approches mais il est tenté de dissocier au maximum les 2 types de règles afin de répondre à l'un des défis du projet, à savoir permettre aux utilisateurs médicaux d'adapter eux-mêmes la solution au domaine d'application qui les concerne. La base de faits est en cours de développement et consiste à reproduire le modèle conceptuel élaboré par le LBBE lors de la procédure d'annotation manuelle, détaillée dans Gicquel et al. [6].

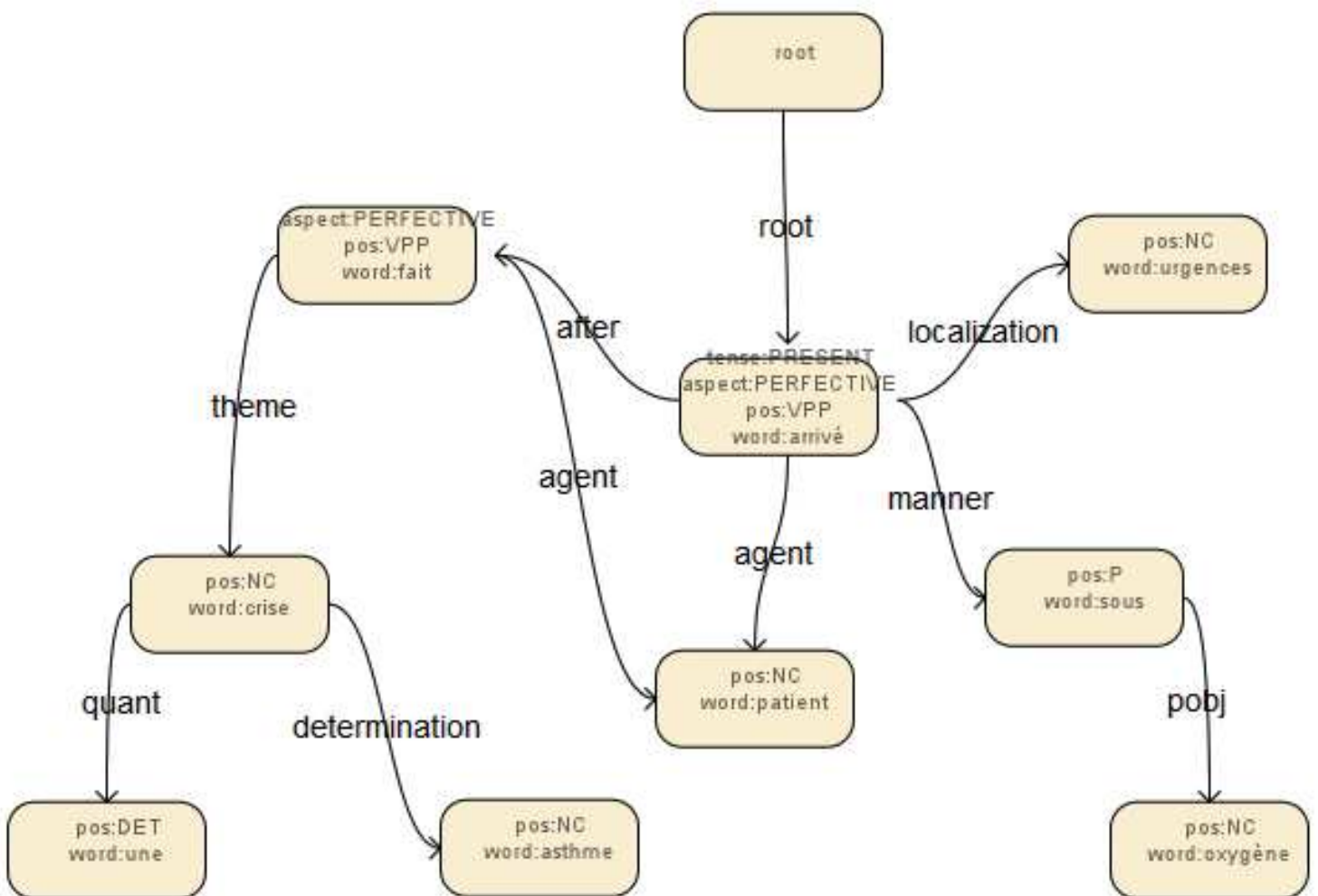


Figure 3 : Exemple de graphe sémantique

Deux approches ont été retenues pour le développement du module « système expert ». Une approche traditionnelle (classique) via l'utilisation d'un outil du marché (BRMS: Business Rules Management System) et une approche orientée Recherche qui étudie le Web sémantique et ce qu'il peut apporter dans un système expert.

5 Discussion

Le projet SYNODOS est donc un projet à visée très opérationnelle et dont l'un des objectifs majeurs est d'intégrer et d'interfacer des technologies encore très expérimentales (serveur terminologique, analyseur sémantique, systèmes expert) dans une solution utilisable en production hospitalière. Le développement n'est pas encore terminé à ce stade pour en évaluer les performances globales mais il est prévu de tester les fonctionnalités de la solution SYNODOS en mode production dès 2015 au Centre Léon Bérard (centre régional de cancérologie de Rhône-Alpes).

L'intérêt d'un tel outil dans le fonctionnement hospitalier est multiple : phénotypage des patients dans le cadre d'études épidémiologiques, screening de sujets éligibles pour diverses interventions, amélioration de la qualité, aide à la décision thérapeutique, évaluation de la prise en charge etc.

