



**HAL**  
open science

# Une approche pour la conception de systèmes d'aide à la décision médicale basés sur un raisonnement mixte à base de connaissance

Lamine Benmimoune

## ► To cite this version:

Lamine Benmimoune. Une approche pour la conception de systèmes d'aide à la décision médicale basés sur un raisonnement mixte à base de connaissance. Ingénierie assistée par ordinateur. Université de Technologie de Belfort-Montbéliard, 2016. Français. NNT : 2016BELF0307 . tel-02003476

**HAL Id: tel-02003476**

**<https://theses.hal.science/tel-02003476v1>**

Submitted on 1 Feb 2019

**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.

# SPIM

## Thèse de Doctorat



école doctorale sciences pour l'ingénieur et microtechniques

UNIVERSITÉ DE TECHNOLOGIE BELFORT-MONTBÉLIARD

# Une approche pour la construction de systèmes d'aide à la décision médicale basés sur un raisonnement mixte à base de connaissances

■ LAMINE BENMIMOUNE



# SPIM

## Thèse de Doctorat



école doctorale sciences pour l'ingénieur et microtechniques  
UNIVERSITÉ DE TECHNOLOGIE BELFORT-MONTBÉLIARD

N° X | X | X

THÈSE présentée par

LAMINE BENMIMOUNE

pour obtenir le

Grade de Docteur de

l'Université de Technologie de Belfort-Montbéliard

Spécialité : **Informatique**

Une approche pour la construction de systèmes d'aide à la  
décision médicale basés sur un raisonnement mixte à base  
de connaissances

Unité de Recherche :

Institut de Recherche sur les Transports, l'Énergie et la Société (IRTES)

Soutenue publiquement le 10 décembre 2016 devant le Jury composé de :

M. LADJEL BELLATRECHE	Rapporteur	Professeur des Universités, ENSMA, LIAS/ISAE-ENSMA
MME. CORINE CAUVET	Rapporteur	Professeur des Universités, Université Aix-Marseille
MME. CÉCILIA ZANNI-MERK	Examinateur	Professeur des Universités, INSA Rouen
M. AMIR HAJJAM EL HASSANI	Directeur de thèse	Maître de conférences HDR, Université de Technologie Belfort-Montbéliard
MME. PARISA GHODOUS	Co-directrice	Professeur des Universités, Université Lyon 1



# REMERCIEMENTS

Par ces quelques lignes, je souhaite remercier toutes les personnes qui ont contribué de près ou de loin à l'aboutissement de cette thèse.

Je tiens à exprimer ma profonde gratitude à monsieur Amir HAJJAM, Maître de conférence, HDR à Université de technologie de Belfort-Montbéliard, pour avoir dirigé et encadré mes recherches. Je le remercie pour la confiance qu'il m'a accordé en me proposant le sujet de ma thèse, et pour tous les moyens qu'il a mis à ma disposition pour mener à bien mes travaux. Ses précieuses remarques, son exigence et ses commentaires ont permis d'améliorer la qualité de mes travaux. Je lui remercie vivement pour son aide précieuse dans la relecture et la correction de ce manuscrit.

Je tiens à remercier également, ma codirectrice de thèse, madame Parisa GHODOUS, Professeur à l'Université Lyon 1, pour avoir co-encadré mes recherches. Je la remercie pour son aide précieuse dans la relecture et la correction de ma thèse, ses précieux conseils de tous ordres, et sa confiance. Son expérience et ses compétences ont permis l'accomplissement de ce travail.

Mes sentiments chaleureux de gratitude et de remerciements à monsieur Mohamed HAJJAM, mon directeur technique à l'entreprise Newel où j'ai passé 3 agréables années, qui m'a fourni une assistance et des conseils précieux tout au long de cette expérience avec beaucoup de patience, de savoirs et d'écoute.

Je remercie vivement monsieur Ladjel BELLATRECHE, Professeur à l'ENSMA, madame Corine CAUVET, professeur à l'université Aix-Marseille, pour avoir accepté d'être rapporteurs de ce mémoire, et pour l'honneur qu'ils me font en participant au jury. Merci également à madame Cécilia ZANNI-MERK, professeur à l'INSA d'avoir accepté d'examiner ce travail et de faire partie du jury.

Je tiens à remercier les partenaires médicaux des projets E-CARE, REPOSE, BRAVE et EHPAD et spécialement médecins M. Emmanuel ANDRÈS, M. Samy TALHA, M. Benoît ROMAIN et M. Martin HUBNER pour leur apport en matière d'expertise médicale.

J'aimerais remercier l'équipe Newel qui m'a accompagnée et m'a soutenu, et les moments inoubliables qui ont marqué nos merveilleuses trois ans passées ensemble. Je remercie également toutes les équipes administratives de l'UTBM qui ont assuré le bon déroulement de cette thèse.

Je remercie du fond du cœur mes parents, mes frères et mes amis qui n'ont jamais cessé de croire en moi pendant toutes mes années d'études et qui m'ont toujours encouragé à

aller de l'avant.

Enfin et surtout, je remercie avec grand amour mon épouse qui m'a soutenue dans les moments difficiles et incertains. Elle était la lumière étincelle qui se fortifiait jour après jour et qui animait mon chemin avec plein d'émotions, d'enthousiasme et de bonheur.

## Résumé

Afin d'accompagner les professionnels de santé dans leur démarche clinique, plusieurs systèmes de suivi et de prise en charge médicale ont été construits et déployés dans le milieu hospitalier. Ces systèmes permettent principalement de collecter des données médicales sur les patients, de les analyser et de présenter les résultats de différentes manières. Ils représentent un appui et une aide aux professionnels de santé dans leur prise de décision par rapport à l'évolution de l'état de santé des patients suivis. L'utilisation de tels systèmes nécessite systématiquement une adaptation à la fois au domaine médical concerné et au mode d'intervention. Il est nécessaire, dans un milieu hospitalier, que ces systèmes puissent s'adapter et évoluer d'une manière simple, en limitant toute maintenance corrective ou évolutive. Ils doivent être en mesure de prendre en compte dynamiquement des connaissances théoriques et empiriques du domaine issues des experts médicaux.

Afin de répondre à ces exigences, nous avons proposé une approche pour la construction d'un système d'aide à la décision médicale capable de s'adapter au domaine médical concerné et au mode d'intervention approprié pour assister les professionnels de santé dans leur démarche clinique. Cette approche permet notamment l'organisation de la collecte des données médicales, en tenant compte du contexte du patient, la représentation des connaissances du domaine à base d'ontologies ainsi que leur exploitation associée aux guides de bonnes pratiques et à l'expérience clinique.

Dans la continuité des travaux précédemment réalisés au sein de notre équipe de recherche, nous avons choisi d'enrichir, avec notre approche, la plateforme E-care qui est dédiée au suivi et à la détection précoce de toute anomalie de l'état de patients atteints de maladies chroniques. Nous avons pu ainsi adapter facilement la plateforme E-care aux différentes expérimentations qui sont été menées notamment dans des EPHAD de la Mutualité Française en Anjou-Mayenne, au CHU de Haute-pierre et au CHUV à Lausanne.

Les résultats de ces expérimentations ont montré l'efficacité de l'approche proposée. L'adaptation de la plateforme par rapport au domaine et au mode d'intervention de chacune de ces expérimentations se limite à de la simple configuration. De plus, l'approche proposée a suscité l'intérêt du personnel médical par rapport à l'organisation de la collecte des données, qui tient compte du contexte du patient, et par rapport à l'exploitation des connaissances médicales qui apporte aux professionnels de santé une assistance pour une meilleure prise de décision.

Mots clés : Système d'aide à la décision médicale ; Ontologie ; Raisonnement à base de cas ; Raisonnement à base de règle ; Suivi médical.



## Abstract

To support health professionals in their clinical processes, several monitoring and medical care systems have been built and deployed in the hospital setting. These systems are mainly used to collect medical data on patients, analyze and present the outcomes in different ways. They represent support and assistance to health professionals in their decision making regarding the evolution in the health status of the patients followed. The use of such systems always requires an adaptation to both the medical field and the mode of intervention. It is necessary, in a hospital setting, to adapt and evolve these systems in a simple manner, limiting any corrective or evolutionary maintenance. Moreover, these systems should be able to consider dynamically the domain knowledge from medical experts.

To meet these requirements, we proposed an approach for the construction of a medical decision support system (MDSS). This MDSS can adapt to the medical field and to the appropriate mode of intervention to assist health professionals in their clinical processes.

This approach allows especially the organization of the medical data collection by taking into account the patient's context, the ontology-based knowledge representation of the domain and permits the exploitation of the medical guidelines and the clinical experience.

In continuity of our research team's previous work, we chose to expand with our approach, the E-care platform which is dedicated to monitoring and early detection of any abnormality of the health status of patients with chronic diseases. We were able to adapt easily the E-care platform for the various experiments that have been conducted, including EPHAD of the Mutualité Française in Anjou-Mayenne, Hautepierre hospital and Lausanne hospital (CHUV).

The outcomes of these experiments have shown the effectiveness of the proposed approach. Where, the adaptation of the platform regarding to the domain and mode of intervention of each of these experiments is limited to the simple configuration. Furthermore, the proposed approach has attracted the interest of the medical staff regarding the organization of the medical data collection, and the exploitation of the medical knowledge which brings assistance to the health professionals for better decision making.

Key words : Medical decision support system, Ontology, Data collection, Case-based reasoning, Rule-based reasoning, medical monitoring.

## Publications scientifiques

### Chapitres de Livres

[1] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres, Samy Talha, and Mohamed Hajjam. Ontology-based information gathering system for patients with chronic diseases : Lifestyle questionnaire design. In *Progress in Artificial Intelligence* , volume 9273 of *Lecture Notes in Computer Science* , pages 110–115. Springer International Publishing, 2015.

[2] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres and Mohamed Hajjam. Ontology-Based Contextual Information Gathering Tool for collecting patients’ data before, during and after a digestive surgery. *Handbook on smart healthcare* (à paraître).

### Revues

[1] Hicham ELASRI, Abderrahim SEKKAKI, Amir HAJJAM, Lamine Benmimoune, Samy TALHA, Emmanuel ANDRES. Ontologies et intégration des connaissances pour un suivi polypathologique. In *Medecine Therapeutique*, vol. 20(2), pp. 67-78, John Libbey DOI : 10.1684/met.2014.0449. 2014.

[2] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres and Mohamed Case-Based Reasoning for Providing Personalized Clinical Pathways to Enhance Patients’ Recovery after Digestive Surgery (en cours)

### Conférences internationales

[1] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres, Samy Talha, and Mohamed Hajjam. Ontology-based medical decision support system to enhance chronic patients’ lifestyle within E-care telemonitoring platform. In *Enabling Health Informatics Applications*, volume 213 of *Studies in Health Technology and Informatics* , pages 279 – 282. IOS Press, 2015.

[2] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres, Samy Talha, and Mohamed Hajjam. Hybrid reasoning-based medical platform to assist clinicians in their clinical reasoning process. In *6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–5, July 2015.

[3] Lamine Benmimoune, Amir Hajjam, Parisa Ghodous, Emmanuel Andres, Samy Talha, and Mohamed Hajjam. E-care : a home health monitoring platform to enhance chronic patients’ health status based on data collection and knowledge-based reasoning. *15th International conference E-society 2017* (à paraître).



# TABLE DES MATIÈRES

<b>I</b>	<b>Introduction générale</b>	<b>5</b>
I.1	Contexte et problématique . . . . .	5
I.2	Objectifs . . . . .	8
I.3	Organisation de la thèse . . . . .	10
<b>I</b>	<b>Synthèse des travaux de l'état de l'art</b>	<b>13</b>
<b>II</b>	<b>Représentation des connaissances médicales</b>	<b>15</b>
II.1	Introduction . . . . .	15
II.2	Typologie des ressources termino-ontologiques . . . . .	16
II.2.1	Terminologie . . . . .	16
II.2.2	Classification . . . . .	16
II.2.3	Nomenclature . . . . .	16
II.2.4	Thésaurus . . . . .	17
II.2.5	Ontologie . . . . .	17
II.2.5.1	Définition . . . . .	17
II.2.5.2	Classification des ontologies . . . . .	18
II.2.5.3	Langages et outils ontologiques . . . . .	19
II.3	Quelques ressources termino-ontologiques du domaine médical . . . . .	22
II.3.1	Thésaurus MeSH . . . . .	22
II.3.2	Nomenclature SNOMED . . . . .	23
II.3.3	CIM . . . . .	24
II.3.4	ATC . . . . .	25
II.3.5	Méta-thésaurus UMLS . . . . .	26
II.3.6	GALEN . . . . .	27
II.4	Conclusion . . . . .	27

<b>III Acquisition de données médicales</b>	<b>29</b>
III.1 Introduction . . . . .	29
III.2 Outils d'acquisition de données pour les SADM . . . . .	30
III.2.1 Formulaires électroniques . . . . .	30
III.2.2 Capteurs connectés . . . . .	31
III.2.3 Questionnaires informatisés . . . . .	32
III.2.3.1 Composition de questionnaire . . . . .	32
III.2.3.2 Utilisation des questionnaires comme outil d'acquisition pour les SADM . . . . .	33
III.3 Conclusion . . . . .	34
<b>IV Raisonnement à base de connaissances</b>	<b>37</b>
IV.1 Introduction . . . . .	37
IV.2 Raisonnement à partir de cas (RàPC) . . . . .	38
IV.2.1 Qu'est-ce que le RàPC . . . . .	38
IV.2.2 Notions de base . . . . .	40
IV.2.2.1 Notion de cas . . . . .	40
IV.2.2.2 Notion de similarité . . . . .	40
IV.2.2.3 La similarité globale . . . . .	44
IV.2.3 Pondération des attributs . . . . .	45
IV.2.4 SADM fondé sur RàPC . . . . .	46
IV.2.4.1 Architecture . . . . .	46
IV.2.4.2 Exemples de SADM fondés sur un RàPC . . . . .	46
IV.2.5 Avantages et inconvénients . . . . .	48
IV.3 Raisonnement à partir de règles (RàPR) . . . . .	48
IV.3.1 Qu'est-ce que RàPR . . . . .	48
IV.3.2 Notions de base . . . . .	49
IV.3.2.1 Règles . . . . .	49
IV.3.2.2 Faits . . . . .	49
IV.3.3 SADM fondé sur RàPR . . . . .	49
IV.3.3.1 Architecture . . . . .	49

IV.3.3.2 Exemples de SADM fondés sur un RàPR . . . . .	51
IV.3.4 Avantages et inconvénients . . . . .	52
IV.4 Les systèmes fondés sur un raisonnement mixte . . . . .	52
IV.5 Conclusion . . . . .	53
<b>II Contribution</b>	<b>55</b>
<b>V Contribution à la représentation des connaissances médicales</b>	<b>57</b>
V.1 Introduction . . . . .	57
V.2 Représentation de connaissance de domaine . . . . .	58
V.2.1 Représentation des connaissances médicales . . . . .	58
V.2.1.1 Ontologie CIM-10 . . . . .	58
V.2.1.2 Ontologie ATC . . . . .	59
V.2.2 Représentation des connaissances liées au domaine d'intervention du SADM . . . . .	60
V.3 Représentation du profil du patient . . . . .	62
V.4 Conclusion . . . . .	65
<b>VI Contribution à l'acquisition de données médicales</b>	<b>67</b>
VI.1 Architecture . . . . .	68
VI.1.1 Module de configuration . . . . .	68
VI.1.1.1 Gestion de questionnaires . . . . .	69
VI.1.1.2 Gestion de l'historique des interrogatoires . . . . .	69
VI.1.2 Module de collecte . . . . .	70
VI.2 Modélisation des connaissances . . . . .	70
VI.2.1 Ontologie Questionnaire . . . . .	70
VI.2.2 Ontologie Interrogatoire . . . . .	73
VI.2.3 Création d'un questionnaire guidée par l'ontologie de domaine DTOnto	74
VI.3 Déroulement de la collecte des données . . . . .	75
VI.3.1 Mode d'acquisition séquentiel . . . . .	75
VI.3.1.1 Créer un nouvel interrogatoire . . . . .	76
VI.3.1.2 Modifier un interrogatoire . . . . .	77

VI.3.1.3	Reprendre un interrogatoire . . . . .	78
VI.3.2	Mode d'acquisition instantané . . . . .	78
VI.3.2.1	Création d'un nouvel d'un interrogatoire . . . . .	79
VI.3.2.2	Modifier un interrogatoire . . . . .	80
VI.3.2.3	Reprendre un interrogatoire . . . . .	81
VI.4	Conclusion . . . . .	81
<b>VII</b>	<b>Contribution dans le raisonnement à base de connaissances</b>	<b>83</b>
VII.1	Introduction . . . . .	83
VII.2	Raisonnement à base de règles . . . . .	84
VII.2.1	Construction de la base de règles . . . . .	86
VII.2.2	Construction de la base de faits . . . . .	87
VII.2.3	Processus de raisonnement à base de règle . . . . .	88
VII.3	Raisonnement à Partir des cas . . . . .	89
VII.3.1	Représentation de cas . . . . .	90
VII.3.1.1	Descripteur du problème . . . . .	90
VII.3.1.2	Solution . . . . .	90
VII.3.1.3	Résultat et évaluation . . . . .	91
VII.3.2	Processus de raisonnement . . . . .	92
VII.3.2.1	Recherche des cas les plus similaires . . . . .	92
VII.3.2.2	Étape de réutilisation . . . . .	96
VII.3.2.3	Étape de mémorisation . . . . .	97
VII.3.3	Modes d'exécution de raisonnement . . . . .	97
VII.3.3.1	Exécution à la demande . . . . .	97
VII.3.3.2	Exécution à la collecte de données . . . . .	97
VII.4	Conclusion . . . . .	99
<b>III</b>	<b>Implémentation et expérimentation</b>	<b>101</b>
<b>VIII</b>	<b>Implémentation</b>	<b>103</b>
VIII.1	Introduction . . . . .	104

VIII.2	Plateforme E-care . . . . .	104
VIII.2.1	Architecture générale de la plateforme E-care . . . . .	104
VIII.2.2	Rôles de la plateforme E-care . . . . .	106
VIII.2.3	Langages et outils utilisés pour le développement de la plateforme . . . . .	106
VIII.3	Acquisition de connaissances spécifiques du domaine d'intervention . . . . .	107
VIII.3.1	Construction de l'ontologie DTOnto . . . . .	107
VIII.3.2	Création d'instances de l'ontologie DTOnto via le portail web . . . . .	107
VIII.4	Profil du patient . . . . .	109
VIII.4.1	Construction de l'ontologie PPO . . . . .	109
VIII.4.1.1	Création de classes . . . . .	109
VIII.4.1.2	Création de propriétés . . . . .	109
VIII.4.2	Création de compte patient . . . . .	110
VIII.4.3	Renseignement des antécédents médicaux . . . . .	111
VIII.5	Outil d'acquisition de données . . . . .	112
VIII.5.1	Construction des ontologies . . . . .	112
VIII.5.1.1	Création de l'ontologie Questionnaire . . . . .	112
VIII.5.1.2	Création de l'ontologie Interrogatoire . . . . .	114
VIII.5.1.3	Construction de l'ontologie protocole de collecte . . . . .	115
VIII.5.2	Création de questionnaires via le portail web . . . . .	116
VIII.5.2.1	Création de questionnaire . . . . .	116
VIII.5.2.2	Définition de protocole questionnaire . . . . .	119
VIII.5.3	Module de collecte de données . . . . .	120
VIII.5.3.1	Collecte de données via le portail web . . . . .	120
VIII.5.3.2	Collecte de données via l'application mobile . . . . .	122
VIII.5.4	Module de l'historique des interrogatoires . . . . .	124
VIII.6	Moteur de raisonnement à base de règles . . . . .	125
VIII.6.1	Gestion des règles d'inférence . . . . .	125
VIII.6.2	Exécution des règles . . . . .	126
VIII.6.3	Inférence des données à partir des règles . . . . .	127
VIII.6.3.1	Indicateurs . . . . .	127
VIII.6.3.2	Alertes . . . . .	128



VIII.6.3.3	Recommandations . . . . .	128
VIII.7	Moteur de raisonnement à base de cas . . . . .	129
VIII.7.1	Gestion des thérapies médicales . . . . .	129
VIII.7.2	Recherche des cas similaires . . . . .	130
VIII.7.2.1	Recherche des profils . . . . .	130
VIII.7.2.2	Recommandations des thérapies . . . . .	131
VIII.8	Conclusion . . . . .	133
<b>IX</b>	<b>Expérimentation et évaluation</b>	<b>135</b>
IX.1	Introduction . . . . .	135
IX.2	Expérimentation . . . . .	136
IX.2.1	Projet Repose . . . . .	136
IX.2.1.1	Définition de contexte du patient . . . . .	136
IX.2.1.2	Définition des connaissances du domaine . . . . .	137
IX.2.1.3	Création de questions . . . . .	137
IX.2.1.4	Définition des règles d'inférence . . . . .	137
IX.2.2	Projet Brave . . . . .	138
IX.2.2.1	Définition de contexte de patient . . . . .	139
IX.2.2.2	Définition des connaissances du domaine . . . . .	139
IX.2.2.3	Création de questions . . . . .	140
IX.2.2.4	Définition de règles d'inférence . . . . .	140
IX.2.3	Projet Chariot EHPAD . . . . .	140
IX.2.3.1	Définition du contexte de résident . . . . .	141
IX.2.3.2	Définition des connaissances de domaine . . . . .	141
IX.2.3.3	Création des questions . . . . .	141
IX.2.3.4	Définition de règles d'inférence . . . . .	142
IX.3	Évaluation . . . . .	142
IX.3.1	Collecte de données . . . . .	142
IX.3.2	Raisonnement à base de cas . . . . .	143
IX.3.2.1	Évaluation de deux modes de calcul des similarités . . . . .	143
IX.3.2.2	Évaluation expérimentale des similarités . . . . .	144

IX.3.2.3 Évaluation des similarités en milieu hospitalier . . . . .	145
IX.3.3 Raisonnement à base de règles . . . . .	146
IX.4 Conclusion . . . . .	146
<b>IV Conclusion générale</b>	<b>149</b>
<b>X Conclusion générale</b>	<b>151</b>
X.1 Synthèse des contributions . . . . .	151
X.2 Perspectives . . . . .	152



# TABLE DES FIGURES

I.1	Modèle global d'un SADM . . . . .	7
II.1	Emboîtement des définitions génériques de ressources sémantiques (DA-SILVA, 2007) . . . . .	21
II.2	Extrait de l'arborescence C (domaine Maladie) de MeSH . . . . .	22
II.3	La classification complète de la metformine dans ATC . . . . .	26
III.1	L'ontologie de questionnaire proposée dans (Bouamrane et al., 2008b) (vue à travers Protégé – OWL) . . . . .	33
IV.1	Processus de raisonnement à base de cas (Aamodt and Plaza, 1994) . . . . .	38
IV.2	Différentes structures d'un cas (Begum, 2009) . . . . .	40
IV.3	Deux concepts et leur subsumant commun le plus spécifique dans une taxinomie. . . . .	43
IV.4	Modèle global d'un système à base de cas . . . . .	46
IV.5	Modèle global d'un système à base de règles . . . . .	50
V.1	Représentation simplifiée de l'ontologie DTOnto . . . . .	61
V.2	Extrait de l'ontologie DTOnto dans le domaine de la chirurgie . . . . .	61
V.3	Représentation simplifiée de l'ontologie Patient Profile Ontology (PPO) . . . . .	63
V.4	Exemple d'instanciation du profil du patient . . . . .	64
VI.1	Architecture de l'outil d'acquisition de données . . . . .	68
VI.2	Représentation simplifié de l'ontologie questionnaire . . . . .	71
VI.3	Exemple de la modélisation d'une question adaptative . . . . .	73
VI.4	Représentation simplifiée de l'ontologie interrogatoire . . . . .	74
VI.5	Questionnaire guidé par l'ontologie de domaine DTOnto . . . . .	75
VI.6	Déroulement de l'interrogatoire en mode séquentiel à l'aide d'une pile . . . . .	76
VI.7	Déroulement de l'interrogatoire en mode instantané à l'aide d'une liste . . . . .	79

VII.1	Représentation d'une règle . . . . .	86
VII.2	Exemple de représentation de règle pour la détection d'anomalie pour un insuffisant cardiaque . . . . .	87
VII.3	Processus d'inférence . . . . .	88
VII.4	Architecture orientée service pour la délocalisation du raisonneur. . . . .	89
VII.5	Exemple d'une représentation simplifiée de deux parties problème-solution d'un cas . . . . .	91
VII.6	Architecture orienté service pour le raisonnement à base de cas . . . . .	98
VIII.	Architecture globale de la plateforme E-care . . . . .	105
VIII.	Création de classes de l'ontologie DTonto sous Protégé . . . . .	107
VIII.	Ajout d'une connaissance spécifique du domaine (type de donnée) depuis le portail web . . . . .	108
VIII.	Ajout d'un type de mesure depuis le portail web . . . . .	108
VIII.	Création de classes pour l'ontologie profil patient . . . . .	109
VIII.	Propriétés d'objet de l'ontologie Profile patient . . . . .	110
VIII.	Création de compte patient . . . . .	110
VIII.	Historique médical du patient . . . . .	111
VIII.	Chargement de la liste des pathologies depuis l'ontologie CIM-10. . . . .	112
VIII.	Création de classes de l'ontologie questionnaire sous Protégé. . . . .	113
VIII.	Création de propriétés de l'ontologie questionnaire sous Protégé . . . . .	113
VIII.	Création de classes de l'ontologie interrogatoire sous Protégé . . . . .	114
VIII.	Création de propriété d'objet de l'ontologie interrogatoire sous Protégé . . . . .	114
VIII.	Création de classes de l'ontologie protocole sous Protégé . . . . .	115
VIII.	Création de propriétés d'objet de l'ontologie protocole sous Protégé . . . . .	116
VIII.	Étape 1 de la création d'un questionnaire via le portail web : renseignement des informations générales . . . . .	117
VIII.	Étape 2 de la création d'un questionnaire via le portail web : création de sous-questionnaires . . . . .	117
VIII.	Étape 3 de la création d'un questionnaire via le portail web : création des questions . . . . .	118
VIII.	Configuration de protocole pour le questionnaire . . . . .	119
VIII.	Affichage des questionnaires en fonction du statut du patient . . . . .	120

VIII.21	Interrogatoire en mode séquentiel via le portail web de la plateforme . . . .	121
VIII.22	Interrogatoire en mode instantané via le portail web de la plateforme . . . .	122
VIII.23	Chargement des questionnaires en fonction du statut du patient sur l'appli- cation mobile de la plateforme. . . . .	122
VIII.24	Interrogatoire en mode séquentiel via l'application mobile de la plateforme .	123
VIII.25	Interrogatoire en mode instantané via l'application mobile de la plateforme	124
VIII.26	Liste de l'historiques des interrogatoires par patient . . . . .	124
VIII.27	Affichage d'un historique d'interrogatoire . . . . .	125
VIII.28	Éditeur de règle via le portail web . . . . .	126
VIII.29	Exemple d'une notification encapsulée dans un objet JSON . . . . .	126
VIII.30	Affichage des indicateurs et détail de calcul de score pour l'indicateur ERAS	127
VIII.31	Affichage des alertes sur le portail web . . . . .	128
VIII.32	Affichage des recommandations sur l'application mobile de la plateforme . .	129
VIII.33	Ajout d'une nouvelle thérapie . . . . .	130
VIII.34	Recherche des profils les plus similaires . . . . .	131
VIII.35	Recommandation de thérapies . . . . .	132
VIII.36	Adaptation d'une thérapie recommandée . . . . .	133
IX.1	Le temps de la recherche des cas (Millisecondes) en fonction du nombre de critères sélectionnés selon les deux modes d'exécutions . . . . .	143
IX.2	Évaluation des résultats de similarités en milieu hospitalier . . . . .	146



# INTRODUCTION GÉNÉRALE

## I.1 Contexte et problématique

Afin d'assurer la pérennisation de notre système de santé universel, tout en dispensant des soins efficaces et toujours de meilleure qualité, il devient indispensable de disposer de nouveaux systèmes informatiques de suivi et de prise en charge des patients dans toutes démarches de soins appliquées à ces derniers.

Les travaux de recherche dans le domaine de l'intelligence artificielle appliquée à la médecine ont permis l'émergence des systèmes d'aide à la décision médicale (SADM). Ces derniers sont conçus pour aider les cliniciens à prendre de meilleures décisions médicales (Shortliffe and Cimino., 2006; Eta S, 2007; Adam Wright, 2009). Ils permettent de leur fournir les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées, correctement filtrées et présentées (J.L. Renaud-Salis, 2010).

Afin d'améliorer la qualité des soins des patients, différents SADM de suivi et de prise en charge médicale ont été construits et déployés dans le milieu hospitalier. Ces SADM accompagnent les professionnels de santé dans leurs démarches de soins et leurs prises de décisions. En effet, ces systèmes permettent principalement de collecter des données médicales sur les patients, de les analyser et de présenter les résultats en fonction du domaine médical et du mode d'intervention du système (télésurveillance médicale, diagnostics médicaux, prise en charge thérapeutique, prescription médicamenteuse, etc.) (Frederic, 2008; Pearson et al., 2009).

Les résultats fournis présentent un appui et une aide aux professionnels de santé dans leur prise de décision par rapport à l'évolution de l'état de santé des patients suivis. Ils peuvent être présentés sous forme : de recommandations, de suggestions ou de conseils (Hussein et al., 2012), de prédictions de risques pour certaines maladies (J. Soni, 2011), de diagnostics différentiels (Z. El Balaa, 2003; P. Ramnarayan, 2004; Chakraborty et al., 2011) ou bien sous forme d'alertes. Une alerte peut être déclenchée, par exemple, au moment de la prescription d'un médicament présentant un risque de toxicité rénale chez un patient dont le taux de créatinine est élevé (J.L. Renaud-Salis, 2010). Ces alertes peuvent être utilisées aussi lors de la surveillance des patients atteints de maladies chroniques, afin de prévenir



le corps médical lorsque le système détecte une anomalie ou une situation dangereuse chez le patient (Benyahia et al., 2012; Minutolo et al., 2010).

Le modèle global d'un SADM tel qu'il est présenté dans la littérature (Spooner, 2007; Guilan kong, 2008) se compose essentiellement de trois couches (voir la figure I.1) :

- **Couche d'acquisition des données** : cette couche permet l'acquisition et l'introduction de données au système. Ces données sont relatives au patient, elles peuvent être de différentes natures : données médicales (signes, symptômes, antécédents médicaux, données physiologiques, etc.), données personnelles (sexe, âge, profession, etc.), données sur l'hygiène de vie (activité physique, tensions artérielles, taux de glycémies, etc.), etc. La manière dont les données sont introduites varie d'un système à l'autre. Certains systèmes imposent aux médecins de sélectionner les données depuis un vocabulaire contrôlé à l'aide de formulaires de saisie (Spooner, 2007; J. Soni, 2011). D'autres systèmes extraient ces données depuis les dossiers médicaux électroniques (DMÉs) (Hussein et al., 2012; Ha and Zhang, 2010). Des capteurs médicaux (tensiomètre, oxymètre, pèse-personne, etc.) (Benyahia et al., 2012, 2013) et des questionnaires informatisés peuvent aussi être utilisés pour l'acquisition de données médicales (Bouamrane et al., 2008b,a; Sherimon, 2013; Sherimon et al., 2014b).
- **Couche décisionnelle** : appelée aussi couche de raisonnement, c'est la partie intelligente du système qui permet, à partir des données introduites par la couche d'acquisition de données, de fournir des résultats à la couche de présentation. Nous pouvons distinguer un ou plusieurs modes de raisonnement. Ces modes de raisonnement permettant d'arriver à un résultat voire à plusieurs résultats pouvant être intriqués. Généralement les modes de raisonnement font appel à différents aspects de l'intelligence artificielle et des techniques de fouille de données dont le but est de simuler le raisonnement humain et, en l'occurrence, le raisonnement des experts médicaux.
- **Couche de présentation des données** : cette couche permet de présenter, à l'utilisateur, les résultats fournis par le système. Plusieurs types et formes de résultats peuvent être donnés en fonction du mode d'intervention du SADM.

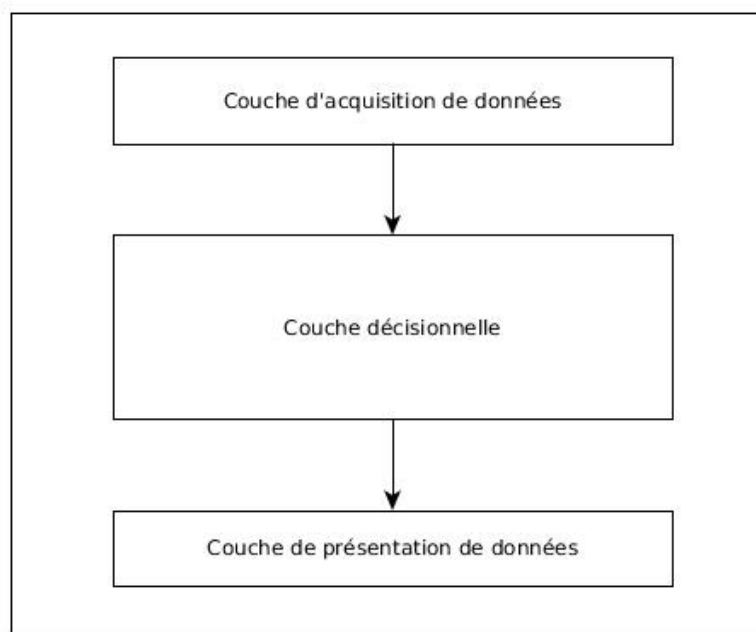


FIGURE I.1 – Modèle global d'un SADM

Plusieurs études ont été menées dans cet axe de recherche parmi lesquelles (Dinevski et al., 2011; Eta S, 2007) qui proposent de catégoriser les SADM selon leurs modes de raisonnement et le type de connaissances qu'ils utilisent. Nous distinguons parmi deux grandes catégories de SADM à savoir : (i) les SADM à base d'apprentissage automatique : ce type de SADM se base sur l'apprentissage automatique de la machine et emploie des modèles statistiques (réseaux de neurones, algorithmes génétiques, arbres de décision, etc.) et (ii) les SADM à base de connaissances : ce type de SADM exploite les connaissances des experts du domaine pour la résolution de problèmes cliniques.

Dans cette thèse, nous nous intéressons à la construction des SADM à base de connaissances. En effet, les SADM à base de connaissances sont très répandus dans le domaine médical et les résultats fournis par ces derniers sont plus crédibles auprès des professionnels de santé, en raison de l'exploitation de leurs connaissances du domaine.

De plus, dans les systèmes à base de connaissances, les connaissances sont généralement séparées du moteur de raisonnement. Ceci présente un énorme avantage dans la maintenabilité du système. La mise à jour des connaissances n'impacte ni la structure du système ni le mode de raisonnement employé.

Les SADM fondés sur le raisonnement à base de connaissances sont constitués essentiellement de deux parties :

- **La Base de connaissances** qui contient les connaissances du domaine médical nécessaires à la démarche médicale (des classifications de pathologies, de symptômes, de traitements, etc.). Cette base peut contenir également des

procédures qui permettent d'engendrer des résultats tels que : l'élaboration d'un diagnostic, la définition des protocoles de soin, la prise en charge thérapeutique, etc. (Alsun, 2012). Ces procédures peuvent être issues des guides de bonnes pratiques (GBPs) représentés sous forme de règles et les expériences cliniques représentées sous forme de cas (Wang and Tansel, 2013).

Le raisonnement sur des connaissances implique leur formalisation selon un certain modèle de représentation. De nombreux modèles de représentation de connaissances, que l'on trouve dans la littérature, utilisent une représentation à l'aide d'ontologies (Bouamrane et al., 2011; Benyahia et al., 2012). L'utilisation des ontologies permet de gérer efficacement un vaste répertoire de connaissances du domaine qui sont représentées par des concepts et des relations entre concepts. Ces connaissances tiennent compte de la classification, signes et symptômes des maladies, des interventions chirurgicales, de la classification des morbidités, etc. (Bouamrane et al., 2011).

La base de connaissance contient aussi les caractéristiques du patient relatives à sa situation clinique. Ces caractéristiques sont appelées des faits et constituent la base de faits sur laquelle le système se base pour fournir des résultats.

- **Le Moteur de raisonnement** appelé aussi moteur d'inférence : c'est un processus qui permet de fournir des solutions à un problème en faisant appel soit à différentes règles, soit à des cas. Il exploite les connaissances de la base de connaissances pour raisonner sur le problème.

Cependant, plusieurs exigences sont à relever dans la construction de tels systèmes où la compréhension du domaine médical doit être maîtrisée. L'adaptation du système, selon le domaine médical et son mode d'intervention, doit être faite de manière simple et rapide. De plus, le système doit être alimenté d'une façon permanente par des connaissances de domaine issues des experts médicaux, limitant ainsi la maintenance récurrente pour chaque connaissance ajoutée.

Afin de répondre à toutes ces exigences, la construction d'un système d'aide à la décision médicale générique et ouvert est primordiale afin qu'il puisse s'adapter au domaine médical concerné et au mode d'intervention approprié.

## I.2 Objectifs

L'objectif de nos travaux de recherche vise à définir une approche pour la construction d'un système d'aide à la décision médicale à base de connaissances, ouvert et générique. Ce système a pour but d'assister les professionnels de santé dans leur démarche de soin. Il permet principalement la représentation des connaissances de domaine sous format ontologique facilement interprétables et exploitables, l'organisation de la collecte des données médicales en tenant compte du contexte du patient et l'exploitation des guides de bonnes pratiques et de l'expérience clinique partagée.

Plus précisément, nos contributions dans le domaine vont porter sur trois volets majeurs :

- **Représentation des connaissances** : nous proposons une représentation sémantique des connaissances basée sur des ontologies à deux niveaux :
  - Le premier niveau concerne la modélisation des connaissances du domaine médical basée sur des ontologies de domaine construites à partir des ressources existantes dans la littérature. Les ontologies permettent d’offrir une représentation à la fois exhaustive et détaillée sur le domaine médical considéré. Une telle représentation permet d’assurer le contrôle du vocabulaire et des termes ainsi que de faciliter le partage des connaissances. De plus, l’utilisation croissante des ontologies, dans le domaine médical, entraîne une disponibilité importante de ces dernières.
  - Le deuxième niveau concerne la modélisation du profil du patient, communément appelée dans l’environnement hospitalier le Dossier Médical Électronique (DMÉ). Pour la modélisation du profil du patient, nous nous sommes reposés sur une approche de représentation développée dans un travail précédent au sein de notre équipe dans (Benyahia, 2015) et nous l’avons étendue pour une plus grande généralité. Cette représentation est à base d’ontologie où les données sont représentées par des concepts issus d’une ontologie de domaine. Ceci permet d’offrir une plus grande souplesse pour la représentation du profil et son évolution, par de nouveaux concepts, ne nécessite aucune intervention sur la structure.
- **Acquisition de données médicales** : toute démarche de soin médicale nécessite un interrogatoire qui doit avoir lieu entre le clinicien et le patient. Cet interrogatoire représente un élément primordial pour la prise en charge du patient. Il permet le plus souvent, à lui seul, de poser un diagnostic ou tout au moins d’orienter le clinicien dans sa réflexion. Nous visons ici à organiser la collecte de données en proposons un outil d’acquisition basé sur un mécanisme de questions/ réponses. Cet outil est construit à base d’ontologie et permet de créer et de fournir automatiquement des questionnaires adaptatifs et sensibles au contexte du patient interrogé et à l’évolution de son état de santé. Il est évolutif et offre la possibilité de changer la structure des questionnaires de manière dynamique quand le besoin se présente.
- **Raisonnement mixte à base de connaissances** : nous proposons l’exploitation des deux sources de connaissances dont l’une complète l’autre : une source théorique (i.e. les guides de bonnes pratiques) et une source pratique (i.e. l’expérience clinique). La disponibilité de ces deux sources dans le même système nous permet d’utiliser à la fois un raisonnement à partir des règles (RàPR) et un raisonnement à partir de cas (RàPC), ce qui permet d’affiner le résultat proposé par le système.
- **Exploitation des guides de bonnes pratiques (GBP)** : nous exploitons les GBPs qui sont considérés comme une source crédible offrant une synthèse des différentes modalités de prise en charge médicale. Ils permettent de fournir les connaissances théoriques (règles) à notre système pour leur appliquer un raison-

nement à partir des règles (RàPR). L'originalité de notre approche réside dans l'intégration d'une ontologie de domaine pour la définition des règles d'inférences indépendamment du domaine d'application du système et de son mode d'intervention.

- **Exploitation de l'expérience** : nous exploitons un raisonnement à partir de cas (RàPC) pour la capitalisation de l'expérience et la résolution de nouveaux cas cliniques. L'originalité de notre approche réside ici dans la définition d'une structure évolutive de cas guidée par une ontologie de domaine qui permet la construction d'un système plus ouvert.

### I.3 Organisation de la thèse

Cette thèse est constituée de trois parties principales : la première présente la synthèse des travaux de l'état de l'art, la seconde partie détaille nos principales contributions et la troisième décrit l'implémentation de nos travaux ainsi que les différentes expérimentations et évaluations menées.

La thèse est organisée comme suit :

- Le chapitre 1 : « Introduction générale »  
Il présente le contexte, les problématiques de recherche abordées et les contributions issues de nos travaux.

La première partie de cette thèse, intitulée *Synthèse des travaux de l'état de l'art*, présente le contexte de nos travaux.

Compte tenu du cadre de notre thèse, nous avons axé l'état de l'art sur les travaux de construction de systèmes d'aide à la décision médicale. Cette partie englobe trois chapitres :

- Le chapitre 2 : « Représentation des connaissances médicale »  
Il présente les notions de base relatives à la représentation des connaissances médicales pour les SADM à base de connaissances. Il présente ainsi une revue de l'état de l'art sur les différentes approches proposées pour les représenter.
- Le chapitre 3 : « Acquisition de données médicales »  
Il présente une synthèse de l'état de l'art sur les différents outils d'acquisition de données médicales au sein d'un SADM.
- Le chapitre 4 : « Raisonnement à base de connaissances »  
Il présente deux modes de raisonnement à base de connaissances : le raisonnement à base de règles et le raisonnement à base de cas. Pour chacun de ces modes, un état de l'art est proposé.

La deuxième partie de cette thèse, intitulée *Contribution*, présente nos contributions relatives à la construction d'un système d'aide à la décision médicale à base de connaissances dédié au suivi médical des patients. Cette partie englobe trois chapitres :

- Le chapitre 5 : « Contribution à la représentation des connaissances médicales »

Il est consacré à notre première contribution. Il définit la problématique et les motivations de l'approche de modélisation de connaissances de domaine proposée.

Il présente cette approche ainsi que la représentation du profil du patient.

- Le chapitre 6 : « Contribution à l'acquisition des données médicales »

Il présente la deuxième partie de nos contributions, relative à l'acquisition de données médicales au sein d'un SADM. Il décrit notre approche de construction d'un outil de collecte de données, basé sur l'utilisation des ontologies, et le mécanisme de questions/réponses sensible au contexte.

- Le chapitre 7 : « Contribution au raisonnement à base de connaissances »

Il présente la troisième partie de nos contributions relatives à l'intégration d'un mode de raisonnement à base de connaissances génériques. Il présente notre approche pour l'exploitation des guides de bonnes pratiques, sous forme de règles, et les expériences cliniques, sous forme de cas.

La troisième partie de cette thèse, intitulée *Implémentation et expérimentations*, dresse le cadre expérimental de nos travaux et présente les résultats obtenus de l'évaluation des approches proposées. Cette partie englobe deux chapitres :

- Le chapitre 8 : « Implémentation »

Il présente l'implémentation de nos travaux au sein de la plateforme E-care qui est dédiée au suivi médical des patients.

- Le chapitre 9 : « Expérimentations et évaluations »

Il présente les différentes expérimentations menées pour l'évaluation du système implémenté ainsi que les différents résultats obtenus.

Le chapitre 10 conclut la thèse, discute de l'impact de nos contributions et, pour finir, présente nos perspectives pour un futur travail.



# I

## SYNTHÈSE DES TRAVAUX DE L'ÉTAT DE L'ART





# II

## REPRÉSENTATION DES CONNAISSANCES MÉDICALES

Dans ce chapitre nous présentant un état de l'art sur les différentes techniques de représentation de connaissances médicales.

### Sommaire

---

<b>II.1 Introduction</b> . . . . .	<b>15</b>
<b>II.2 Typologie des ressources termino-ontologiques</b> . . . . .	<b>16</b>
II.2.1 Terminologie . . . . .	16
II.2.2 Classification . . . . .	16
II.2.3 Nomenclature . . . . .	16
II.2.4 Thésaurus . . . . .	17
II.2.5 Ontologie . . . . .	17
<b>II.3 Quelques ressources termino-ontologiques du domaine médical</b> . . . . .	<b>22</b>
II.3.1 Thésaurus MeSH . . . . .	22
II.3.2 Nomenclature SNOMED . . . . .	23
II.3.3 CIM . . . . .	24
II.3.4 ATC . . . . .	25
II.3.5 Méta-thésaurus UMLS . . . . .	26
II.3.6 GALEN . . . . .	27
<b>II.4 Conclusion</b> . . . . .	<b>27</b>

---

### II.1 Introduction

Depuis les années cinquante les technologies numériques ont donné lieu à un accroissement considérable du volume d'information. Pour faciliter la représentation et l'accès à une telle quantité d'informations, plusieurs ressources terminologiques ont été conçues. Ces ressources ont évolué d'une liste de termes techniques d'un domaine (une terminologie) à un ensemble structuré de termes et de concepts représentant le sens d'un champ d'informations (une ontologie).

Dans ce qui suit, nous donnons quelques définitions des notions élémentaires en ingénierie de la connaissance et les vocabulaires contrôlés les plus utilisés dans le domaine médical.

## II.2 Typologie des ressources termino-ontologiques

### II.2.1 Terminologie

Une terminologie est un ensemble de termes, rigoureusement définis, spécifiques à une science, à une technique ou encore à un domaine particulier de l'activité humaine (Larousse<sup>1</sup>, 2015). Une terminologie assure l'uniformisation des termes d'un domaine par la notion de "**concept**". L'intérêt majeur d'une terminologie est de réduire, voire supprimer, l'ambiguïté qui existe entre les termes d'un domaine pour faciliter l'échange et le partage de connaissances (Dinh, 2012). En effet, une terminologie de référence spécifie une norme pour un domaine donné, le sens de chaque terme est figé et il n'y a qu'une interprétation possible pour l'utilisateur. Il existe des terminologies de natures diverses adaptées aux différents objectifs de traitement de l'information : classification, nomenclature, thésaurus et ontologie.