Ce projet présente toutefois un certain nombre de limites lié aux verrous technologiques existants : les ressources terminologiques mises à disposition pour l'indexation automatique des concepts médicaux bien que très riches, ne répondent pas de façon exhaustive et avec le niveau de précision souhaité à tous les domaines d'application. Par exemple, dans le cadre du domaine d'application sur la détection des infections nosocomiales, il a été mis en évidence la nécessité de développer une terminologie spécifique additionnelle, qui n'a pas pu être générée de manière automatisée avec les techniques classiques d'ingénierie des connaissances. Cette terminologie est donc enrichie manuellement en fonction de l'annotation manuelle du corpus, ce qui pose le problème de sa complétion. Selon les domaines d'application, les futurs utilisateurs de la solution pourront être confrontés à cette limite. L'étiquetage temporel des concepts médicaux dans le parcours de soins du patient est un autre élément indispensable à l'exploitation ultérieure de l'information médicale extraite. La pertinence de cet étiquetage temporel sera évalué dans le cadre du projet ; elle dépend à la fois de la qualité de l'analyse sémantique des documents mais aussi de la qualité des « règles de transition ». Plus généralement, les 2 challenges les plus importants du projet sont d'une part définir un modèle conceptuel suffisamment générique et d'autre part obtenir des « règles de transition » très diverses pour la représentation des données dans la base de faits afin que leur analyse ultérieure corresponde de la manière la plus exacte possible à la situation médicale décrite dans le DPI. Afin d'enrichir le modèle et les règles en fonction des utilisations ultérieures par domaine d'application, il est prévu une interface utilisateur permettant d'ajouter des éléments à la base de connaissances et de nouvelles règles expertes. L'évaluation de la solution SYNODOS prévue en 2015, permettra de préciser l'impact réel de ces limites.

D'autres limites sont inhérentes aux propriétés de ce type de données (incomplètes, complexes, biaisées). Hripcsak et al. ont notamment souligné les limites de l'usage secondaire du DPI à des fins de phénotypage [10]. Les perspectives seront d'étudier l'impact de ces propriétés sur la qualité du phénotypage.

6 Conclusion

La solution SYNODOS constitue un projet innovant dans la mesure où diverses technologies terminologiques, sémantiques et expertes sont assemblées pour traiter et analyser l'information médicale très riche des dossiers médicaux mais dont les données sont essentiellement textuelles. La conception de la solution que nous venons de décrire, vise à exploiter ces données pour tout domaine d'application médicale. Ce caractère générique devrait être assuré d'un point de vue fonctionnel par le fait que ce sont les utilisateurs de la solution qui choisiront et adapteront eux-mêmes les événements médicaux d'intérêt qu'ils souhaitent analyser dans leurs dossiers médicaux.

Remerciements

Ces travaux ont été financés par l'Agence Nationale de Recherche, dans le cadre d'un programme TECSAN (projet SYNODOS n° ANR-12-TECS-0006).

Références

- [1] Olive F, Gomez F, Schott AM, Remontet L, Bossard N, et al. (2011) [Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: a longitudinal approach becomes possible]. *Rev Epidemiol Sante Publique* 59: 53-58.
- [2] Defossez G, Rollet A, Dameron O, Ingrand P Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer. *BMC Med Inform Decis Mak* 14: 24.
- [3] Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, et al. (2013) The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 15: 761-771.
- [4] Thompson WK, Rasmussen LV, Pacheco JA, Peissig PL, Denny JC, et al. An evaluation of the NQF Quality Data Model for representing Electronic Health Record driven phenotyping algorithms. *AMIA Annu Symp Proc* 2012: 911-920.
- [5] Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DC, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. *J Am Med Inform Assoc* 20: e341-348.
- [6] Gicquel Q, Kergourlay I, Gerbier-Colomban S, Chariout S, Bittar A, et al. (2014) Annotation methods to develop and evaluate a medical expert system based on natural language processing in electronic medical records. MIE. Istanbul, Turkey.
- [7] JBoss Drools Business Logic Integration Platform. <http://www.jboss.org/drools>
- [8] Proux D, Hagège C, Gicquel Q, Kergourlay I, Pereira S, et al. (2012) ALADIN : Développement d'un outil sémantique d'analyse des documents textuels médicaux pour la détection d'infections associées aux soins IRBM, *Ingenierie et Recherche BioMedicale* 33: 137-142.
- [9] Metzger M, Gicquel Q, Kergourlay I, Cluze C, Grandbastien B, et al. (2011) Codage standardisé de données médicales textuelles à l'aide d'un serveur multi-terminologique de santé : exemple d'application en épidémiologie hospitalière In: Degoulet P, Fieschi M, editors. *Systèmes d'information pour l'amélioration de la qualité en santé : comptes rendus des quatorzièmes Journées francophones d'informatique médicale*, Tunis, 23 et 24 septembre 2011. Paris: Springer-Verlag France. pp. 109-120.
- [10] Hripcsak G, Albers DJ (2013) Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* 20: 117-121.

Adresse de correspondance

Marie-Hélène Metzger, UCBL-CNRS UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, Villeurbanne F-69622, France.

E-mail : marie-helene.metzger@chu-lyon.fr

URL : <http://www.synodos.fr>