### II.2.2 Classification

Une classification est la distribution systématique par classes d'objets ou de notions ayant des caractères communs afin d'en faciliter l'étude (Bourigault, 2004). Suivant les objets considérés (maladies, traitements, actes médicaux, etc.), différentes classifications ont été développés. Elles sont utilisées dans tous les domaines d'activités humaines comme la biologie, la médecine, l'économie, l'informatique, etc. Les classifications qui portent sur un domaine restreint sont généralement bien admises par les experts du domaine. Les classifications de nature universelle sont, de ce fait, l'objet de nombreuses critiques. Elles apportent cependant un éclairage sur la nature de la connaissance (Dinh, 2012). Dans le domaine médical, la Classification Internationale des Maladies (CIM) et la Classification Commune des Actes Médicaux (CCAM) sont de bons exemples de classifications hiérarchiques bien qu'elles n'aient pas le même niveau de profondeur.

### II.2.3 Nomenclature

Une nomenclature est un ensemble de termes en usage dans une science, un art, ou relatifs à un sujet donné, présentés selon une classification méthodique (Larousse, 2015)<sup>2</sup>. Elle désigne une instance de classification (code, tableau, liste, règles d'attribution d'identité, etc.) faisant autorité et servant de référence à une discipline donnée (wikipedia<sup>3</sup>, 2015) .

---

1. <http://www.larousse.fr>

2. <http://www.larousse.fr/dictionnaires/francais/nomenclature/54811>

3. <https://fr.wikipedia.org/wiki/Nomenclature>

Elle permet de décrire les concepts d'un domaine de manière complète sans se restreindre à un objectif spécifique. Une nomenclature importante dans le domaine médical est la Nomenclature SNOMED (Systematized Nomenclature of Medicine–Clinical Terms). Celle-ci est une nomenclature pluri-axiale et multi-domaines couvrant tous les champs de la médecine, de la dentisterie humaine ainsi que de la médecine vétérinaire.

## II.2.4 Thésaurus

Un thésaurus est une liste organisée de termes contrôlés et représentant les concepts d'un domaine de la connaissance (wikipedia<sup>4</sup>, 2015). Ces termes sont organisés sous sur une structuration hiérarchisée et reliés entre eux par des relations sémantiques.

Plus concrètement, un thésaurus est un ensemble structuré de termes d'un vocabulaire, par exemple les termes techniques utilisés en médecine, représentés de façon normalisée par des descripteurs ou des mots-clés. En général, un thésaurus peut être construit automatiquement à partir d'une collection de documents ou manuellement par les experts (linguistes, documentalistes, bibliothécaires).

La construction automatique d'un thésaurus ne garantit pas la qualité des termes désignant les concepts ni les relations entre eux, car ses performances dépendent totalement de la qualité des méthodes du traitement du langage naturel et de la qualité de la collection.

La construction manuelle assure une bonne qualité des termes qui constituent des unités sémantiques, appelées concepts. Par contre, la construction du thésaurus nécessite l'intervention des experts de domaine ce qui la rend assez coûteuses notamment en termes d'expertise et de temps (Dinh, 2012).

## II.2.5 Ontologie

### II.2.5.1 Définition

En philosophie, le terme "**ontologie**" fait référence à l'étude de l'être en tant qu'être, de ses modalités et de ses propriétés (définition proposée par Aristote). Ce terme a été repris en science de l'information et en informatique dans les années quatre-vingt-dix. Sa première définition a été donnée par (Gruber, 1993) : "une ontologie est une spécification d'une conceptualisation". Une autre définition est donnée par (Studer et al., 1998) : "Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée".

Plus formellement, une ontologie est une description des concepts et des relations qui existent pour un objet ou un ensemble d'objets. L'objectif majeur d'une ontologie est donc de partager et de réutiliser des connaissances propres à un domaine donné (Dinh, 2012). De ce fait, une ontologie constitue en soi un modèle de données représentatif, conçue pour

---

4. <https://fr.wikipedia.org/wiki/Thésaurus>

modéliser un ensemble de connaissances d'un domaine spécifique, comme par exemple le domaine de la médecine, le génie logiciel, l'aviation, etc. Une ontologie décrit généralement un ensemble de primitives considérées comme ses principaux constituants :

- **Un concept**, appelé aussi classe, peut représenter un objet matériel (par exemple, une pathologie), une notion (par exemple, la quantité) ou bien une idée. Selon (Gómez-Pérez, 1999) les concepts peuvent être classifiés selon plusieurs niveaux :
  - Niveau d'abstraction (concret ou abstrait) ;
  - Niveau d'atomicité (élémentaire ou composée) ;
  - Niveau de réalité (réel ou fictif).
- **Un attribut**, appelé aussi propriété, fonctionnalité, caractéristique ou paramètre, est associé à un objet et représente des connaissances simples de type numérique (entier, réel, etc.), énumération, chaîne de caractères, etc.
- **Une relation**, reflète les liens que les concepts peuvent avoir entre eux. Les relations incluent, notamment, les associations suivantes :
  - Sous-classe-de (généralisation – spécialisation) ;
  - Partie-de (agrégation ou composition) ;
  - Associée-à ;
  - Instance- de, etc.
- **Un événement**, indique un ou des changements des attributs ou des relations.

### II.2.5.2 Classification des ontologies

Les ontologies peuvent être classifiées selon plusieurs dimensions. Parmi celles-ci, nous citons les classifications suivantes :

#### Classification selon le niveau de formalisation

Dans (Uschold and Gruninger, 1996) les auteurs proposent de classier les ontologies en quatre catégories suivant leurs niveaux de formalisation :

- **Informelles** : Ontologies opérationnelles dans un langage naturel (sémantique ouverte) ;
- **Semi- informelles** : Ontologies définies dans un langage naturel structuré et limité ;
- **Semi-formelles** : Ontologies utilisant un langage artificiel défini formellement ;
- **Formelles** : Ontologies utilisant un langage artificiel contenant une sémantique formelle, ainsi que des théorèmes et des preuves des propriétés telles que la robustesse et l'exhaustivité.

#### Classification selon l'objet de conceptualisation

On distingue des ontologies de différents niveaux de généralité :

- **Ontologies supérieures ou de haut niveau**. Ce type d'ontologie décrit des

concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. qui ne dépendent pas d'un problème ou d'un domaine particulier (Lando, 2006).

Guarino (Guarino, 1997) et Sowa (Sowa, 1995) intègrent les fondements philosophiques comme étant des principes à suivre pour la construction des ontologies de haut niveau ou supérieures. Sowa introduit deux concepts importants, **Continuant** et **Occurrent**, et obtient douze catégories supérieures avec la combinaison de sept propriétés primitives.

L'ontologie supérieure de Guarino consiste en deux mondes : une ontologie des Particuliers (choses qui existent dans le monde) et une ontologie des Universels comprenant les concepts nécessaires à décrire les Particuliers (Psyché et al., 2003).

- **Ontologies de domaine.** Cette ontologie contient un ensemble de vocabulaires et de concepts qui décrivent un domaine d'application ou monde cible (médecine, architecture, aviation, etc.). Elle permet de créer des modèles d'objets du monde cible. L'ontologie de domaine peut être considérée comme une méta-description d'une représentation des connaissances, c'est-à-dire une sorte de méta-modèle de connaissances dont les concepts et propriétés sont de type déclaratif (Psyché et al., 2003). La plupart des ontologies existantes sont des ontologies du domaine.
- **Ontologies de tâches.** Ce type d'ontologies est utilisé pour conceptualiser des tâches spécifiques dans un système (tâches de diagnostic, tâches d'enseignement, etc.) (Mizoguchi and Bourdeau, 2000). Elle contient un ensemble de vocabulaires et de concepts qui décrivent une structure de réalisation des tâches indépendante du domaine (Psyché et al., 2003).
- **Ontologies d'application.** Ce type d'ontologie contient des concepts spécifiques à un domaine donné (Psyché et al., 2003). Autrement dit les concepts correspondent souvent aux rôles joués par les entités du domaine tout en exécutant une certaine activité (Maedche and Staab, 2001). L'auteur dans (Coulet, 2007) souligne que lors de la construction ou le choix d'une ontologie, il est important d'avoir à l'esprit que plus le niveau d'abstraction choisi est proche de l'application moins l'ontologie est réutilisable, mais plus elle est adaptée.

### II.2.5.3 Langages et outils ontologiques

Dans le contexte du Web sémantique, plusieurs langages ont été développés pour la formalisation et l'interrogation des ontologies. Certains sont recommandés par la W3C<sup>5</sup>. Dans la suite de cette section, nous étudierons plus particulièrement les formats RDF, RDF Schéma, OWL et SPARQL.

- **RDF** (Resource Description Framework) est un modèle de données destiné à décrire formellement les ressources Web et leurs métadonnées, de façon à permettre le

---

5. [www.w3.org](http://www.w3.org)

traitement automatique de telles descriptions. Initialement recommandé par le W3C dans le but de standardiser les définitions et les usages des métadonnées.

Les ressources du Web sont l'élément de base de RDF. Chaque ressource est caractérisée par une URI (Uniform Resource Identifier, soit littéralement identifiant uniforme de ressource). L'utilisation des URI pour désigner les ressources décrites par un graphe RDF a plusieurs avantages. Elles permettent de partager le même vocabulaire tout en évitant les ambiguïtés de noms à plusieurs applications (Gueffaz, 2012). Un graphe peut contenir deux types de littéral :

- **Un littéral simple** : est constitué d'une chaîne de caractères, appelée forme lexicale, et d'un attribut facultatif indiquant la langue.
- **Un littéral typé** : est formé d'une chaîne de caractères et d'une URI indiquant un type qui sera utilisé pour décoder cette chaîne.

Un document structuré en RDF est un ensemble de triples :

- **Le sujet** : représente la ressource à décrire ;
- **Le prédicat** : représente un type de propriété applicable à cette ressource ;
- **L'objet** : représente une donnée ou une autre ressource : c'est la valeur de la propriété.

Le modèle des données RDF ne précise que le mode de description des données mais ne fournit pas la déclaration des propriétés spécifiques au domaine ni la manière de définir ces propriétés avec d'autres ressources (Gueffaz, 2012). RDFS (RDF Schema)<sup>6</sup> a pour but d'étendre RDF en fournissant des éléments de bases (classes, sous-classes, propriétés, sous-propriétés, etc.) à la représentation des connaissances notamment à la définition des ontologies et de vocabulaires pour les descriptions RDF.

- **OWL** Web Ontology Language (OWL)<sup>7</sup> créée par W3C en 2001, est un langage de représentation des connaissances construit sur le modèle de données de RDF. Il fournit les moyens pour définir des ontologies web structurées. Sa deuxième version est devenue une recommandation du W3C fin 2012. OWL est composé de trois sous-langages, d'expressivité croissante, à savoir :
  - **OWL Lite** est le sous langage de OWL le plus simple. Il est destiné aux utilisateurs qui ont besoin de modéliser des ontologies simples ayant une hiérarchie de concepts simples. Sa complexité formelle étant peu élevée, il permet d'implémenter facilement des raisonneurs corrects et complets (Benyahia, 2015).
  - **OWL DL** est plus complexe que OWL Lite qui permet une expressivité bien plus importante. Il est fondé sur la logique descriptive (d'où son nom, OWL Description Logics) et permet de garantir la complétude des raisonnements (toutes les inférences sont calculables) et leur décidabilité (leur calcul se fait en une durée finie) (Lacot, 2005). OWL DL comprend toutes les constructions de lan-

---

6. [www.w3.org/TR/rdf-schema/](http://www.w3.org/TR/rdf-schema/)

7. [www.w3.org/2001/sw/wiki/OWL](http://www.w3.org/2001/sw/wiki/OWL)

gage OWL, mais elles ne peuvent être utilisées que sous certaines restrictions (par exemple, les restrictions numériques ne peuvent pas être appliquées sur les propriétés transitives).

- **OWL Full** est la version la plus complexe d'OWL avec le plus haut niveau d'expressivité. OWL Full est destiné aux situations où il est important d'avoir un haut niveau de capacité de description (Lacot, 2005). OWL Full offre cependant des mécanismes intéressants, comme par exemple la possibilité d'étendre le vocabulaire par défaut de OWL. Dans (Lacot, 2005) l'auteur souligne qu'une dépendance hiérarchique existe entre ces trois sous langages : toute ontologie OWL Lite valide est également une ontologie OWL DL valide, et toute ontologie OWL DL valide est également une ontologie OWL Full valide.
- **Protégé**<sup>8</sup> est un éditeur d'ontologies distribué en open source par l'université en informatique médicale de Stanford. Protégé n'est pas un outil spécialement dédié à OWL, mais un éditeur hautement extensible, capable de manipuler des formats très divers. En quelques années, cet éditeur s'est imposé comme la référence, avec une communauté d'utilisateurs extrêmement importante et active. Il offre une multitude de fonctionnalités permettant en particulier de : vérifier la cohérence et la consistance de la structure ontologique à l'aide des raisonneurs intégrés, créer des axiomes formels de manière intuitive, accéder aux ontologies par des interfaces graphiques évoluées, comparer et fusionner des ontologies, etc.

Du point de vue structurel, l'emboîtement des ressources sémantiques représenté dans la figure II.1 évolue d'un arbre taxonomique vers un graphe ontologique.

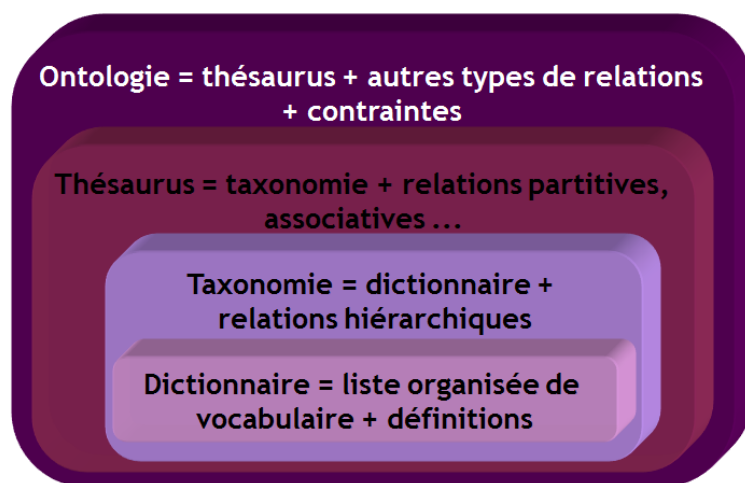


FIGURE II.1 – Emboîtement des définitions génériques de ressources sémantiques (DA-SILVA, 2007)

8. [protege.stanford.edu](http://protege.stanford.edu)



## II.3 Quelques ressources termino-ontologiques du domaine médical

### II.3.1 Thésaurus MeSH

Le thésaurus MeSH<sup>9</sup> (Medical Subject Headings) a été publié en 1954 et mis à jour chaque année par la NLM<sup>10</sup> (The National Library of Medicine). Il est utilisé pour indexer, classer et rechercher des documents biomédicaux (notamment ceux de la base MEDLINE). La traduction de MeSH vers le français est assurée par l'INSERM<sup>11</sup> (Institut National de la Santé Et de la Recherche Médicale) qui met la version bilingue à la disposition de la communauté francophone. La version française de MeSH sert de thésaurus au site CISMef (le site du Catalogue et Index des Sites Médicaux Francophones). MeSH est composé essentiellement de termes qui désignent les concepts biomédicaux, des relations, des descripteurs et des qualificatifs (Dimitrieski et al., 2016). Le rôle de ces éléments est le suivant :

```

C <maladies>
  C04 <tumeurs>
  C18 <métabolisme et nutrition, maladies>
    C18.452 <métabolisme, maladies>
      C18.452.090 <amyloïdose>
      C18.452.394 <troubles du métabolisme glucidique>
        C18.452.394.750 <diabète>
          C18.452.394.750.124 <diabète de type 1>
            C18.452.394.750.124.960 <wolfram, syndrome>
            C18.452.394.750.149 <diabète de type 2>
            C18.452.394.750.774 <état prédiabétique>
          C18.452.394.937 <glycosurie>
        C18.654 <troubles nutrition>
    C23 <troubles liés environnement>
  
```

FIGURE II.2 – Extrait de l’arborescence C (domaine Maladie) de MeSH

- **Terme** : comprend un mot ou un ensemble de mots exprimant une notion particulière ;
- **Concept** : comprend un ou plusieurs termes synonymes et portes le nom d’un de ces termes, dit terme préféré ;
- **Relation** : permet de relier sémantiquement deux concepts. Il existe deux types de relations : les relations hiérarchiques et les relations associatives (associé à). Les relations hiérarchiques dans MeSH sont représentées par un code reflétant l’arborescence auquel le concept appartient (voir figure II.2). Elles peuvent prendre plusieurs sens :
  1. relation “**est-une-partie-de**” (Méronymie), par exemple le concept “doigt” (A01.378.800.667.430) est une partie de “main” (A01.378.800.667).
  2. relation “**est-un-type-de**” (Hyponymie), par exemple le concept pouce

9. <http://mesh.inserm.fr>

10. <http://www.nlm.nih.gov>

11. <http://www.inserm.fr>

(A01.378.800.667.430.705) est un type de “doigt” (A01.378.800.667.430).

3. relation “**est-sémantiquement-proche-de**” (Aboutness), par exemple le concept “sécurité” (G03.850.110.060.075) est sémantiquement proche de “accidents” (G03.850.110).

— **Descripteur** : est constitué d’un ou de plusieurs concepts de significations proches et porte le nom d’un de ces concepts, dit préféré. Les autres concepts, dits subordonnés, sont reliés avec le concept préféré par une relation sémantique. Cette relation peut être soit hiérarchique (générique ou spécifique), soit associative (associé).

Le MeSH comprend 27 149 descripteurs répartis en 16 catégories recouvrant la biologie, la médecine et les domaines connexes

Chaque catégorie de descripteurs est structurée en arborescence hiérarchique et peut comprendre jusqu’à onze niveaux hiérarchiques. Chaque descripteur est représenté par un code alphanumérique, la lettre indiquant la catégorie et la séquence numérique précisant la localisation dans la hiérarchie (voir figure II.2). Certains descripteurs ont plusieurs localisations, au sein de la même catégorie ou de catégories différentes, et plusieurs codes alphanumériques représentant chacun une localisation (Dinh, 2012).

— **Qualificatif** : les qualificatifs sont des entrées MeSH qui peuvent être utilisées seules ou associées à un descripteur pour un aspect particulier (Dinh, 2012). Par exemple, le sens du descripteur “diabète de type 2” associé au qualificatif “thérapeutique” (diabète de type 2/thérapeutique) évoque le traitement du diabète non insulino-dépendant. Alors que seul, le descripteur “diabète de type 2” fait référence à la maladie en général.

### II.3.2 Nomenclature SNOMED

La SNOMED<sup>12</sup> (Systematized Nomenclature of Medicine) est une nomenclature pluri-axiale et multi-domaines couvrant tous les champs de la médecine, de la dentisterie humaine ainsi que de la médecine vétérinaire.

SNOMED-CT (SNOMED Clinical Terms) a été créé en 1999 par la fusion, l’expansion et la restructuration des deux terminologies à grande échelle : SNOMED RT (SNOMED terminologie de référence), développé par le College of American Pathologists (CAP), et la Clinical Terms Version 3 (CTV3), développée par le National Health Service du Royaume-Uni (NHS). Elle représente la dernière version de la nomenclature et comporte 11 axes.

Dans chaque axe, les concepts sont représentés par un code, une série de termes, des synonymes et des définitions. La version française comporte 97 485 concepts désignés par 144 796 termes. Par ailleurs, chaque axe représente une hiérarchie simple de concepts qui peuvent représenter une combinaison de concepts.

12. [www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)

Par exemple, le concept “pemphigoïde bulleux” (D0-10431) est la combinaison des concepts “peau, SAI, cutané” (T-01000), “acantholyse” (M-51551) et “ampoule, SAI ” (M36-760).

### II.3.3 CIM

CIM<sup>13</sup> (Classification Internationale des Maladies ou, en anglais, International Statistical Classification of Diseases and Related Health Problems - ICD) a été conçue par l’Organisation Mondiale de la Santé (OMS) et est utilisée à travers le monde pour codifier et classer les maladies ainsi qu’une très vaste variété de signes, de symptômes, de lésions traumatiques, d’empoisonnements, de circonstances sociales et de causes externes de blessures.

La CIM a été traduite en 43 langues. Environ 117 pays l’utilisent pour communiquer les données de mortalité, les causes de mortalité et les indicateurs principaux de l’état de santé. La version la plus récente étant la 10<sup>ème</sup> révision (CIM-10, publiée en 1993), elle représente une norme internationale pour décrire l’information sur les diagnostics cliniques. Il s’agit d’une classification mono-axiale qui comporte vingt-deux chapitres depuis 2006. Chaque chapitre est divisé en catégories associées à un code à trois caractères, par exemple : asthme J45. La majorité des catégories propose un niveau de détail supplémentaire dont le code est précisé par un quatrième caractère (séparé des trois premiers par un point), par exemple : asthme allergique J45.0. Chaque entité (maladie) ne correspond qu’à un seul code, les ambiguïtés de classement étant levées par les règles d’exclusion. Les vingt-deux chapitres avec l’indication des codes des premières et dernières catégories qu’ils contiennent sont représentés dans le tableau suivant (voir tableau II.1 ).

Chaque entité (maladie) ne correspond qu’à un seul code, les ambiguïtés de classement étant levées par les règles d’exclusion.

---

13. <http://www.who.int/classifications/icd>

Chapitre	Bloc	Titre
I	A00–B99	Certaines maladies infectieuses et parasitaires
II	C00–D48	Tumeurs
III	D50–D89	Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire
IV	E00–E90	Maladies endocriniennes, nutritionnelles et métaboliques
V	F00–F99	Troubles mentaux et du comportement
VI	G00–G99	Maladies du système nerveux
VII	H00–H59	Maladies de l’œil et de ses annexes
VIII	H60–H95	Maladies de l’oreille et de l’apophyse mastoïde
IX	I00–I99	Maladies de l’appareil circulatoire
X	J00–	Maladies de l’appareil respiratoire
XI	K00–K93	Maladies de l’appareil digestif
XII	L00–L99	Maladies de la peau et du tissu cellulaire sous-cutané
XIII	M00–M99	Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif
XIV	N00–N99	Maladies de l’appareil génito-urinaire
XV	O00–O99	Grossesse, accouchement et puerpéralité
XVI	P00–P96	Certaines affections dont l’origine se situe dans la période périnatale
XVII	Q00–Q99	Malformations congénitales et anomalies chromosomiques
XVIII	R00–R99	Symptômes, signes et résultats anormaux d’examens cliniques et de laboratoire, non classés ailleurs
XIX	S00–T98	Lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes
XX	V01–Y98	Causes externes de morbidité et de
XXI	Z00–Z99	Facteurs influant sur l’état de santé et motifs de recours aux services de santé
XXII	U00–U99	Codes d’utilisation particulière

TABLE II.1: Liste des chapitres principaux du CIM

### II.3.4 ATC

ATC<sup>14</sup>(anatomique, thérapeutique et chimique) est une classification de médicament publiée et contrôlée par le Collaborating Centre for Drug Statistics Methodology de l’OMS. Les médicaments dans l’ATC sont divisés en différents groupes selon l’organe ou le système sur lequel ils agissent et/ou leurs caractéristiques thérapeutiques et chimiques. La classification ATC repose sur cinq niveaux de classement qui correspondent aux organes (ou systèmes d’organes) cibles, et aux propriétés thérapeutiques, pharmacologiques et chimiques des différents produits.

La forme générale du code d’une molécule est LCCLCC, où L représente une lettre et C un chiffre (exemple : A01AA01). Chaque lettre et chaque doublet de chiffres représentent un niveau successif. Le premier niveau (première lettre) correspond au groupe anato-

14. <http://www.whocc.no/atc/>

mique parmi 14 différents. Le deuxième niveau (deux premiers chiffres) correspond au sous-groupe pharmacologique ou thérapeutique principal. Le troisième et le quatrième niveau (deuxième et troisième lettres) correspondent à des sous-groupes chimiques, pharmacologiques ou thérapeutiques. Le cinquième et dernier niveau (deux derniers chiffres) indique la substance chimique.

Par exemple la classification complète de la 'metformine' est illustrée par la structure du code suivante (voir figure II.3).

```

A <Voies digestives et métabolisme>
  A10 <Médicaments du diabète>
    A10B <Antidiabétiques sauf insulines>
      A10BA <Biguanides>
        A10BA02 <Metformine>
  
```

FIGURE II.3 – La classification complète de la metformine dans ATC

Les principaux groupes du ATC sont représentés dans le tableau II.2.

Groupe	Titre
A	Système digestif et métabolisme
B	Sang et organes hématopoïétiques
C	Système cardio-vasculaire
D	Dermatologie
G	Système génito-urinaire et hormones sexuelles
H	Hormones systémiques, à l'exclusion des hormones sexuelles et des insulines
J	Anti-infectieux (usage systémique)
L	Antinéoplasiques et agents immunomodulants
M	Système musculo-squelettique
N	Système nerveux
P	Antiparasitaires, insecticides et répulsifs
R	Système respiratoire
S	Organes sensoriels
V	Divers

TABLE II.2: Liste des 14 groupes principaux du ATC

### II.3.5 Méta-thésaurus UMLS

UMLS<sup>15</sup> (Unified Medical Language System) est un système qui recueille plusieurs vocabulaires contrôlés dans les sciences biomédicales mise en place par la NLM (United States National Library of Medicine). Il réunit trois bases de connaissances : Méta-thésaurus, Réseau sémantique, et SPECIALIST Lexicon. Elles peuvent être utilisées séparément ou ensemble.

15. [www.nlm.nih.gov/research/umls](http://www.nlm.nih.gov/research/umls)

Chacune de ces bases constitue une unité dans l'UMLS :

- **Méta-thésaurus** : est une base qui unifie des concepts issus de plus de 150 terminologies biomédicales (dont MeSH, SNOMED et CIM-10) en 17 langues. UMLS regroupe les différents termes synonymes (issus des différentes terminologies) sous un même concept. Chaque concept est caractérisé par un code unique, un terme préféré, un ou plusieurs types sémantiques (défini(s) dans le réseau sémantique de l'UMLS) et une ou plusieurs définitions. Le méta-thésaurus contient aussi des relations entre les concepts (par exemple des relations de hiérarchie, de proximité, etc.). Les versions françaises de la CIM-10 et du SNOMED n'y sont pas encore intégrées. Cependant, la version française du MeSH y est intégrée et elle est régulièrement mise à jour (Dinh, 2012).
- **Réseau sémantique** : est une base qui fournit une catégorisation cohérente de 133 concepts représentés dans la base méta-thésaurus et 54 relations sémantiques entre ces concepts.
- **SPECIALIST Lexicon** : cette base fournit des informations lexicales nécessaires au traitement automatique de la langue (TALN). Chacune de ces informations possède une forme de base (lemme), une catégorie syntaxique, un code unique et éventuellement des variantes orthographiques (Dinh, 2012).

### II.3.6 GALEN

GALEN (Rector A. L., 1996; Rector and Rogers, 2006), est une ontologie haut niveau pour la représentation des concepts du domaine médical. La version initiale de GALEN a été développée en 1995, elle comptait une hiérarchie de plus de 4000 concepts. Actuellement, plus de 52 000 concepts sont recensés. Les concepts de l'ontologie GALEN sont organisés d'une manière hiérarchique formant un réseau sémantique où les différents concepts sont reliés entre eux essentiellement par la relation sémantique "sorte-de". Chaque concept de cette ontologie est complété d'une déclaration des relations qui peuvent le lier à d'autres concepts. Les concepts et relations peuvent être combinés à volonté pour créer de nouveaux concepts structurés.

## II.4 Conclusion

Ce chapitre donne un aperçu sur l'état de l'art des principales ressources termino-ontologiques les plus utilisées dans le domaine médical.

Dans un premier temps, nous avons abordé les différents types de ressources médicales où nous avons montré que les taxonomies et les thésaurus lient leurs concepts propres via les relations spécifiques à chacune, mais n'expriment pas de contraintes pour imposer un sens unique et formel pour un domaine d'application donné. Tandis que les ontologies permettent la définition des relations de types divers et d'exprimer des contraintes. Nous

pouvons dire qu'une ontologie est une ressource plus formelle que celles qu'elle contient. Pour cela elle peut être utilisée pour réaliser des traitements plus complexes tels que l'inférence.

Dans un second temps, nous nous sommes intéressés à quelques ressources utilisées dans la littérature. Nous avons montré leur intérêt dans la représentation des connaissances médicales par la construction des SADM à base de connaissances. Certaines ressources telles que Méta-thésaurus UMLS et Thésaurus MeSH proposent une représentation plus complète sur les connaissances du domaine. Elles offrent plus de relations entre concepts et de définitions utilisés principalement dans les applications de TALN (traitement automatique du langage naturel) et de recherche d'information biomédicale. Etant donné que ces ressources ne font pas l'objet de notre thèse, notre choix s'est porté sur les classifications CIM-10 et ATC qui répondent parfaitement à nos besoins pour la représentation des antécédents et des traitements médicaux. Elles facilitent leur implémentation et proposent une structure indexée et hiérarchisée. Ceci est très utile pour le calcul des similarités sémantiques pour le raisonnement à base de cas.

## ACQUISITION DE DONNÉES MÉDICALES

Dans ce chapitre nous présentons un état de l'art sur les différents modes et outils d'acquisition de données médicales pour les systèmes d'aide à la décision médicale (SADM).

### Sommaire

---

<b>III.1 Introduction</b> . . . . .	<b>29</b>
<b>III.2 Outils d'acquisition de données pour les SADM</b> . . . . .	<b>30</b>
III.2.1 Formulaire électroniques . . . . .	30
III.2.2 Capteurs connectés . . . . .	31
III.2.3 Questionnaires informatisés . . . . .	32
<b>III.3 Conclusion</b> . . . . .	<b>34</b>

---

### III.1 Introduction

Dans le domaine médical, la collecte des données du patient constitue la première étape de la démarche clinique. Celle-ci est l'instrument logique par excellence pour organiser le travail du corps médical. Les données collectées permettent de se situer par rapport aux problèmes du patient et de poser le jugement clinique. De plus, ces données permettent de planifier les interventions nécessaires, d'assurer la surveillance clinique tout au long d'un épisode de soins, de contribuer aux décisions médicales et enfin d'évaluer les soins reçus (Phaneuf, 2012).

De même, sans l'acquisition de données médicales formatées et structurées décrivant la situation clinique du patient, les systèmes d'aide à la décision médicale ne pourraient pas fonctionner. L'objectif des outils d'acquisition de données est donc de faciliter la prise en compte des données médicales tout en s'intégrant dans la démarche médicale des médecins. Ils vont leur permettre une acquisition simple et rapide des données pertinentes dans le contexte spécifique du patient (Jean louis et al., 2010).



## III.2 Outils d'acquisition de données pour les SADM

Plusieurs outils d'acquisition de données peuvent être intégrés au sein des SADM afin de permettre l'introduction des données médicales des patients de manière structurée. Ces outils reposent généralement sur l'utilisation de formulaires électroniques, de capteurs connectés ou encore de questionnaires informatisés.

Les outils d'acquisition de données cliniques peuvent être déclenchés par différents événements : à la demande du clinicien ou par sélection d'un problème du patient et au choix d'un outil d'acquisition à partir de menus. Ils peuvent aussi être déclenchés de manière automatique, par exemple en cas de données manquantes lors de la prescription d'un médicament (mesure de la créatinine par exemple). Dans cette section nous présentons trois outils d'acquisition de données qui peuvent être intégrés dans un SADM :

### III.2.1 Formulaire électroniques

En informatique, un formulaire est un espace de saisie dans l'interface utilisateur, pouvant comporter plusieurs champs à savoir : texte, case à cocher, liste de sélection, boutons, etc. Les formulaires sont largement utilisés dans les systèmes d'information de santé pour l'acquisition de données liées aux patients. La plupart des SADM introduisent les données médicales des patients par le biais de formulaires électroniques. Dans (J. Soni, 2011), les auteurs proposent un SADM pour les prédictions de risque pour des maladies cardiovasculaires. Les données cliniques des patients sont introduites via une interface composée d'un formulaire contenant 14 caractéristiques telle que : âge, sexe, type de la douleur thoracique, tension artérielle, pouls, etc. Dans (Z. El Balaa, 2003; Z. El Balaa and Uziel, 2003), les auteurs proposent un SADM pour le diagnostic des malformations fœtales. Ce système permet l'introduction de données cliniques à l'aide d'un formulaire unique. Celui-ci contient tous les attributs cliniques nécessaires pour la description de la situation clinique de la maman et de son fœtus.

Les auteurs dans (Chakraborty et al., 2011) proposent un SADM appliqué au diagnostic de la grippe porcine. Les données des patients sont introduites par le biais d'un formulaire contenant des informations sur le patient et des signes vitaux telles que : âge, température, nausée, etc.

Dans (P. Ramnarayan, 2004), les auteurs proposent un SADM appliqué au diagnostic de plusieurs pathologies. Les données cliniques du patient (signes vitaux, symptômes, tests laboratoire, etc.) sont exprimées à l'aide d'un texte libre par le biais d'un formulaire informatisé. Ce système utilise le traitement automatique du langage naturel (TALN) pour interpréter les données saisies et pour proposer des diagnostics appropriés.

L'un des principaux inconvénients de l'utilisation des formulaires est la maintenabilité. Celle-ci est généralement coûteuse étant donné que la structure des formulaires est figée.

La maintenance d'un formulaire nécessite de revoir la structure de la base de données, la couche d'échange de données et l'interface graphique. Les formulaires sont souvent développés pour un besoin bien spécifique. Cependant, si nous souhaitons concevoir des SADM ouverts et génériques, l'utilisation des formulaires figés rendent cette tâche impossible en raison de leur structure. Un formulaire conçu pour un SADM appliqué au suivi des patients souffrants d'un diabète ne peut pas être réutilisé pour un SADM appliqué au suivi des insuffisants cardiaques. Les données collectées ne sont pas forcément les mêmes (dans le premier cas, le taux de la glycémie est indispensable alors que ce n'est pas forcément nécessaire dans le deuxième cas).

### III.2.2 Capteurs connectés

Durant les quinze dernières années, le monde des capteurs a évolué. Dans le domaine médical des capteurs ont été conçus pour l'acquisition de données physiologiques (activité physique, tension artérielle, pouls, etc.). Le but étant de mesurer précisément les données physiologiques tout en ayant la possibilité de les transmettre à des systèmes médicaux distants via des standards de communication sans fil (ex. Bluetooth, infra-rouge). Dans la littérature, plusieurs SADM dédiés à la télésurveillance médicale ont utilisé des capteurs médicaux comme outils d'acquisition de données médicales. Parmi eux nous citons (Benyahia et al., 2012, 2013) où les auteurs proposent une plateforme de télésurveillance médicale pour le suivi des insuffisants cardiaques à domicile. La plateforme proposée collecte des données physiologiques sur les patients (poids, pouls, tension artérielle, température, etc.) à l'aide des capteurs médicaux (balance, tensiomètre, thermomètre, etc.) et les transmet par la suite à un serveur médical pour les analyser à l'aide d'un moteur d'inférence. (Franco et al., 2010b) et (Franco et al., 2010a) ont travaillé sur une étude de télésurveillance médicale à domicile pour des personnes atteintes d'Alzheimer. Leur SADM permet la détection des dérives des rythmes nycthémeraux à partir des données de localisation. Ces données sont capturées par des capteurs infrarouges passifs afin d'alerter le corps médical quand une anomalie est détectée. (Shahriyar et al., 2010) proposent un SADM pour la télésurveillance médicale appelé Intelligent Mobile Health Monitoring System (IMHMS). Ce système utilise des capteurs implantables ou portables pour l'acquisition de données physiologiques comme la température corporelle, l'ECG et la saturation en oxygène. D'autres capteurs pour collecter des données sur le comportement du patient comme un microphone et un accéléromètre sont aussi utilisés. Ces données sont ensuite transmises vers un serveur médical. Le serveur traite les données en se basant sur des techniques de fouille de données pour détecter des situations dangereuses et alerte les médecins.

L'utilisation des capteurs comme outil d'acquisition de données a démontré leur grand intérêt dans le domaine de la surveillance médicale. Ces outils permettent de collecter des données sur le patient et son environnement de manière autonome sans l'intervention des cliniciens.

Cependant, l'utilisation des capteurs est coûteuse et demande une maintenance qui nécessite parfois la maîtrise de plusieurs technologies utilisées pour leurs fonctionnements. De plus, la plupart des capteurs permettent seulement de collecter des données quantitatives telles que les mesures physiologiques (poids, pouls, tension artérielle, etc.) ou des données liées à l'environnement du patient (température de la chambre, taux d'humidité, etc.).

Toutefois, les données qualitatives telles que l'hygiène de vie, le niveau de douleur, la nutrition, etc. ne peuvent pas être collectées par des capteurs. Il est donc nécessaire d'utiliser d'autres outils d'acquisition de données qui permettent d'interroger directement le patient.

### III.2.3 Questionnaires informatisés

Les questionnaires, sont considérés comme étant des types de formulaires conçus pour recueillir des données. Ils ont été inventés par la société statistique de Londres en 1838. Bien qu'ils soient souvent conçus pour une analyse statistique des réponses, les questionnaires sont très utilisés dans le domaine médical pour recueillir des informations sur les patients.

Les questionnaires informatisés peuvent être considérés comme étant la version électronique des questionnaires qui permettent la collecte de données à l'aide d'un ordinateur. Ils présentent plusieurs avantages par rapport aux questionnaires papiers en raison de leur efficacité. La saisie des données est automatisée et va créer ainsi une plus grande précision des informations saisies (Bouamrane et al., 2008b) et une rapidité d'exécution (K. K. Boyer and Jackson., 2002).

Ils peuvent ainsi être intégrés comme des outils d'acquisition de données dans des systèmes médicaux pour la collecte des données médicales du patient tout au long des épisodes de soin.

#### III.2.3.1 Composition de questionnaire

Un questionnaire est généralement constitué d'une succession de questions qui peuvent être indépendantes ou liées les unes aux autres. La typologie des questions est étudiée suivant différents points de vue, principalement la formulation de la question, le format de la réponse ou encore le nombre de choix possibles. Les types de questions les plus communs sont des :

- Question à choix simple : une question fermée avec un choix à sélectionner parmi plusieurs réponses prédéfinies.
- Question à choix multiple : une question fermée avec plusieurs choix à sélectionner parmi plusieurs réponses prédéfinies.
- Question conditionnée : une question conditionnée par une ou plusieurs réponse(s) précédente(s).
- Question ouverte : une question qui laisse au répondant la liberté de choisir ses

propres mots. Ce type de questions est peu approprié pour analyser les réponses et fait souvent appel aux techniques de traitement de langage naturel.

- Question échelle : ce type de question est très répandu dans le domaine de la psychanalyse, la personne interrogée exprime son degré d'accord ou de désaccord vis-à-vis d'une affirmation.

### III.2.3.2 Utilisation des questionnaires comme outil d'acquisition pour les SADM

Dans le domaine de la e-santé, plusieurs travaux de recherches ont porté sur l'informatisation des questionnaires afin de les intégrer dans les SADM en tant qu'outil d'acquisition de données.

Les travaux de Bouamrane et al. (Bouamrane et al., 2008b,a) proposent un modèle générique pour la collecte de données sensibles au contexte. Ce modèle est basé sur l'utilisation des ontologies permettant de modéliser les questionnaires. L'ontologie questionnaire est composée de quatre types de propriétés :

- **Propriétés structurales** : utilisées pour décrire la structure du questionnaire.
- **Propriétés de composition** : utilisées pour combiner les questions ensemble.
- **Propriétés de type** : utilisées pour connaître la nature des questions et les réponses.
- **Propriétés adaptative** : utilisées pour déterminer le comportement dynamique du questionnaire.

Le modèle proposé dans (Bouamrane et al., 2008c) est implémenté en tant que module de collecte de données médicales des patients afin d'évaluer les risques préopératoires (voir figure III.1).

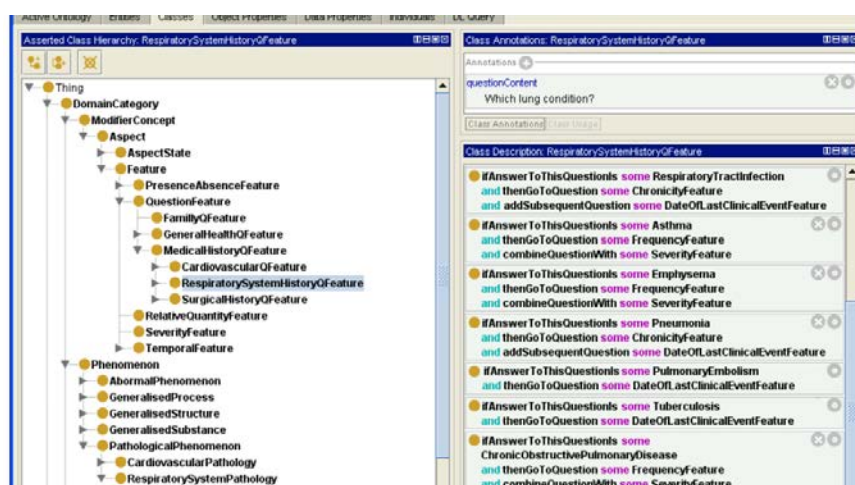


FIGURE III.1 – L'ontologie de questionnaire proposée dans (Bouamrane et al., 2008b) (vue à travers Protégé – OWL)

L'ontologie de questionnaire proposée (voir figure III.1) est organisée en trois

sous-questionnaires : informations personnelles, informations médicales génériques et antécédents médicaux. Le sous-questionnaire d'antécédents médicaux contient : l'histoire cardiovasculaire, les antécédents de troubles respiratoires, les antécédents chirurgicaux, etc. La partition sémantique du questionnaire a été faite conformément à l'ontologie GALEN (Rector and Rogers, 2006).

Sherimon et al. (Sherimon, 2013) (Sherimon et al., 2014b) (Sherimon et al., 2014a) proposent un questionnaire à base d'ontologie en se basant sur les travaux de Bouamrane et al. (Bouamrane et al., 2008b). Cette ontologie est intégrée dans un SADM qui permet de prédire les risques d'hypertension en collectant les antécédents médicaux des patients à l'aide des questionnaires informatisés.

Farooq et al (Farooq et al., 2011) propose un SADM pour l'évaluation des risques cardiovasculaires en interrogeant le patient sur ses douleurs thoraciques. Pour la construction de l'outil permettant l'acquisition de données, les auteurs se sont basés sur les travaux de Bouamrane et al. (Bouamrane et al., 2008b).

Alipour (Alipour, 2014) propose une approche pour la construction d'un système de collecte de données à base d'ontologie et de moteur d'inférence Pellet. Cette approche consiste à modéliser les questionnaires à l'aide d'une ontologie, alors que le moteur Pellet permet la sélection des questions.

Bien que les travaux présentés précédemment ont permis de construire des outils d'acquisition de données à base d'ontologies, la création des questionnaires reste compliquée à réaliser vu l'absence d'un vrai outil de configuration. Elle nécessite en effet la maîtrise de l'éditeur Protégé, ce qui n'est pas à la portée de tout le monde, notamment des cliniciens. De plus, les modèles de questionnaires et les connaissances de domaines sont créés ensemble dans une même ontologie ce qui rend l'adaptation des questionnaires moins flexible et plus compliquée à maintenir.

### III.3 Conclusion

Nous avons présenté dans ce chapitre un état de l'art sur l'acquisition de données médicales dans lequel nous avons souligné, dans un premier temps, son importance pour l'élaboration des raisonnements cliniques. Ensuite, nous nous sommes intéressés à l'informatisation de l'acquisition, par le biais des outils tels que : les formulaires électroniques, les capteurs connectés et les questionnaires informatisés. Ces outils peuvent être utilisés pour introduire des données sur le patient et de son environnement dans les systèmes d'aide à la décision médicale.

Dans un second temps, nous avons cité pour chacun des outils quelques travaux relatifs où nous avons souligné les avantages et les inconvénients en termes de facilité d'implémentation et de maintenabilité. Les problématiques liées à ces outils et aux travaux de recherches proposés dans la littérature ont été également présentés. Il nous a apparu

primordial de proposer un outil d'acquisition de données médicales adaptable, évolutif et facile à maintenir pour rendre le SADM le plus générique possible.



# IV

## RAISONNEMENT À BASE DE CONNAISSANCES

Dans ce chapitre nous présentons un état de l'art sur des travaux de recherche menés dans le raisonnement à base de connaissances

### Sommaire

---

<b>IV.1 Introduction</b> . . . . .	<b>37</b>
<b>IV.2 Raisonnement à partir de cas (RàPC)</b> . . . . .	<b>38</b>
IV.2.1 Qu'est-ce que le RàPC . . . . .	38
IV.2.2 Notions de base . . . . .	40
IV.2.3 Pondération des attributs . . . . .	45
IV.2.4 SADM fondé sur RàPC . . . . .	46
IV.2.5 Avantages et inconvénients . . . . .	48
<b>IV.3 Raisonnement à partir de règles (RàPR)</b> . . . . .	<b>48</b>
IV.3.1 Qu'est-ce que RàPR . . . . .	48
IV.3.2 Notions de base . . . . .	49
IV.3.3 SADM fondé sur RàPR . . . . .	49
IV.3.4 Avantages et inconvénients . . . . .	52
<b>IV.4 Les systèmes fondés sur un raisonnement mixte</b> . . . . .	<b>52</b>
<b>IV.5 Conclusion</b> . . . . .	<b>53</b>

---

### IV.1 Introduction

Un raisonnement à base de connaissances est un mode de raisonnement qui s'appuie sur des connaissances relatives à un domaine particulier afin de résoudre des problèmes qui se posent dans ce domaine. Ces connaissances sont issues des connaissances des experts du domaine et elles sont stockées dans une base de connaissances sous un format exploitable (F. Le Ber and Napoli, 2006).

Parmi les types de raisonnements existants, nous nous intéressons plus particulièrement à deux types de raisonnements : le raisonnement à partir de cas (RàPC) et le raisonnement à partir de règles (RàPR) qui font l'objet de notre étude.



## IV.2 Raisonnement à partir de cas (RàPC)

### IV.2.1 Qu'est-ce que le RàPC

Le raisonnement à partir de cas (RàPC) est un paradigme de raisonnement consistant à interpréter un nouveau problème en le comparant avec des problèmes similaires rencontrés et résolus par le passé (Alsun, 2012). Ce type de raisonnement simule le raisonnement humain pour la résolution des problèmes de la vie quotidienne. Une personne fait naturellement appel à son expérience et se remémore les situations semblables déjà rencontrées. Elle les compare ensuite à la situation courante afin de construire une nouvelle solution qui, à son tour, s'ajoutera à son expérience (Watson, 1998).

En médecine, le RàPC est un moyen très répandu qui apporte aux médecins une aide précieuse dans leurs démarches cliniques, ceci et d'autant plus intéressant quand les médecins sont confrontés à un cas complexe. Ce moyen tend à résoudre un problème auquel est confronté le médecin au cours de son analyse en s'appuyant sur les solutions des cas similaires déjà traités (Alsun, 2012).

Ce type de raisonnement dépend de la base de cas comportant des cas déjà résolus. La base de cas doit comporter un nombre important de cas bien structurés et indexés afin de permettre une recherche simple, facile et rapide quand le besoin se présente (Alsun, 2012).

Le RàPC passe par 4 étapes. La figure IV.1 illustre le cycle de raisonnement tel qu'il a été proposé dans (Aamodt and Plaza, 1994).

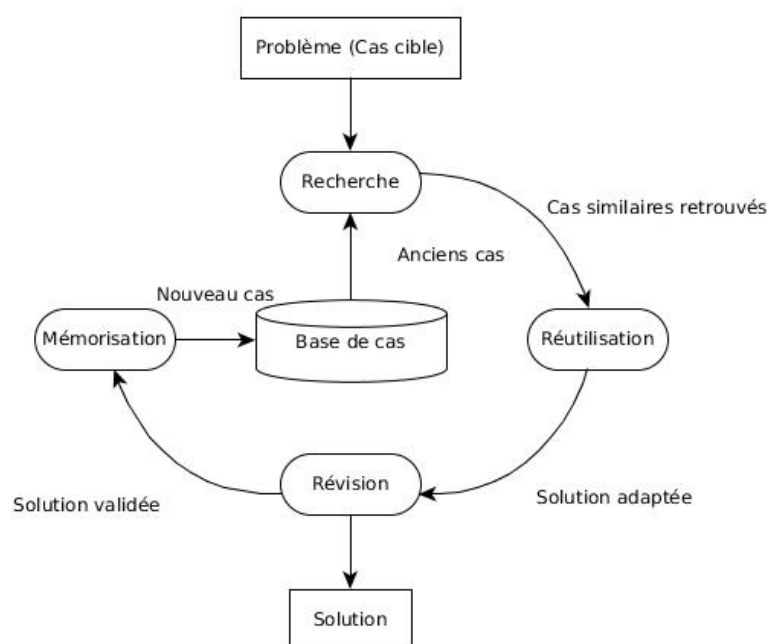


FIGURE IV.1 – Processus de raisonnement à base de cas (Aamodt and Plaza, 1994)

1. **Recherche** : cette phase permet d'effectuer une recherche dans la base de cas afin d'extraire les cas les plus similaires à un cas cible. La recherche des cas similaires est une phase majeure dans un RàPC où la mise en correspondance entre deux cas joue un rôle vital. Cette étape est essentielle, surtout dans les applications médicales car le manque de cas similaires peut réduire l'efficacité de la décision.

Dans (Begum, 2009) l'auteur estime que la fiabilité et la précision des systèmes de diagnostic à base de cas dépendent fortement de la manière de stocker les cas, de rechercher les cas pertinents et de classer les résultats. En effet, les cas trouvés sont classés en fonction de leur degré de similarité et le système propose les mieux classés comme une solution à la situation courante.

2. **Réutilisation et révision** : les cas extraits sont envoyés à l'étape de réutilisation (voir figure IV.1) où la solution d'un cas passé est souvent adaptée afin de trouver une solution au cas cible. Un utilisateur peut adapter les solutions en combinant par exemple deux ou plusieurs solutions de la liste des cas proposés pour développer une solution au problème de cas cible (Begum, 2009). Cette adaptation doit se faire par des experts du domaine, c'est à dire des cliniciens du domaine médical considéré. Le clinicien / expert détermine si la solution proposée est plausible au problème posé. Il pourrait éventuellement modifier la solution avant qu'elle ne soit approuvée. Ensuite, le cas adapté est envoyé à l'étape de révision où l'exactitude de la solution adaptée est vérifiée manuellement puis présentée comme une solution confirmée pour le nouveau problème (cas cible).

Par ailleurs, dans les systèmes d'aide à la décision médicale, la possibilité d'adaptation n'est généralement pas permise car souvent ces systèmes proposent les meilleurs cas au clinicien comme une suggestion de solutions au problème posé sans lui donner la possibilité de les adapter (Watson, 1997).

3. **Mémorisation** : c'est la dernière phase du cycle de RàPC. Dans cette phase, le nouveau cas résolu est ajouté à la base dans l'espoir que sa solution puisse à son tour aider le même utilisateur ou tout autre utilisateur dans la résolution des nouveaux cas similaires.

Il est à noter que la mémorisation du nouveau cas résolu peut être faite manuellement selon la décision de l'expert de domaine (Begum, 2009).

Dans (Alsun, 2012) l'auteur montre que l'efficacité d'un tel mode de raisonnement en termes de rapidité, de fiabilité de résultats et de facilité de calcul dépend de plusieurs facteurs : le mode de représentation de connaissances (représentation des cas), les méthodes à utiliser pour calculer le degré de similarité entre deux cas et l'exploitation des cas similaires.

## IV.2.2 Notions de base

### IV.2.2.1 Notion de cas

Un cas représente une pièce maîtresse de la connaissance et joue un rôle primordial dans le processus de le RàPC. La représentation des cas prend une place importante dans la réalisation d'un système à base de cas. Elle peut être structurée de différentes manières et, généralement, constituée en deux parties : (i) une description du problème et (ii) une solution. Pour une bonne évaluation du cas cible (le nouveau cas à traiter), les cas peuvent ainsi contenir une partie résultats (voir figure IV.2).

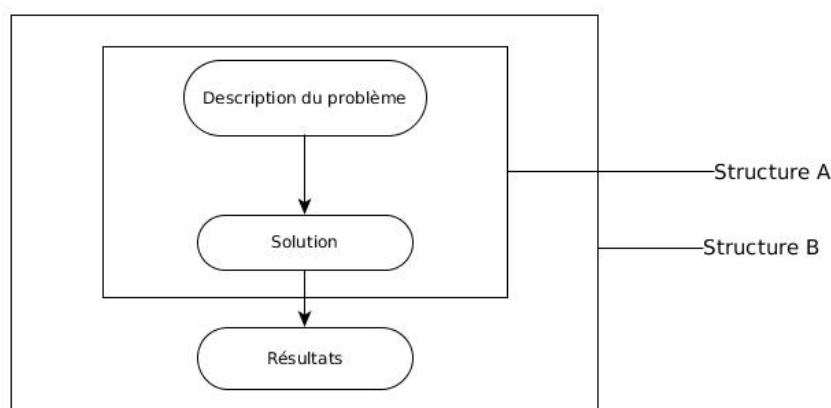


FIGURE IV.2 – Différentes structures d'un cas (Begum, 2009)

### IV.2.2.2 Notion de similarité

La notion de similarité joue un rôle primordial dans le RàPC étant donné que la phase de recherche en dépend fortement pour trouver les cas les plus similaires au cas cible (Alsun, 2012).

Plusieurs mesures de similarité ont été proposées dans la littérature, elles se distinguent selon les types de données auxquelles elles s'appliquent. Les deux grands types de données simples sont les données symboliques et les données numériques. A cela s'ajoutent d'autres types de données (séries de donnée temporelles, évaluations, durée, date, etc.).

**Généralités** Les mesures de similarité sont très largement utilisées dans divers domaines. Par conséquent, leur terminologie varie considérablement (ressemblance, proximité, etc.) et l'auteur d'une mesure particulière n'est pas forcément unique. Malgré la confusion qui règne dans la littérature, certaines conditions sur les mesures de similarité sont plus fréquemment mentionnées que d'autres.

Soit  $X$  l'espace des données ou univers, une mesure de similarité est usuellement définie de

la manière suivante :

**Définition** Une mesure de similarité  $Sim$  est une fonction  $X \times X \rightarrow R^+$  qui satisfait les propriétés suivantes :

- Positivité :  $\forall x, y \in X, Sim(x, y) \geq 0$
- Symétrie :  $\forall x, y \in X, Sim(x, y) = Sim(y, x)$
- Maximalité :  $\forall x, y \in X, Sim(x, x) \geq Sim(x, y)$

D'autres propriétés peuvent être requises comme la normalisation qui impose que les valeurs appartiennent à l'intervalle  $[0, 1]$ .

Des mesures non normalisées peuvent subir une transformation de normalisation pour obtenir leur version normalisée à l'aide de la fonction de transformation  $f(Sim) = (Sim + 1)/2$ . Dans la suite, nous considérons le cadre des mesures normalisées. Les similarités peuvent être exprimées sous forme de distances (dissimilarité) et vice-versa. En appliquant la formule  $Dist = 1 - Sim$  (respectivement  $Sim = 1 - Dist$ ). Pour cela,  $Sim$  et  $Dist \in [0, 1]$ . La notion de similarité comporte trois aspects : la similarité locale, la similarité globale et la pondération des attributs.

**La similarité locale** Cette similarité est calculée pour chaque attribut, elle dépend de la nature de l'attribut en question :

- **Similarité symbolique.** Cette similarité est stricte, elle vaut donc 0 ou 1. Elle vaut 1 si les attributs sont égaux, sinon 0. Elle peut être écrite de cette forme :

$$sim(a, b) = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{sinon} \end{cases} \quad (IV.1)$$

- **Similarité numérique.** Cette similarité est utilisée pour calculer le degré de similarité entre deux valeurs numériques. Elle peut être écrite de cette forme :

$$Sim(a, b) = 1 - \frac{|a - b|}{D} \quad (IV.2)$$

Où  $D$  est la valeur maximale que  $a$  et  $b$  peuvent prendre.

- **Similarité entre des séries de données.** Plusieurs méthodes proposent de mesurer le degré de similarité entre des séries de données.
- **Distance euclidienne.** Pour deux vecteurs  $\vec{u}$  et  $\vec{v}$  de taille  $N$  avec  $\vec{u} = \{x_1, x_2, x_3, \dots, x_n\}$  et  $\vec{v} = \{y_1, y_2, y_3, \dots, y_n\}$ . La distance  $ED$  est exprimée comme suit :

$$d_{ED}(\vec{u}, \vec{v}) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (IV.3)$$

Comme la distance euclidienne n'a pas de borne supérieure et que sa valeur augmente avec le nombre de descripteurs  $N$ , il est conseillé de calculer la distance

ED normalisée :

$$d_{ED}(\vec{u}, \vec{v}) = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (x_i - y_i)^2} \quad (\text{IV.4})$$

Cette distance ignore les dépendances temporelles entre les différentes séries de données. Ceci ne permet donc pas de comparer la variation de série de données temporelle (Najmeddine et al., 2012)

- **Distance de Mahalanobis.** C'est une mesure de distance proposée par Prasanta Chandra Mahalanobis en 1936. Elle est basée sur la corrélation entre des variables par lesquelles différents modèles peuvent être identifiés et analysés. C'est une manière utile de déterminer la similarité entre une série de données. Elle diffère de la distance euclidienne par le fait qu'elle prend en compte la variance et la corrélation de la série de données.

Soit deux vecteurs aléatoires  $\vec{u}$  et  $\vec{v}$  de même taille  $N$  avec  $\vec{u} = \{x_1, x_2, x_3, \dots, x_n\}$  et  $\vec{v} = \{y_1, y_2, y_3, \dots, y_n\}$ . La distance est exprimée comme suit :

$$d_{\vec{u}, \vec{v}}(\vec{u}, \vec{v}) = \sqrt{(\vec{u} - \vec{v})^T \vartheta^{-1} (\vec{u} - \vec{v})} \quad (\text{IV.5})$$

Où  $\vartheta$  est une matrice de covariance. Si la matrice de covariance est la matrice identité, cette distance est simplement la distance euclidienne. Si la matrice de covariance est diagonale, on obtient la distance euclidienne normalisée :

$$d_M(\vec{u}, \vec{v}) = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{\vartheta_i^2}} \quad (\text{IV.6})$$

Où  $\vartheta_i$  est l'écart type de  $x_i$  sur la série de données.

- **Similarité sémantique.**

**Similarité à distance taxonomique.** (Rada et al., 1989) ont suggéré de calculer la similarité sémantique en se basant sur les liens taxonomiques « is-a » dans un réseau sémantique. Ce calcul est basé sur les liens hiérarchiques de spécialisation/généralisation. Pour évaluer la similarité sémantique dans une taxonomie, il est nécessaire de calculer la distance entre les concepts par le chemin le plus court. Les auteurs soulignent que cette proposition est valable pour tous les liens de type hiérarchique (is-a, kind-of, part-of, etc.) mais doit être adaptée pour d'autres types de liens.

La similarité proposée est calculée comme suit :

$$sim(C_1, C_2) = depth(C_1) + depth(C_2) \quad (\text{IV.7})$$

Où  $depth(C_i)$  est le nombre d'arcs entre  $(C_i)$  et la racine en passant par le subsumant commun le plus spécifique (LCS).

(Wu and Palmer, 1994) ont proposé une mesure de similarité entre concepts quel-

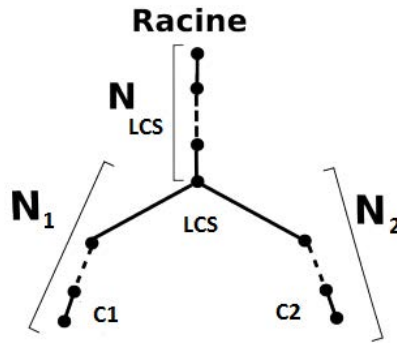


FIGURE IV.3 – Deux concepts et leur subsumant commun le plus spécifique dans une taxinomie.

conques  $C_1$  et  $C_2$  dans une taxinomie par rapport à la distance qui les sépare et par rapport à leur subsumant commun le plus spécifique. La similarité entre  $C_1$  et  $C_2$  est :

$$sim(C_1, C_2) = \frac{2 \times N_{lcs}}{N_1 + N_2 + 2 \times N_{lcs}} \quad (IV.8)$$

Ou plus formellement :

$$sim(C_1, C_2) = \frac{2 \times depth(LCS)}{depth(C_1) + depth(C_2)} \quad (IV.9)$$

Où  $LCS$  est le subsumant commun le plus spécifique,  $depth(LCS)$  est la longueur du chemin entre  $LCS$  et la racine de la hiérarchie,  $depth(C_i)$  est le nombre d'arcs entre  $(C_i)$  et la racine en passant par  $LCS$ .

(Leacock and Chodorow, 1998) se basent également sur la mesure de Rada, mais au lieu de normaliser par la profondeur relative de la taxinomie par rapport aux concepts, ils choisissent une normalisation par rapport à la profondeur totale de la taxinomie  $D$  et normalisent avec un logarithme.

$$sim(C_1, C_2) = -\log\left(\frac{depth(C_1) + depth(C_2)}{2 \times D}\right) \quad (IV.10)$$

**Similarité à base de contenu informationnel** Les mesures de similarité proposées (Resnik, 1995) et (Jiang and Conrath, 1997) sont fondées sur la notion de contenu informationnel. Elle utilise l'ontologie et le corpus d'une manière conjointe. Le contenu informationnel d'un concept traduit la pertinence d'un concept dans le corpus en tenant compte de la fréquence de son apparition ainsi que de la fréquence d'apparition des concepts qu'il subsume. On dit qu'un concept  $C_1$  subsume un concept  $C_2$  si  $C_2$  est plus spécifique que  $C_1$ . Le contenu informationnel peut se calculer par la formule suivante :

$$CI(c) = -\log(P(c)) \quad (IV.11)$$

Où  $P(c)$  est la probabilité de retrouver une instance du concept  $c$ . Ces probabilités sont calculées par :  $frequency(c)/N$  où  $N$  est le nombre total de concepts.

— **Similarité entre deux ensembles**

**L'indice de Jaccard.** Proposé en 1901 par Paul Jaccard, cet indice permet de mesurer la similarité entre deux ensembles d'objets. Il consiste à diviser le nombre d'objets commun par le nombre d'objets distincts dans les deux ensembles. Autrement dit c'est le rapport entre le cardinal (la taille) de l'intersection des ensembles considérés et le cardinal de l'union des ensembles. Soit deux ensembles  $A$  et  $B$ , l'indice est défini comme suit :

$$J(A, B) = \frac{A \cap B}{A \cup B} \quad (IV.12)$$

Cependant cet indice ne prend pas en considération les relations sémantiques qui peuvent être les éléments d'ensemble.

**Best Match Average (BMA).** Proposé par (Wu et al., 2013) pour calculer la similarité entre deux ensembles de gènes en se basant sur l'ontologie GO<sup>1</sup>. Cette dernière est utilisée pour déterminer le score de similarité entre les gènes.

Soit  $A$  et  $B$  deux ensembles de données qui n'ont pas forcément la même cardinalité. La similarité BMA est définie comme suit :

$$sim_{BMA}(A, B) = \frac{\sum_{k=1}^m \max_{1 \leq j \leq n} sim_{sem}(a_k, b_j) + \sum_{j=1}^n \max_{1 \leq k \leq m} sim_{sem}(b_j, a_k)}{m + n} \quad (IV.13)$$

Où  $m = |A|$  et  $n = |B|$ .  $sim(a_i, b_j)$  est la similarité sémantique entre  $a_k$  et  $b_j$  appartenant à la même classe sémantique.

### IV.2.2.3 La similarité globale

En calculant la similarité pour chaque pair d'attributs dont les valeurs sont connues, on obtient un ensemble de similarités locales. Un système de raisonnement à partir de cas doit calculer la similarité globale pour chaque cas candidat. Ceci permet d'ordonner les cas mémorisés selon leur degré de ressemblance avec le cas à traiter.

On ne peut pas avoir une mesure de similarité appropriée pour tous les domaines. Les systèmes de raisonnement à partir de cas utilisent en général différentes mesures de simi-

1. <http://geneontology.org>

larités globales pour assurer l'efficacité de la tâche de recherche des cas similaires.

Voici, dans ce qui suit, quelques méthodes qui permettent de mesurer le degré de similarité entre deux cas  $A$  et  $B$ , à  $N$  attributs chacun.

— **Manhattan.**

$$sim(A, B) = \frac{1}{P} \sum_{i=1}^P sim_i(a_i, b_i) \quad (IV.14)$$

— **Manhattan pondérée.**

$$sim(A, B) = \frac{1}{P} \sum_{i=1}^P w_i \times sim_i(a_i, b_i) \quad (IV.15)$$

— **Euclidienne.**

$$sim(A, B) = \left( \frac{1}{P} \sum_{i=1}^P sim_i(a_i, b_i)^2 \right)^{1/2} \quad (IV.16)$$

— **Minkowski.**

$$sim(A, B) = \left( \frac{1}{P} \sum_{i=1}^P sim_i(a_i, b_i)^r \right)^{1/r} \quad (IV.17)$$

— **Minkowski pondérée.**

$$sim(A, B) = \left( \frac{1}{P} \sum_{i=1}^P w_i \times sim_i(a_i, b_i)^r \right)^{1/r} \quad (IV.18)$$

Où  $P$  est le nombre d'attributs et  $w_i$  est le poids du  $i^{\text{ème}}$  attribut avec  $\sum_{i=1}^P w_i = 1$ .  $sim_i$  est la similarité locale calculée pour l'attribut  $i$ .

Dans le cas où  $\sum_{i=1}^P w_i \neq 1$ . La similarité pondérée doit être divisée par la somme des poids des attributs, donc la mesure de similarité Manhattan devient :

$$sim(A, B) = \frac{\sum_{i=1}^P w_i \times sim_i(a_i, b_i)}{\sum_{i=1}^P w_i} \quad (IV.19)$$

### IV.2.3 Pondération des attributs

Consiste à attribuer à chaque attribut un coefficient numérique qui permet de pondérer son influence sur la mesure de similarité. La manière avec laquelle les attributs sont pondérés est cruciale pour que la mesure de similarité soit pertinente.

On distingue deux familles de méthodes de pondération automatiques :

— **La pondération par apprentissage :** les poids des attributs qui ont eu tendance



à prédire les bonnes solutions sont incrémentés, tandis que les poids des attributs en désaccords sont diminués.

- **La pondération par analyse statistique** : la pondération est évaluée à partir des probabilités d'apparition des attributs dans les différentes classes.

## IV.2.4 SADM fondé sur RàPC

### IV.2.4.1 Architecture

Un système fondé sur un RàPC se compose essentiellement de deux parties. La figure IV.4 illustre le modèle global d'un système à base de cas.

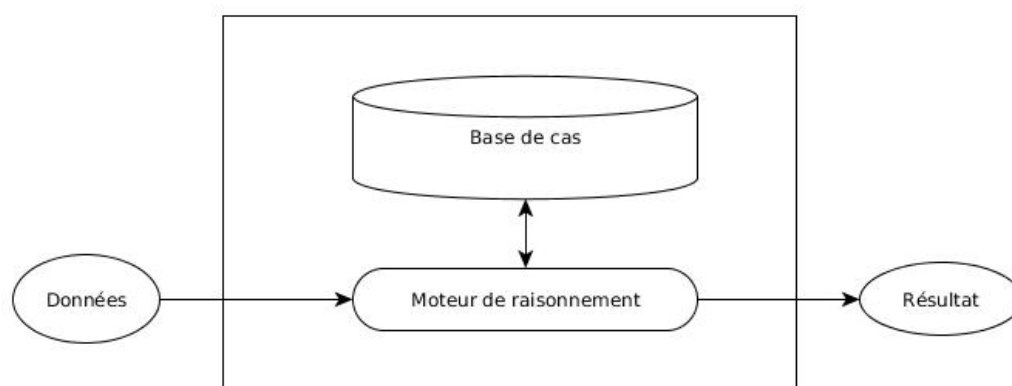


FIGURE IV.4 – Modèle global d'un système à base de cas

- **La base de cas** : contient les cas déjà résolus dans le passé, structurés et indexés. Cette base peut s'enrichir au fur et à mesure avec des nouveaux cas résolus.
- **Le Moteur de raisonnement** : c'est la partie intelligente du système. Le moteur de raisonnement permet à partir des données introduites et en se basant sur des connaissances de fournir des résultats.

### IV.2.4.2 Exemples de SADM fondés sur un RàPC

FM-Ultranet (Z. El Balaa, 2003; Z. El Balaa and Uziel, 2003) est un système d'aide au diagnostic à base de cas appliqué à la gynécologie pour la détection précoce des malformations fœtales. Ce système permet d'aider les échographistes dans leur diagnostic en exploitant leur propre expérience à travers des cas cliniques similaires déjà diagnostiqués. Les cas sont représentés par des objets et de manière hiérarchique. La hiérarchie est organisée en plusieurs concepts et sous-concepts. Chaque concept est représenté par un ou plusieurs attributs. Les attributs se composent des connaissances relatives à la structure anatomique du fœtus, complétées par des données cliniques sur le patient et des connaissances appro-

priées du domaine. La mesure de la similarité entre les cas implique des mesures médicales spécifiques développées par des gynécologues.

SFDA (Chakraborty et al., 2011) est un système d'aide au diagnostic à base de cas appliqué au diagnostic de la grippe porcine. Ce système permet de prédire si le patient est touché par la grippe porcine ou non en fonction des symptômes présents chez le patient et son âge. Dans la base de cas, chaque cas est représenté par un ensemble d'attributs (âge, température, nausée, etc.) et chaque attribut est associé à un coefficient de pondération (poids). Les cas les plus similaires sont retrouvés à l'aide d'une fonction de similarité qui permet de déterminer le degré de similarité entre le cas cible et chacun des cas de la base en se basant sur l'algorithme K-NN (Nearest Neighbour Algorithm).

MedCase (Wang and Tansel, 2013) est une plate-forme open source qui permet aux professionnels de santé de concevoir des SADM distribués fondés sur le RàPC. La représentation des cas est à base d'ontologie ce qui rend la construction et le partage de cas plus facile par les systèmes distribués.

La conception de MedCase a été basée sur des travaux de recherches antérieurs présentés dans (Hsien-Tseng Wang, 2013) où les auteurs proposent la construction d'un système d'aide au diagnostic médical fondée sur RàPC en utilisant le web sémantique comme technologie de représentation des connaissances. Les données d'entrées au système sont : (i) les connaissances médicales simplifiées (ontologie de maladies et symptômes pour le pré-diagnostic) et (ii) les symptômes et les signes présents chez le patient consulté. Le mécanisme de raisonnement passe par deux étapes. La première étape consiste en l'application d'un simple algorithme d'appariement pour recueillir les maladies candidates pour le diagnostic. Ensuite un RàPC est utilisé pour affiner la décision du diagnostic. La base de cas est structurée par un modèle ontologique de connaissances, dont le processus de recherche de cas similaires est basé sur le calcul de la similarité sémantique.

Dans (P. Pesl and Georgiou, 2015) les auteurs présentent une architecture d'un SADM à base de cas pour le calcul du bolus d'insuline nécessaire pour des personnes atteintes d'un diabète de type 1. Ce système fournit des recommandations personnalisées en distinguant différents scénarios de diabète et en ajustant automatiquement ses paramètres au fil du temps.

Le système proposé est composé de deux principales composantes :

- une composante mobile dédiée aux patients permettant la saisie manuelle du taux de glucose et des variables qui influent sur les niveaux de glucose dans le sang. Elle fournit en temps réel des conseils sur le bolus d'insuline nécessaire.
- une composante dédiée aux cliniciens pour la révision clinique et la supervision des adaptations automatiques des paramètres du calculateur de bolus.

Le système se base sur un raisonnement à base de cas pour la personnalisation de bolus d'insuline en utilisant des cas déjà traités par le passé et pour l'apprentissage automatique à partir des nouveaux événements.

### IV.2.5 Avantages et inconvénients

Le RàPC a des avantages majeurs bien connus. Dans (Amélie Cordier, 2009) les auteurs soulignent que les systèmes fondés sur un RàPC ne construisent pas des solutions à partir de rien, ils s'appuient sur des cas déjà traités par le passé. Par conséquent, ces systèmes n'ont pas besoin de disposer des connaissances complètes sur le domaine pour commencer à raisonner. Ils commencent avec très peu de connaissances et ils sont capables d'apprendre et de s'améliorer avec chaque nouveau cas traité. De plus, ce type de système prend en compte l'expérience clinique des médecins.

Pour toutes ces raisons, le RàPC est considéré comme un puissant paradigme de raisonnement facile à mettre en œuvre.

Cependant, malgré ces avantages, le RàPC souffre de quelques problèmes. Selon Amélie Cordier et al (Amélie Cordier, 2009) dans un RàPC, les cas sont souvent considérés comme une structure figée qui n'est généralement pas conçue pour évoluer. Par ailleurs, si l'on souhaite faire évoluer la représentation d'une expérience, les contraintes de la structure de cas nous limitent. Ce manque de flexibilité de la représentation des connaissances est sans aucun doute une véritable limite du RàPC. En outre, selon (Isabel Estevez, 2006) un des inconvénients majeurs du RàPC réside dans la difficulté d'indexation des cas.

## IV.3 Raisonnement à partir de règles (RàPR)

### IV.3.1 Qu'est-ce que RàPR

Le raisonnement à partir de règles (RàPR) utilise les connaissances des experts sous forme de règles. Ces règles sont définies par des experts du domaine et ensuite combinées dans le but d'imiter les processus de raisonnement de ces experts (Dinevski et al., 2011).

Le raisonnement à partir des règles présente certains avantages : tout d'abord, il est facile à comprendre et à interpréter. De plus il est assez naturel car l'être humain raisonne souvent sous forme de règle. Il présente un autre avantage : c'est sa modulabilité. En effet, il est possible d'ajouter et d'enlever des règles d'une manière simple sans avoir à modifier l'architecture du système ni à la structure de données.

Le RàPR a été utilisé la première fois en 1976 dans le système MYCIN par Shortliffe (Shortliffe, 1976). Il fut le premier système d'aide au diagnostic à base de règles qui permettait de choisir la thérapie antimicrobienne la plus appropriée pour un patient atteint d'une maladie du sang.

## IV.3.2 Notions de base

### IV.3.2.1 Règles

Les règles sont une technique de représentation des connaissances sur un domaine d'application. Une règle est constituée d'une ou plusieurs prémisses (appelées aussi conditions ou antécédents) reliées à une ou plusieurs conclusions (Chniti, 2013). Elle est généralement écrite sous la forme SI - ALORS. Les prémisses sont décrites dans la partie SI de la règle et représentent des faits ou des situations, tandis que les conclusions sont décrites dans la partie ALORS et représentent soit de nouveaux faits soit des actions à appliquer. Selon (Chniti, 2013) les règles peuvent être catégorisées selon les connaissances modélisées ou selon l'impact des conclusions modélisées comme suit :

- **Les règles logiques.** Cette catégorie de règles permet d'inférer de nouveaux faits lorsque les prémisses sont vérifiées.
- **Les règles de production.** Cette catégorie de règles représente les conditions nécessaires pour effectuer une action. Dans ce cas, la règle décrit dans sa prémisse un ensemble de situations initiales qui peuvent se produire et la partie conclusion modifie cet ensemble en exécutant une fonction.
- **Les règles probabilistes.** Cette catégorie de règles attribue un facteur de certitude aux prémisses et aux conclusions d'une règle. Les conclusions inférées à partir d'une prémisse probabiliste ne peuvent être considérées comme "certainement vrai" et les prémisses qui sont "absolument certaines" peuvent acheminer à des conclusions incertaines.
- **Les règles floues.** Cette catégorie de règles représente des ensembles flous en utilisant des termes qui peuvent être interprétés différemment selon le contexte.
- **Les règles métier.** Cette catégorie de règles définit un aspect métier. Une règle métier permet de décrire les critères de décision et les actions à mener pour un processus donné.

### IV.3.2.2 Faits

Les faits sont des données considérées comme vraies et sur lesquelles le moteur de règles se base pour inférer des nouvelles données. En médecine les faits représentent une situation clinique d'un patient. Par exemple : Age=55 ans ; Sexe =Homme et Température=38.2°C.

## IV.3.3 SADM fondé sur RàPR

### IV.3.3.1 Architecture

Les systèmes à base de connaissances fondés sur un RàPR sont généralement constitués de trois parties (a) Une base de règles ; (b) Une base de faits ; (c) Un moteur d'inférence.

La figure IV.5 illustre le modèle global d'un système à base de règle.

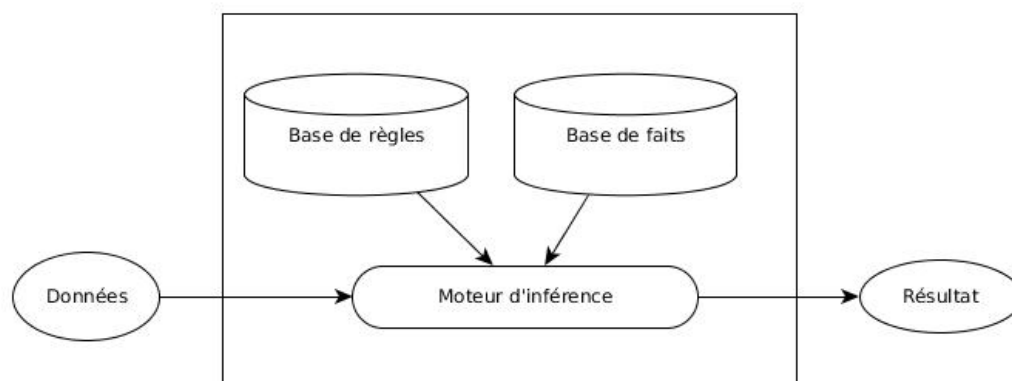


FIGURE IV.5 – Modèle global d'un système à base de règles

- **Base de règles** : elle contient des connaissances sous la forme d'un ensemble de règles de production représentée par des conditions type « SI <liste de faits> ALORS <liste de faits> », et/ou des associations probabilistes entre les données cliniques du patient (Dinevski et al., 2011).

Notons que les faits sont les symptômes observés et les données disponibles sur le patient. L'utilisation de certaines règles permet de déduire de nouveaux faits, qui à leur tour permettent, à travers d'autres règles, de produire de nouveaux faits, et ainsi de suite jusqu'à ce que des faits répondant aux questions posées soient obtenus ou qu'aucun fait nouveau ne puisse plus être produit.

La méthode d'encodage de ces connaissances doit correspondre à la conception du moteur d'inférence utilisé. Par exemple si le moteur d'inférence est basé sur le raisonnement Bayésien, la base de règles doit contenir des règles sous forme de probabilités a priori, conditionnelles et postérieures pour chaque type de connaissance (maladies, symptômes, etc.) et des conclusions (Spooner, 2007).

- **Base de faits** : elle contient un ensemble de connaissances appelées "faits" considérés comme vrais. À partir de ces faits, le moteur d'inférence va appliquer les règles issues de sa base de règles pour déduire de nouveaux faits et ainsi résoudre le problème exposé.
- **Moteur d'inférence** : c'est le processus qui fait coopérer connaissances, faits, et stratégies de résolution de problèmes, afin d'inférer des conclusions. Le moteur d'inférence applique des règles formelles ou informelles à des données (Spooner, 2007). Généralement le raisonnement se fait par règle de production, en combinant et mettant en corrélation les règles avec les données afin de dériver des conclusions.

Généralement un moteur d'inférence fonctionne selon deux mécanismes de raisonnement qui peuvent être combinés :

- Raisonnement en chaînage avant, guidé par les données, au cours duquel les règles

sont utilisées dans le sens « conditions → conclusion ».

- Raisonnement en chaînage arrière, guidé par un but ou un modèle, revenant à utiliser une règle dans le sens « conclusion → conditions ».

Dans les deux cas, le moteur possède un fonctionnement cyclique au cours duquel il sélectionne dans sa base de connaissances les règles les plus appropriées. Il choisit la règle à appliquer, il l'applique et recommence (Haton and Haton, 2016).

### IV.3.3.2 Exemples de SADM fondés sur un RàPR

MYCIN (Shortliffe, 1976) conçu par Shortliffe en 1976, il fut le premier système expert fondé sur un RàPR pour le diagnostic médical, il diagnostiquait des patients touchés par des maladies du sang et leur proposait les thérapies les plus appropriées. Le système était constitué d'une base de règles associées à un coefficient de vraisemblance (poids). Quant au moteur d'inférence, il produisait un chaînage avant simple.

ISABEL (P. Ramnarayan, 2004) est un système d'aide au diagnostic médical développé en 2001 et commercialisé par la société Isabel Healthcare. Ce système se base essentiellement sur l'exploitation des guides de bonnes pratiques (GBPs) qui permettent de suggérer les diagnostics les plus appropriés pour le cas clinique présenté en entrée. La liste de diagnostics retournée par le système est suivie par une liste de médicaments pertinents, de GBPs et des protocoles associés. Ce système dispose d'une base de connaissances couvrant plusieurs pathologies relatives à la médecine interne, la chirurgie, la gynécologie, l'obstétrique, la pédiatrie, la gériatrie, l'oncologie, la toxicologie et le bioterrorisme. Il utilise le traitement automatique du langage naturel (TALN) comme outil de recherche sur les données cliniques du patient (signes, symptômes, tests laboratoire) qui sont exprimées à l'aide d'un texte libre. Il permet ainsi l'extraction de données cliniques depuis des dossiers médicaux électroniques (DMÉs). Le moteur de raisonnement d'ISABEL se base sur l'inférence bayésienne et la théorie de l'information de Shannon (Autonomy, 2009).

PAIRS (Rao, 2000a,b) (Physician assistant Artificial Intelligence System) est un système d'aide au diagnostic médical développé en 2000. Il a été conçu pour aider les médecins à diagnostiquer les cas difficiles en lui fournissant des diagnostics différentiels. La base de connaissances de PAIRS dispose de plus de 35 000 symptômes pour plus 547 maladies de la médecine interne. Elle est construite à partir des guides de bonnes pratiques (GBP). Quant au moteur de raisonnement du système, il est basé sur les méthodes variationnelles et des modèles probabilistes développés par Jaakkola et Jordan dans (Jaakkola and Jordan, 1999). Le système calcule les probabilités pour chacune des maladies à l'aide des règles bayésiennes et donne des diagnostics différentiels, ce qui permet éventuellement au médecin de trouver un diagnostic pertinent.

ATHENA (Goldstein et al., 2004; Lai, 2004) est un système d'aide au diagnostic médical pour l'hypertension développé en 2002 au sein de l'université de Stanford. Fondé sur un RàPR dont les règles sont issues des guides de bonnes pratiques employés au sein

du système. Il propose des recommandations sur le choix approprié de la thérapie. Ce système dispose d'une base de connaissances facilement modifiables qui spécifie les critères d'éligibilité et la stratification des risques chez le patient. Cette base comprend aussi les comorbidités pertinentes et des recommandations des guides de bonnes pratiques. ATHENA utilise l'architecture EON de l'université de Stanford. EON est une architecture à base de connaissances, qui permet de guider les médecins en charge des patients par des protocoles et des directives.

ADVISE est un système d'aide à la prescription pour le traitement des infections bactériennes chez les patients hospitalisés en soins intensifs. Il a été développé en 2002 par (Thursky et al., 2006; Thursky and Mahemoff, 2007). ADVISE est un système à base de règles (Thursky et al., 2006). Ces règles sont construites à partir de ressources documentaires en infectiologie. Elles se basent sur le pathogène, le spectre bactérien et les caractéristiques du patient (âge, poids, sexe, créatinine plasmatique) pour générer des recommandations (Thursky et al., 2006). Le système génère les recommandations en utilisant les résultats des examens microbiologiques obtenus à partir du système d'information hospitalier, ainsi que les données du patient entrées par le médecin (Thursky et al., 2006).

#### **IV.3.4 Avantages et inconvénients**

L'un des principaux avantages de cette approche est qu'elle permet une représentation naturelle des connaissances des experts de domaine sous forme de règles. En 1972 Newell et Simon ont souligné que les règles sont une façon simple, facile à comprendre et naturelle pour modéliser le raisonnement des humains dans la résolution des problèmes. De plus, ce type d'approche sépare la couche connaissance de la couche de raisonnement, et de ce fait le moteur de raisonnement peut être développé dans n'importe quel langage. Un troisième point positif peut être relevé, cette approche permet le traitement des connaissances incomplètes et incertaines.

Malgré une utilisation assez large de cette approche dans les systèmes d'aide à la décision médicale, elle présente quelques inconvénients. Dans (Begum, 2009) Shahina Begum souligne que pour la mise en œuvre d'un tel système le domaine du problème doit être bien compris et constant dans le temps, ce qui n'est pas le cas pour le domaine médical. Ce dernier présente beaucoup de complexité et exploite des connaissances dynamiques qui ne cessent d'évoluer.

### **IV.4 Les systèmes fondés sur un raisonnement mixte**

Afin de pallier les insuffisances de chacune des approches présentées ci-dessus (RàPR et RàPC), plusieurs travaux de recherches ont été menés pour but de combiner les deux modes de raisonnement. Les études menées dans (Berka, 2011; Prentzas and Hatzilygeroudis, 2002) ont démontré que le RàPC et le RàPR peuvent fonctionner conjointement

au sein d'un système d'aide à la décision. Ces deux types de raisonnement peuvent être combinés selon plusieurs modes. Dans (Berka, 2011), Petr Berka distingue trois stratégies principales pour la combinaison du RàPC et le RàPR : La première stratégie consiste en l'application du RàPR puis RàPC. La deuxième stratégie consiste en l'application du RàPC puis RàPR. Quant à la troisième stratégie, elle consiste à l'entrelacement des deux modes de raisonnement. Petr Berka souligne ainsi que l'application du RàPR en premier lieu est appropriée lorsque les règles sont assez efficaces et précises. Cependant, si les règles sont déficientes, l'application du RàPC en premier lieu peut avoir plus de sens.

Dans la littérature, plusieurs travaux de recherches essayent de combiner le RàPC et le RàPR. Ils ont été menés pour la construction des SADM hybrides. Nous citons à titre d'exemple, les travaux de Cabrera et Edye dans (Cabrera and Edye, 2010). Les auteurs présentent une approche hybride appliquée au diagnostic de la méningite bactérienne aiguë. L'approche proposée combine les deux modes de raisonnement où le RàPR est utilisé en premier lieu puis le RàPC s'il est nécessaire. Si le RàPR peut inférer avec succès une conclusion (sans doute), l'application du RàPC n'est pas effectuée. Sinon le RàPC est utilisé pour trouver une solution au cas cible.

## IV.5 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art sur le raisonnement à base de connaissances. Nous avons montré que ce type de raisonnement se base sur les connaissances des experts du domaine et qu'il emploie un mécanisme de raisonnement plus ou moins proche du raisonnement humain. Nous avons cité deux types de raisonnements : (i) le raisonnement à partir de règles (RàPR) et (ii) le raisonnement à partir de cas (RàPC). Pour chaque type de raisonnement, nous avons cité des avantages et des inconvénients et quelques travaux de recherches liés à ces modes de raisonnement. À ce niveau-là, nous avons constaté que ces deux modes de raisonnement peuvent être combinés selon plusieurs stratégies afin de pallier leurs inconvénients.

Nous avons choisi de combiner l'utilisation de ces deux modes de raisonnements dans le même système permet l'exploitation des deux sources de connaissances dont l'une complète l'autre à savoir : les guides de bonnes pratiques et l'expérience clinique.





# II

## CONTRIBUTION



# CONTRIBUTION À LA REPRÉSENTATION DES CONNAISSANCES MÉDICALES

Dans ce chapitre, nous présentons notre contribution concernant la représentation des connaissances au sein du système d'aide à la décision médicale.

## Sommaire

---

<b>V.1 Introduction</b> . . . . .	<b>57</b>
<b>V.2 Représentation de connaissance de domaine</b> . . . . .	<b>58</b>
V.2.1 Représentation des connaissances médicales . . . . .	58
V.2.2 Représentation des connaissances liées au domaine d'intervention du SADM . . . . .	60
<b>V.3 Représentation du profil du patient</b> . . . . .	<b>62</b>
<b>V.4 Conclusion</b> . . . . .	<b>65</b>

---

## V.1 Introduction

La construction d'un système d'aide à la décision médicale (SADM) à base de connaissances implique la modélisation des connaissances dans le domaine considéré. Le problème qui se pose est de trouver les représentations formelles qui permettront non seulement le stockage de ces connaissances mais aussi l'utilisation correcte de celles-ci.

Dans le chapitre II, nous avons présenté un état de l'art sur les techniques de représentation des connaissances médicales. Nous avons présenté ainsi quelques ressources terminologiques utilisées dans la construction des SADM à base de connaissances. Nous présentons dans ce chapitre notre contribution concernant la représentation des connaissances dans les SADM. Nous proposons une représentation générique des connaissances afin de rendre notre système plus ouvert pour accepter des nouvelles connaissances indépendamment de son mode d'intervention. Pour ce faire, nous avons choisi de représenter les connaissances à l'aide d'ontologies. Celles-ci permettent d'apporter une sémantique formelle à notre système et facilitent le partage et la réutilisation des

connaissances.

Pour la construction des ontologies, nous nous sommes basés sur une méthodologie de construction développée dans un précédent travail au sein de notre équipe de recherche (Benyahia, 2015).

## V.2 Représentation de connaissance de domaine

### V.2.1 Représentation des connaissances médicales

Pour la représentation des connaissances médicales, il existe plusieurs ressources dont quelques-unes sont décrites dans le chapitre II. Après l'étude de ces différentes ressources, nous avons choisi d'utiliser des classifications de l'OMS (Organisation Mondiale de la Santé) à savoir :

- La classification CIM-10 (Classification Internationale des Maladies) traduction de la ICD (International Classification of Diseases) pour la représentation des antécédents médicaux.
- La classification ATC (Anatomique Thérapeutique Chimique) traduction de ATC (Anatomical Therapeutic Chemical) pour la représentation des traitements (les médicaments).

Ces classifications répondent parfaitement à nos besoins pour la représentation des antécédents et des traitements médicaux. Elles facilitent leur implémentation et proposent une structure indexée et hiérarchisée. Ceci est très utile pour le calcul des similarités sémantiques dans le raisonnement à base de cas.

#### V.2.1.1 Ontologie CIM-10

La CIM-10 est constituée de vingt-deux chapitres et chaque chapitre est composé de plusieurs Blocs. Les pathologies sont classées du chapitre I jusqu'au chapitre XVII. Tandis que les interventions hospitalières et chirurgicales sont classées dans le chapitre XXI, bloc Z80-Z99.

Nous avons choisi de représenter la classification sous format ontologique dont les classes représentent uniquement les deux premiers niveaux de la CIM-10, à savoir, les chapitres et les blocs.

Le reste de la hiérarchie est représenté par des instances. Ceci est justifié par le fait que le niveau 3 est utilisé pour représenter les antécédents médicaux dans le profil du patient, par exemple, E10 : Diabète sucré insulino-dépendant. Dans cet exemple, E10 est de niveau 3 et peut être utilisé pour dire que le patient a un diabète de type 1 (insulino-dépendant).

La construction des classes a été faite avec un algorithme automatique proposé dans (Be-

nyahia, 2015). Cet algorithme reçoit en entrée la CIM-10 sous format XML et génère en sortie un fichier OWL contenant les classes en utilisant les index des éléments de la CIM-10. Le fichier OWL de sortie est ensuite introduit dans l'outil Protégé pour vérifier l'absence d'erreurs.

Dans l'ontologie créée à partir de la classification CIM-10, nous avons utilisé comme attribut l'annotation RDFS « label ». Cet attribut représente le nom de l'élément dans la CIM-10. Par exemple, l'élément E10 a un attribue « label » ' Diabète sucré insulino-dépendant '. Comme il a été présenté dans la partie précédente, la hiérarchie des classes représente les deux premiers niveaux, Chapitres et Blocs.

Pour représenter la hiérarchie au sein d'un bloc de la CIM-10, une relation de Spécialisation a été créée. Cela permet de dire qu'une instance A de la CIM-10 est une spécification d'une instance B, et de même l'instance B est une généralisation de l'instance A.

Ci-dessous un exemple qui présente la hiérarchie au sein d'un bloc de la CIM-10 :

- Bloc E10-E14 Diabète sucré
  - E10. Diabète sucré insulino-dépendant
    - E10.0 avec coma
    - E10.1 avec acidocétose
    - E10.2 avec complication rénales
    - E10.9 sans complication

Dans l'exemple, E10.2 (Diabète sucré insulino-dépendant avec complication rénales) a une relation spécialisation avec E10 (Diabète sucré insulino-dépendant).

### V.2.1.2 Ontologie ATC

Dans le système ATC, les médicaments sont divisés en plusieurs groupes selon l'organe ou le système sur lequel ils agissent et selon leurs propriétés chimiques, pharmacologiques et thérapeutiques. La classification ATC est basée sur une architecture de 5 niveaux, le premier niveau est composé de 14 groupes principaux selon l'organe ou le système traité (voir chapitre II).

Nous avons choisi de représenter la classification sous format ontologique dont la définition des classes reprend les quatre premiers niveaux de la hiérarchie de l'ATC. Tandis que le cinquième niveau de la hiérarchie est représenté par des instances. Ceci est justifié par le fait que seul le niveau 5 représente les substances chimiques des médicaments.

Comme pour l'ontologie de domaine des antécédents, la construction des classes a été faite avec un algorithme automatique proposé dans (Benyahia, 2015). Cet algorithme reçoit en entrée la ATC sous format XML et génère en sortie un fichier OWL contenant les classes à partir de l'index des éléments de l'ATC.

Comme pour l'ontologie CIM-10, nous avons utilisé comme attribut l'annotation RDFS

« label ». Cet attribut représente le nom de l'élément dans la ATC. Par exemple l'élément A10VA02 a un attribut « label » 'Metformin'.

De la même manière que pour les classes, les instances ont été créées à l'aide d'un algorithme automatique. L'algorithme utilisé pour les classes a été étendu pour générer également les instances et fournir en sortie un fichier OWL. Par la suite, le fichier OWL de sortie a été introduit dans l'outil Protégé.

## V.2.2 Représentation des connaissances liées au domaine d'intervention du SADM

Bien que les ontologies CIM-10 et ATC permettent la représentation des antécédents médicaux, celles-ci ne permettent pas de représenter les concepts spécifiques liés au domaine d'intervention du SADM.

Afin de rendre le système ouvert et générique, nous avons conçu une ontologie de domaine appelée Data type Ontology (DTOnto). L'ontologie DTOnto est une terminologie composée d'un ensemble de termes définis par des experts médicaux. Chaque terme de la terminologie désigne un concept de domaine représentant une connaissance médicale qui peut être, par exemple, un acte médical ou une constante physiologique.

DTOnto permet de représenter des connaissances sur le domaine spécifique de l'intervention du système, par exemple le domaine de la chirurgie digestive pour une réhabilitation améliorée. L'ontologie DTOnto devrait représenter tous les concepts qui sont liés à ce domaine tel que le type de l'opération, les complications, la douleur, la perfusion, etc.

Afin que les concepts (termes) soient typés en fonction de la nature qu'ils peuvent prendre, l'ontologie DTOnto a été structurée avec plusieurs concepts qui représentent chacun un type de données. Ensuite, chaque terme défini est associé au concept approprié qui représente sa nature. Ceci est très utile pour le raisonnement à base de cas où les mesures de similarité dépendent de la nature des types de données.

La figure V.1 illustre une représentation simplifiée de l'ontologie DTOnto.

Comme le montre la figure V.1 l'ontologie DTOnto est composée de plusieurs concepts (classes) qui servent à décrire les types de données de domaine de manière générique :

- **DataType** : est une classe générique qui permet de représenter n'importe quel type de données indépendamment de sa nature. Un type de données est caractérisé simplement par un nom et un poids déterminant son degré d'importance.
- **QualitativeType** : est une sous-classe de la classe DataType. Elle permet de représenter les types de données qualitatives, par exemple le type d'opération.
- **QuantitativeType** : est une sous-classe de la classe DataType. Elle permet de représenter les types de données quantitatives, par exemple le niveau de douleur.
- **DateType** : est une sous-classe de la classe QuantitativeType. Elle permet de représenter les types de données date, par exemple la date de l'opération.
- **DurationType** : est une sous-classe de la classe QuantitativeType. Elle permet

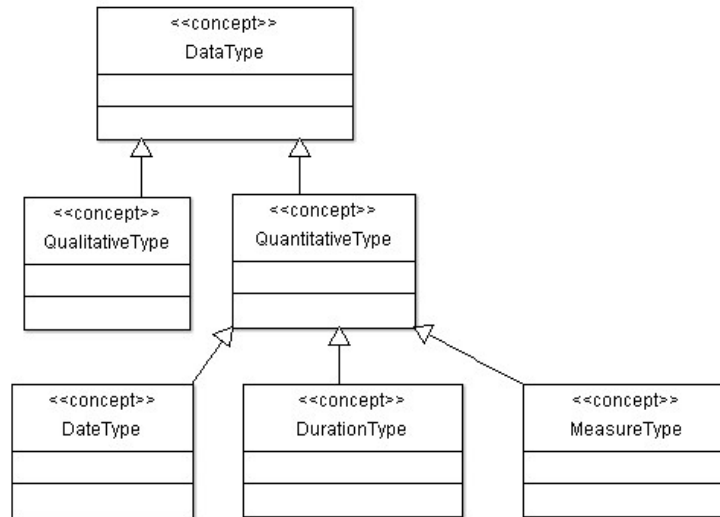


FIGURE V.1 – Représentation simplifiée de l’ontologie DTOnto

- de représenter les types de données durées, par exemple la durée d’hospitalisation.
- **MeasureType** : est une sous-classe de la classe QuantitativeType. Elle permet de représenter les types de données mesures, par exemple les données physiologiques. Ce type de données est un peu particulier car les mesures sont caractérisées par un type de représentation graphique et par une unité de mesure.

La figure V.2 illustre l’utilisation de l’ontologie DTOnto pour la représentation des concepts du domaine chirurgical.

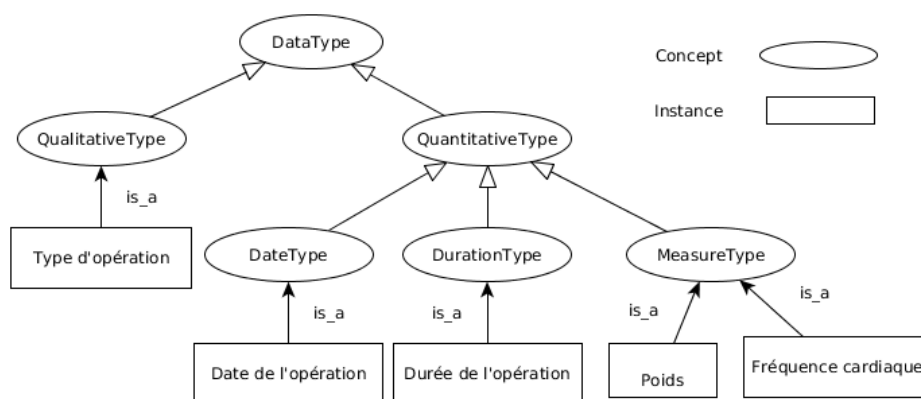


FIGURE V.2 – Extrait de l’ontologie DTOnto dans le domaine de la chirurgie

L’exemple illustré par la figure V.2 montre la représentation des instances (termes) dans l’ontologie en fonction de leurs types. Les instances sont définies en collaboration avec les experts médicaux.



Les intérêts de l'utilisation de l'ontologie DTOnto sont multiples :

- Le typage des termes, représentés dans DTOnto, permet de déterminer la fonction de similarité appropriée dans le raisonnement à base de cas où les mesures de similarité dépendent de la nature des types de données. Par exemple la donnée '2 heures et 30 minutes' est une instance du concept 'durée de l'opération' dont le type de données est une durée.
- DTOnto permet de guider l'outil d'acquisition de données dont les termes représentés dans l'ontologie donnent une signification aux données collectées et, par conséquent, les rendent interprétables et exploitables par le moteur de raisonnement.
- DTOnto offre une représentation évolutive du profil du patient où certains de ses concepts sont liés aux concepts de domaine définis dans DTOnto. En effet, pour enrichir le profil du patient par de nouveaux concepts, il suffit de les rajouter dans l'ontologie de domaine sans avoir à modifier la structure de données du profil.

### V.3 Représentation du profil du patient

Nous proposons une représentation sémantique du profil du patient à l'aide d'une nouvelle ontologie appelée Patient Profile Ontology (PPO). Cette représentation est basée sur l'exploitation des ontologies de domaines décrites précédemment (CIM-10, ATC et DTOnto). Cette représentation est générique, elle ne dépend ni du domaine médical ni du mode d'intervention du système. L'ontologie que nous proposons permet la représentation de toutes les données qui sont en relation avec le patient. Parmi elles, on trouve les données administratives renseignées à l'enregistrement du patient dans le système, les antécédents médicaux du patient et les données collectées au cours du suivi médical du patient.

La figure V.3 illustre une représentation simplifiée de l'ontologie Patient Profile Ontology (PPO).

PPO est composée de plusieurs concepts définis en collaboration avec les experts médicaux. Ces concepts permettent de modéliser tous les éléments constituant un profil générique du patient. Elle est composée du concept **PatientProfile** qui représente le profil du patient. Celui-ci est constitué de deux concepts représentant chacun un type de profil, à savoir, le profil civil et le profil médical.

- **CivilProfile** : ce concept représente les données administratives, c'est-à-dire, des données non médicales. Ces données sont principalement utilisées pour la constitution du dossier administratif du patient et elles sont constituées des données de contact et des données personnelles.
- **ContactData** : ce concept représente les données de contact du patient, ces données comprennent essentiellement l'adresse du patient, ses numéros de téléphone ainsi que les données de ces contacts familiaux et médicaux en cas de

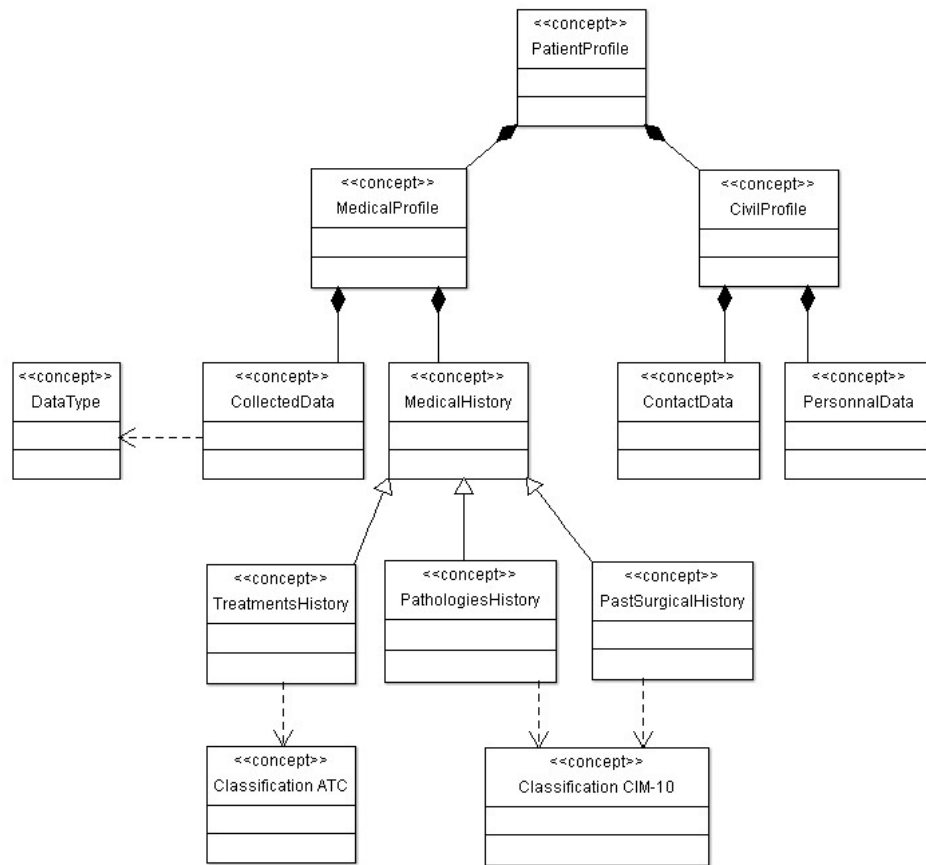


FIGURE V.3 – Représentation simplifiée de l'ontologie Patient Profile Ontology (PPO)

problème.

- **PersonalData** : ce concept représente les données personnelles du patient. Ces données comprennent essentiellement le nom et le prénom du patient ainsi que sa civilité, sa profession, sa date de naissance et son numéro de sécurité sociale.
- **MedicalProfile** : ce concept représente les données médicales du patient. Il est constitué de deux concepts comme suit :
  - **CollectedData** : ce concept représente les données collectées dynamiquement à l'aide de l'outil d'acquisition de données. Ces données sont reliées sémantiquement avec les types de données définis dans l'ontologie de domaine DTonto.
  - **MedicalHistory** : ce concept représente les antécédents médicaux du patient, il est constitué de trois concepts :
    - **TreatmentsHistory** : ce concept représente les traitements prescrits pour le patient. Cette classe a comme attributs le nom du médicament, la dose, la date de début et la date de fin. Cette classe fait référence à l'ontologie de

classification ATC.

- **PathologiesHistory** : ce concept représente l'historiques de maladies du patient. Comme attributs, nous avons la date de début, date de fin et le nom de la maladie. Ce concept fait référence à l'ontologie de classification CIM-10 dont les instances sont liées du chapitre I jusqu'au chapitre XVII de la classification.
- **PastSurgicalHistory** : ce concept représente les interventions chirurgicales que le patient a subi dans le passé. Comme attributs nous avons la date et le nom de l'opération. Ce concept fait référence à l'ontologie de classification CIM-10 dont les instances sont liées uniquement au chapitre XXI, bloc Z80-Z99 de la classification.

Comme le montre la figure V.4 la représentation du profil que nous proposons est flexible et évolutive. En effet, elle peut s'enrichir par des nouveaux concepts issus des ontologies de domaine (DTONto, CIM-10 et ATC) sans avoir à redéfinir toute la structure du profil.

Pour faire évoluer la représentation du profil avec des nouveaux types de données liés aux domaines d'intervention du SADM, il suffit simplement de définir les concepts associés dans l'ontologie de domaine DTONto.

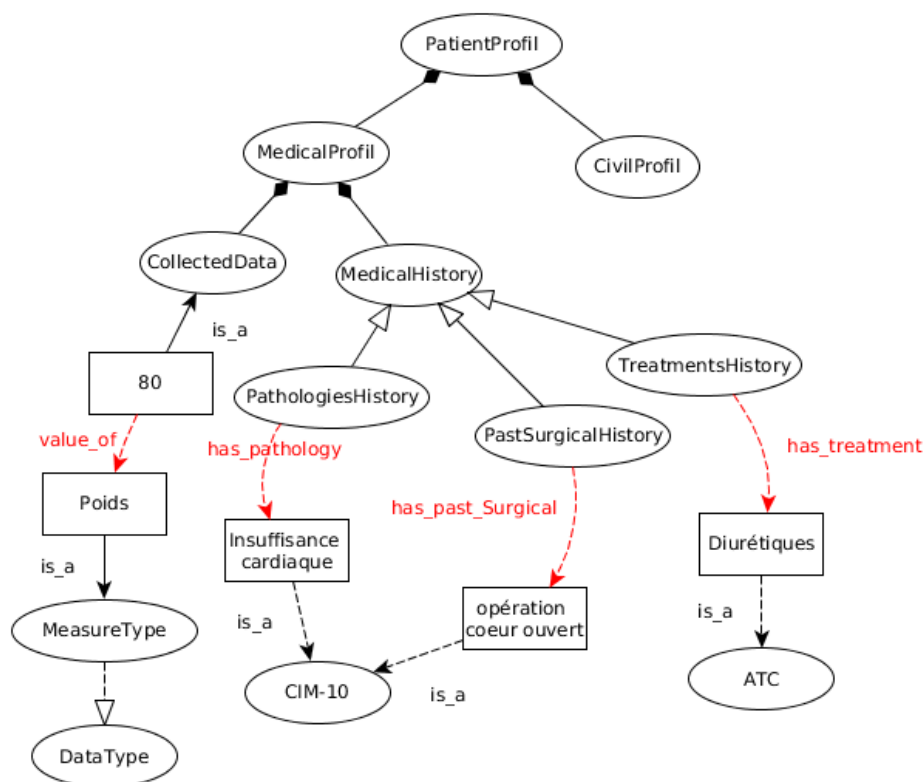


FIGURE V.4 – Exemple d'instanciation du profil du patient

Le profil du patient ainsi représenté permet de rendre le système générique et évolutif

avec une maintenance plus accessible. Ceci est utile étant donné que dans un système dit centré-patient, tel que nous proposons où le profil du patient est un élément central.

## V.4 Conclusion

Dans ce chapitre nous avons présenté notre contribution concernant la représentation des connaissances pour les SADM à base de connaissances et particulièrement la représentation du profil du patient. Pour la représentation des connaissances de domaine, nous avons construit trois ontologies, à savoir, ATC et CIM-10 à partir des classifications existantes pour la représentation des antécédents médicaux et l'ontologie DTOnto pour la représentation des connaissances spécifiques de domaine d'intervention du SADM.

Nous avons ainsi décrit l'ontologie PPO pour la représentation générique du profil du patient. Cette ontologie permet de représenter l'état de santé du patient et son évolution de manière générique, indépendamment de la nature du SADM. Nous avons défini des liens sémantiques entre cette ontologie et les ontologies de domaine (CIM-10, ATC et DTOnto). Ces liens permettent d'avoir une base ontologique liée, où les concepts de l'ontologie PPO vont faire référence à des concepts des ontologies de domaine. Ceci, permet de mettre à jour l'ontologie individuellement sans pour autant modifier les autres ontologies.

L'approche de représentation des connaissances proposée permet donc la construction d'un SADM générique et ouvert qui ne dépend ni du domaine ni de son type d'intervention. L'utilisation d'une telle ontologie de domaine facilite l'acquisition des connaissances et la maintenabilité du système.



# VI

## CONTRIBUTION À L'ACQUISITION DE DONNÉES MÉDICALES

Dans le troisième chapitre de la partie de l'état de l'art, nous avons montré que dans le domaine médical, la collecte des données du patient constitue la première étape de la démarche clinique. Cette dernière permet d'organiser le travail du corps médical et les données collectées permettent de se situer par rapport aux problèmes du patient et de poser le jugement clinique. Nous avons souligné, au même titre que (Phaneuf, 2012), que ces données collectées permettent d'assurer la surveillance clinique tout au long d'un épisode de soins, de contribuer aux décisions médicales et d'évaluer les soins reçus. Principalement l'interrogatoire médical est le mécanisme qui permet cette collecte de données. Il est primordial pour la prise en charge d'un patient. Il permet à lui seul de poser un diagnostic ou tout au moins d'orienter le clinicien dans son processus de raisonnement clinique (Mervoyer, 2009). Nous visons à orienter d'une manière judicieuse et pragmatique le clinicien dans son interrogatoire, en lui apportant les éléments nécessaires. Ces éléments, sont fournis automatiquement, ou à la demande, et ils sont structurés sous forme de questionnaires adaptatifs et contextuels. Les questionnaires sont adaptés dynamiquement, selon les réponses remontées lors de l'interrogatoire, et contextuellement suivant le contexte du patient.

Pour l'acquisition de données des patients pour les SADM, nous avons proposé un outil de collecte de données générique à base d'ontologie. Cet outil permet la collecte des données de patients en se basant sur un mécanisme de question/réponse.

Contrairement aux outils présentés dans la partie de l'état de l'art, l'outil que nous proposons offre plus de flexibilité et de souplesse dans son intégration au sein d'un SADM et il permet :

- La configuration de questionnaires contextuels et adaptatifs indépendamment du domaine ;
- La structuration hiérarchique des questionnaires à travers la définition de liens sémantiques entre les questions et les conditions pour leur affichage ;
- L'historisation des interrogatoires ;

- Une intégration et une maintenabilité intuitives grâce à la structure générique.

## VI.1 Architecture

Dans cette section, nous exposons l'architecture de l'outil d'acquisition de données à base d'ontologies (voir figure VI.1).

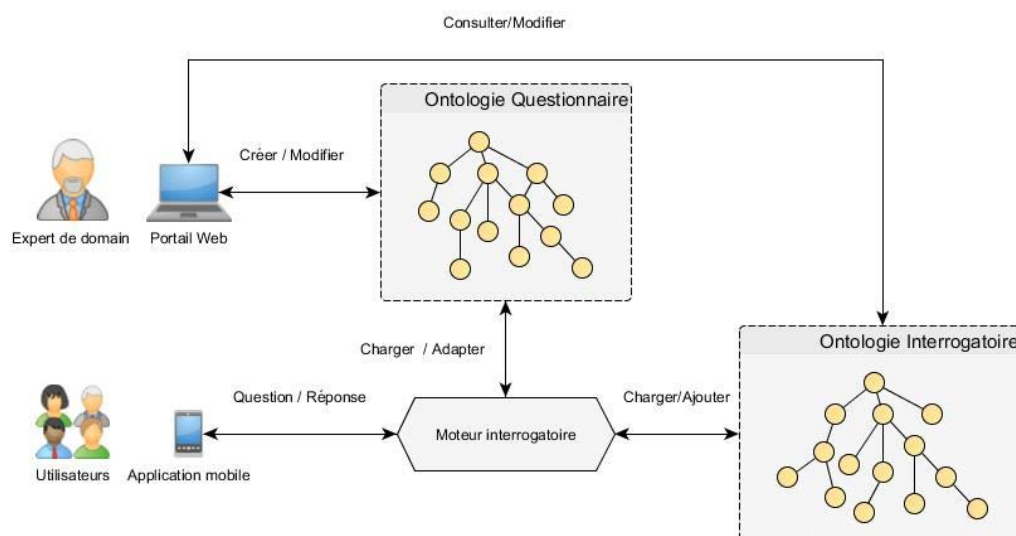


FIGURE VI.1 – Architecture de l'outil d'acquisition de données

Cet outil est constitué essentiellement de trois composantes :

- **Moteur interrogatoire** : permet de dérouler un interrogatoire, de poser les questions suivant le contexte du patient et de les adapter suivant les réponses apportées ;
- **Ontologie questionnaire** : permet de modéliser les modèles de questionnaires à utiliser pour la collecte de données ;
- **Ontologie interrogatoire** : permet d'archiver réponses de questionnaires sous forme d'interrogatoires.

Dans les sections suivantes nous détaillons la structuration de l'outil, proposée sous forme modulaire, ainsi que la représentation des connaissances utilisée pour l'acquisition de données et les modes d'acquisitions proposés.

### VI.1.1 Module de configuration

Ce module est géré par les experts de domaine (les cliniciens). Accessible depuis une interface web (voir figure VI.1), il permet de gérer les questionnaires et l'historique des interrogatoires.

### VI.1.1.1 Gestion de questionnaires

Cette fonctionnalité est dédiée exclusivement aux experts de domaine (cliniciens). Elle permet aux experts de domaine de gérer les questionnaires (i.e. création, modification, consultation et suppression).

La création d'un nouveau questionnaire consiste à définir sa structure en définissant les éléments suivants :

- **Informations générales** : les informations du questionnaire telle que le titre et la description.
- **Contexte** : l'attachement du questionnaire à un contexte médical permet la structuration de la collecte de données. Un contexte médical, appelé aussi statut du patient, peut désigner une étape de prise en charge d'un patient. Par exemple, dans le cadre d'un suivi opératoire des patients, le contexte va permettre d'identifier l'étape dans laquelle se trouve le patient (préopératoire, peropératoire, postopératoire).
- **Protocole de collecte de données** : permet de définir la fréquence et le moment où le questionnaire doit être posé. La fréquence peut désigner simplement le nombre de fois que le questionnaire doit être posé avec ou sans restriction et selon un cycle, précisant les heures et les jours de semaine.
- **Sous-questionnaires** : ces éléments servent principalement à thématiser le questionnaire. Pour un questionnaire sur l'hygiène de vie, par exemple, nous pouvons définir des sous-questionnaires sur les habitudes alimentaires, l'activité physique, le sommeil, le tabagisme, etc.
- **Questions** : la création des questions consiste à définir leurs informations générales, leurs types et éventuellement leurs réponses potentielles. L'outil offre ainsi la possibilité de relier les questions entre elles en définissant des conditions et des relations sémantiques. Par exemple, nous pouvons définir une structure permettant de poser la question « Combien de verres buvez-vous par jour » uniquement lorsque la réponse à la question « Buvez-vous d'alcool ? » a été « Oui ».

Les questionnaires sont évolutifs de manière à permettre aux utilisateurs de modéliser et de faire évoluer la structure des questionnaires à volonté et de manière intuitive.

### VI.1.1.2 Gestion de l'historique des interrogatoires

Cette fonctionnalité offre aux cliniciens la possibilité de suivre l'évolution de l'état de santé de leurs patients.

Toutes les données collectées via les questionnaires sont structurées sous forme d'interrogatoires. Ces derniers sont gérés depuis une application web (portail web) et les cliniciens peuvent consulter les données collectées et les modifier si le besoin se présente. Des indicateurs sur le taux de remplissage des questionnaires permettent aux cliniciens de connaître l'assiduité des patients, la pertinence des données collectées et de valider le respect du



protocole défini.

### VI.1.2 Module de collecte

Ce module assure la collecte des données des patients. Il est accessible depuis le portail Web et depuis l'application mobile (voir figure VI.1). Il peut être utilisé par le corps médical (infirmières et médecins) ou directement par les patients.

Ce module permet principalement de collecter les réponses et de les stocker par la suite sous forme d'interrogatoires.

Le cœur de ce module est le moteur d'interrogatoire qui permet principalement le déroulement des interrogatoires. Ce dernier permet de :

- Charger les questionnaires en fonction du contexte de patient et l'étape dans laquelle il se trouve.
- Adapter les questions suivant les réponses apportées en exploitant les relations sémantiques définies entre les questions lors de la configuration du questionnaire.
- Collecter les réponses et les archiver sous forme d'interrogatoires.

## VI.2 Modélisation des connaissances

L'outil que nous proposons permet de modéliser les questionnaires et les interrogatoires à l'aide de deux ontologies : Questionnaire et Interrogatoire.

### VI.2.1 Ontologie Questionnaire

Cette ontologie permet la modélisation des questionnaires et de l'ensemble des éléments qui les constituent. Pour la modélisation de cette ontologie, nous nous sommes basés sur les travaux présentés dans (Bouamrane et al., 2008a,b,c).

L'ontologie questionnaire, appelé par la suite Questionnaire Ontology (QO), que nous proposons permet une modélisation générique des questionnaires indépendamment de leurs domaines d'application et cela de manière flexible et évolutive. Ces questionnaires peuvent être utilisés dans les différentes étapes de la prise en charge et du suivi médical des patients. QO est composée essentiellement de quatre concepts (voir figure VI.2) :

- **Questionnaire** : ce concept sert à décrire les informations générales d'un questionnaire. Il peut être relié à un contexte (statut de patient) ou rattaché à un protocole de collecte de données.
- **SubQuestionnaire** : ce concept permet de mieux structurer un questionnaire en plusieurs thématiques.
- **Question** : ce concept représente les questions à poser pour la collecte de données. Une question est caractérisée par une étiquette, un ordre, un commentaire et un

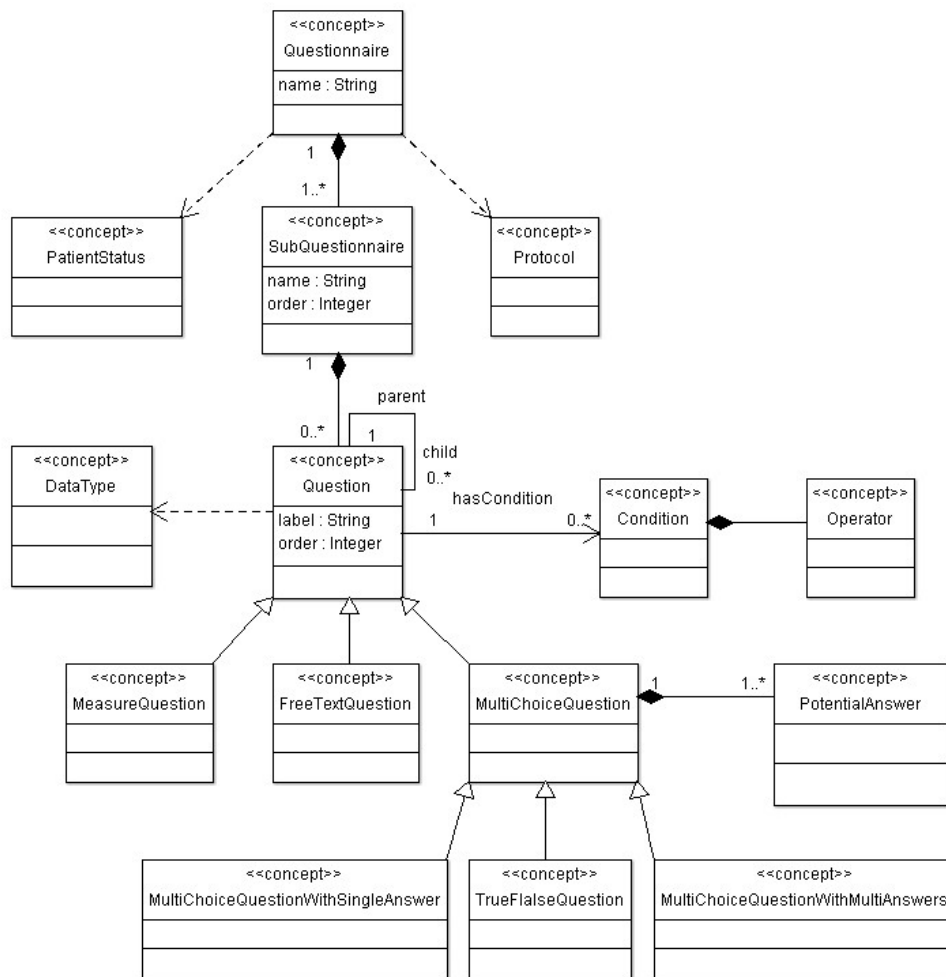


FIGURE VI.2 – Représentation simplifié de l'ontologie questionnaire

type. Plusieurs types de questions sont définis :

- **FreeTextQuestion** : modélise une question ayant comme réponse un texte libre. Ce type de question permet aux utilisateurs d'exprimer leurs réponses dans un texte libre sans aucune contraintes.
- **TrueFalseQuestion** : ce concept modélise une question ayant deux réponses potentielles booléennes 'Oui' et 'Non'.
- **MultiChoiceQuestionWithSingleAnswer** : ce concept modélise une question ayant plusieurs réponses potentielles à choix multiples.
- **MultiChoiceQuestionWithMultiAnswers** : ce concept modélise une question ayant plusieurs réponses potentielles à choix unique.
- **MeasureQuestion** : ce concept modélise une question permettant de collecter une mesure. Une mesure peut être par exemple une données physiologique (poids, température, pouls, etc.). Ce type de question peut comporter une restriction sur le format de données numériques saisies.

- **DateQuestion** : ce concept modélise des questions acceptant une date comme réponse. Ce type de question est utile pour saisir la date d'admission du patient ou la date de son opération dans le cadre d'un suivi opératoire. Ce type de question peut comporter une restriction sur le format de la date saisie.
- **DurationQuestion** : ce concept modélise des questions acceptant une durée comme réponse. Une durée est exprimée en nombre de jours, d'heures, de minutes et de secondes. Ce type de question peut être utilisé par exemple pour saisir la durée de l'opération ou de séjour en soins intensifs.
- **PotentialAnswer** : concept qui modélise les réponses potentielles pour une question. Ces réponses sont définies à la création des questions et elles permettent de contrôler les données saisies. Chaque réponse potentielle est caractérisée par un ordre, une étiquette et une valeur.

Afin de rendre le questionnaire adaptatif en fonction des réponses apportées lors du déroulement de l'interrogatoire, les questions peuvent être reliées entre elles par des liens sémantiques. Ces liens sémantiques (propriétés sémantiques) :

- **hasChild** : cette propriété permet de définir une ou plusieurs questions enfants à une question parent.
- **hasParent** : c'est une propriété inverse à la propriété hasChild. Cette propriété permet de relier une question à une seule question parent.
- **hasSibling** : cette propriété permet de définir un lien sémantique entre toutes questions ayant le même parent.

Pour relier les questions parents à des questions enfants, l'idée consiste à définir des conditions constituées de trois propriétés :

- **ifAnswerToThisQuestionIs** : cette propriété relie la condition à la réponse pour laquelle la question enfant pourrait être posée.
- **thenGoToThisQuestion** : cette propriété relie la condition à la question enfant qui pourrait être posée.
- **hasOperator** : cette propriété relie la condition à l'opérateur. Plusieurs opérateurs sont proposés =, ≠, <, ≤, > et ≥. Pour les données qualitatives, seuls les opérateurs =,≠ sont proposées.

La figure VI.3 illustre la modélisation du comportement adaptatif du questionnaire.

L'exemple de la figure VI.3 peut se traduire comme suit : Si la réponse à la question « fumez-vous ? » est égal à « Oui » Alors poser la question « Combien de cigarettes par jour ? »

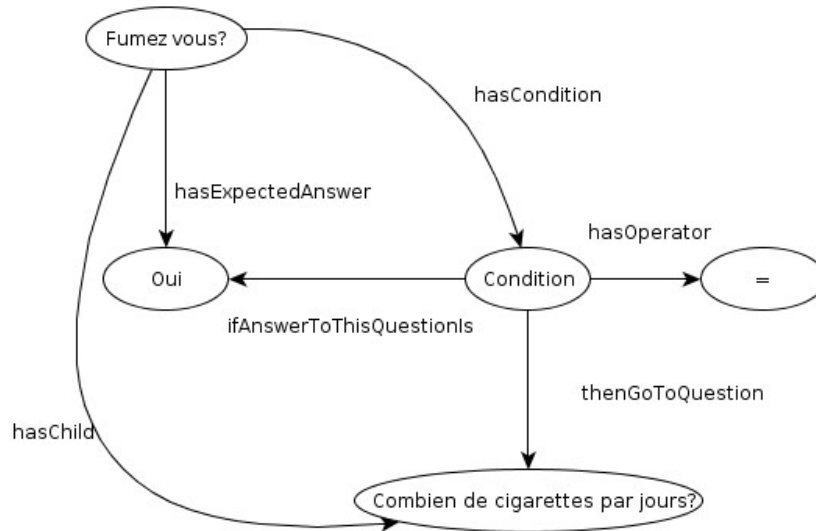


FIGURE VI.3 – Exemple de la modélisation d'une question adaptative

## VI.2.2 Ontologie Interrogatoire

Un interrogatoire peut être vu comme une instance de questionnaire posé à un instant  $T$  à un patient  $P$  par un intervenant  $I$ . Nous avons développé une ontologie appelé Interrogation History Ontology (IHO) (voir figure VI.4). Cette ontologie permet la modélisation d'un interrogatoire médical en décrivant tous les concepts le caractérisant tels que la date de l'interrogatoire, les questions posées, les réponses, la date de saisie des réponses, etc. Elle modélise ainsi toutes les relations sémantiques qui relient un interrogatoire au questionnaire utilisé, au patient interrogé et éventuellement à l'intervenant qui a déroulé l'interrogatoire.

L'ontologie IHO permet ainsi le stockage de tous les interrogatoires médicaux effectués auprès des patients. Ceci est très utiles dans cadre du suivi de patients car il permet aux médecins d'avoir une idée globale sur l'évolution de l'état du patient.

L'ontologie est composée essentiellement de 5 concepts :

- **Interrogation** : concept qui permet de modéliser un interrogatoire médical. Il est caractérisé par une date et heure de début de l'interrogatoire, la date et l'heure de la fin de l'interrogatoire et le taux de remplissage. Ce concept est relié sémantiquement au profil du patient interrogé ainsi qu'à l'intervenant (corps médical) qui a déroulé l'interrogatoire.
- **QuestionnaireHistory** : concept qui permet de modéliser l'historique du questionnaire posé lors de l'interrogatoire. Il fait référence au concept *Questionnaire* de l'ontologie *QO*. Un interrogatoire peut être relié à un ou plusieurs historiques du questionnaire dans le cas où l'interrogatoire a été modifié plusieurs fois.
- **SubQuestionnaireHistory** : concept qui permet de modéliser l'historique du sous-questionnaire posé lors d'interrogatoire. Il fait référence au concept *SubQues-*

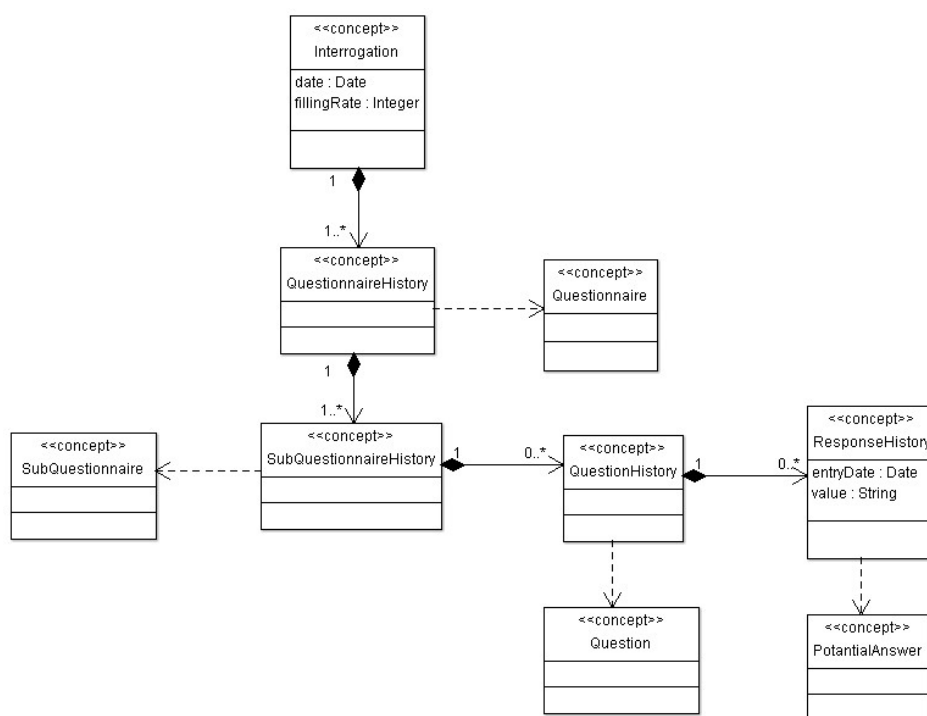


FIGURE VI.4 – Représentation simplifiée de l'ontologie interrogatoire

tionnaire de l'ontologie *QO*.

- **QuestionHistory** : concept qui permet de modéliser l'historique de toutes les questions posées lors d'un interrogatoire. Il fait référence au concept *Question* de l'ontologie *QO*.
- **ResponseHistory** : concept qui modélise les réponses collectées lors d'un interrogatoire. Il fait référence au concept *PotentialAnswer* de l'ontologie *QO* uniquement quand il s'agit d'une réponse à une question ayant des réponses potentielles.

### VI.2.3 Création d'un questionnaire guidée par l'ontologie de domaine DTonto

Nous avons décrit dans le chapitre précédent l'ontologie de domaine DTonto (Data type Ontology), qui offre un vocabulaire contrôlé. Elle est constituée d'une liste de termes définis par les experts de domaine pour la représentation des connaissances spécifiques au domaine d'intervention du SADM. L'ontologie DTonto est utilisée dans l'acquisition de données médicales, afin que celles-ci soient significatives et par conséquent interprétables et exploitables par le moteur de raisonnement.

L'idée consiste à relier sémantiquement chaque question créée à l'instance appropriée dans l'ontologie DTonto (voir figure VI.5).

Ceci permet le contrôle de la création des questionnaires, de façon à s'assurer que les données collectées via les questionnaires soient significatives.

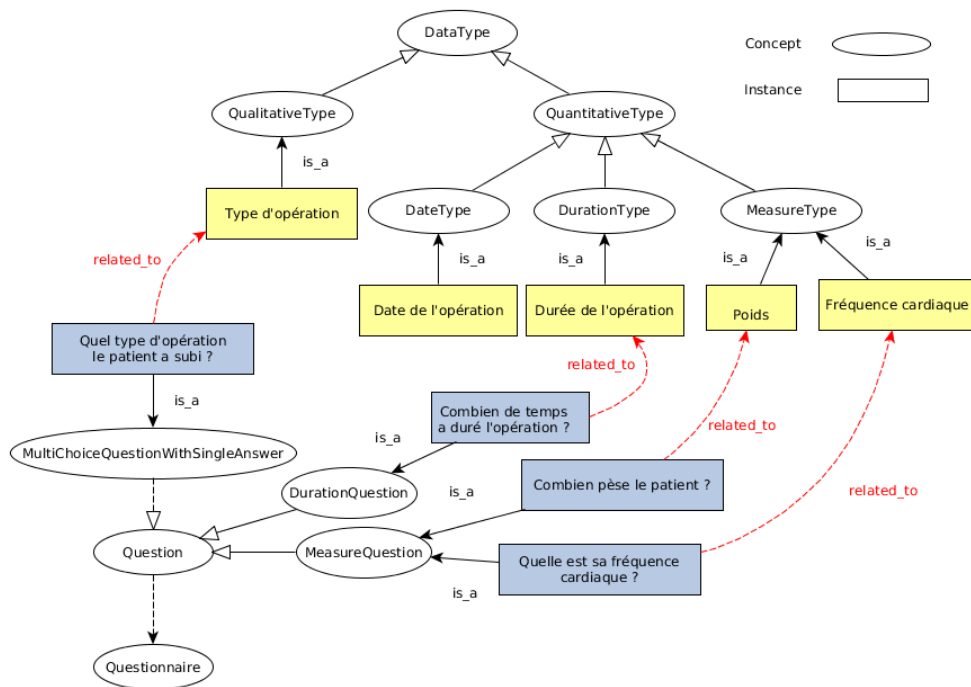


FIGURE VI.5 – Questionnaire guidé par l'ontologie de domaine DTonto

### VI.3 Déroulement de la collecte des données

Pour le déroulement de la collecte des données, nous avons proposé deux modes de collecte à savoir un mode séquentiel et un mode simultané.

#### VI.3.1 Mode d'acquisition séquentiel

Ce mode d'acquisition consiste à poser les questions d'une manière séquentielle. Les questions sont posées les unes après les autres suivant un enchaînement défini par les cliniciens. Ce mode d'acquisition est utilisé généralement par des patients dans le cadre du suivi à domicile.

Nous avons choisi de modéliser la mise en œuvre du questionnaire comme une pile, ou plutôt, comme une pile de piles avec un comportement d'exécution adaptatif comme illustré par la figure VI.6.

Dans la figure VI.6, chaque étape correspond à une itération du système.

La question 1 (Q1) ne possède pas de propriétés d'adaptation et conduit directement à la question 2 (Q2), indépendamment de la réponse (étape 1 → 2). Q2 ne présente pas de propriétés d'adaptation, cependant, la réponse de l'utilisateur ne déclenche pas un appel à d'autres questions et ainsi conduit aussi directement à la question suivante Q3 (étape 2 → 3). À l'étape 3 → 4, contrairement aux questions précédentes, la réponse à la Q3

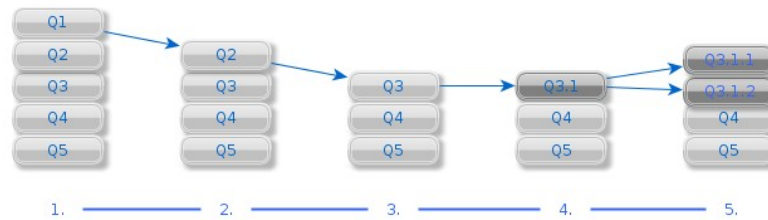


FIGURE VI.6 – Déroulement de l'interrogatoire en mode séquentiel à l'aide d'une pile

déclenche l'appel à une autre question (question enfant). Cette question enfant (Q3.1) est donc placée en tête de la pile (question suivante à apparaître sur l'interface utilisateur). Enfin, la réponse à la Q3.1 déclenche elle aussi un appel à trois questions enfants (Q3.1.1 et Q3.1.2). Ces questions enfants sont placées au sommet de la pile des questions, dans l'ordre défini dans l'ontologie. En fonction des propriétés d'adaptation des questions, le processus d'ajout de questions est répété jusqu'à ce que le moteur d'interrogatoire atteigne le bas de la pile (fin du questionnaire).

### VI.3.1.1 Créer un nouvel interrogatoire

Le lancement d'un nouvel interrogatoire consiste à créer une nouvelle instance du questionnaire. En partant du principe décrit ci-avant. Les questions sont posées les unes après les autres en suivant l'ordre défini par les cliniciens. L'algorithme 1 décrit le déroulement d'un nouvel interrogatoire.

---

#### Algorithme 1 : Nouvel interrogatoire

---

**Données :** **QO** : Questionnaire Ontology ; **IHO** : Interrogation History Ontology (IHO) ;  
**PPO** : Patient Profile Ontology ;

```

1 * StackQ ← loadAllParents( QO ) ; // charger dans la pile les questions parents
2 tant que StackQ ≠ ∅ faire
3   * questioni ← StackQ.pop() ; // charger et supprimer le premier élément de la pile
4   * ask(questioni) ; // poser la question
5   * responsei ← collect() ; // recueillir la réponse
6   * PPO ← responsei ; // sauvegarder la réponse en tant que donnée dans l'ontologie PPO
7   * IHO ← responsei ; // sauvegarder l'historique de la réponse dans l'ontologie IHO
8   * childrenQuestions ← loadChildren(questioni, responsei) ; // charger les questions enfants
   correspondantes à la responsei
9   si childrenQuestions.size() > 0 alors
10  |   * StackQ.push(childrenQuestions) // mettre les questions enfants en tête de la pile

```

---

À la validation de l'interrogatoire, l'instance du questionnaire est complétée avec les réponses de l'utilisateur puis stockée dans l'ontologie interrogatoire (IHO) sous un nouvel interrogatoire. Les données collectées sont ainsi stockées dans l'ontologie PPO qui représente le profil du patient.

### VI.3.1.2 Modifier un interrogatoire

Cette fonctionnalité permet aux utilisateurs de modifier les réponses aux questions déjà saisies. Ceci est important pour la rectification des erreurs de saisies, surtout que les données collectées sont par la suite analysées par le moteur de raisonnement. L'algorithme 2 décrit le déroulement de la modification d'un interrogatoire en mode séquentiel.

---

#### Algorithme 2 : Modification d'un interrogatoire

---

**Données :** **QO** : Questionnaire Ontology ; **IHO** : Interrogation History Ontology ; **PPO** :

Patient Profile Ontology

```

1 * StackQH ← loadHistory( IHO ) ; // charger dans la pile l'historique de questions avec leurs
   réponses
2 tant que StackQH ≠ ∅ faire
3   * questionHistoryi ← StackQH.pop() ; // charger et supprimer le premier élément de la pile
4   * ask(questionHistoryi) ; // poser la questioni et proposer l'ancienne réponse
5   * responsei ← collect() ; // recueillir la nouvelle réponse
6   * PPO ← responsei ; // sauvegarder la nouvelle réponse et modifier la donnée
   correspondante dans le profil
7   * IHO ← responsei ; // sauvegarder l'historique de la réponse dans l'ontologie IHO
8   * childrenQuestions ← loadChildren(questionHistoryi, responsei) // charger les questions
   enfants correspondantes à la reponsei
9   si childrenQuestions.size() > 0 alors
10  | * StackQH.push(childrenQuestions) ; // mettre les questions enfants en tête de la pile

```

---

Partant du principe de la pile décrite ci-avant, les questions et les réponses correspondantes à l'interrogatoire à modifier sont chargées dans la pile. Ensuite elles sont posées les unes après les autres. Pour chaque question posée le moteur d'interrogatoire propose l'ancienne réponse saisie. A la validation de l'interrogatoire, ce dernier est enregistré dans le l'ontologie de l'interrogatoires (IHO) sous une nouvelle version. Quant aux données collectées correspondantes aux réponses modifiées, elles sont modifiées dans le profil du patient.



### VI.3.1.3 Reprendre un interrogatoire

Cette fonctionnalité permet aux utilisateurs de reprendre des interrogatoires non terminés. Dans cette situation, le moteur d'interrogatoire ne propose que les questions auxquelles le patient n'a pas répondu.

---

#### Algorithme 3 : Reprise d'un interrogatoire

---

**Données :** **QO** : Questionnaire History Ontology ; **IHO** : Interrogation Ontology ; **PPO** : Patient Profile Ontology

```

1 * StackQ ← loadUnansweredParents(QO) // charger dans la pile toutes les questions parents
   non répondues
2 tant que Q ≠ ∅ faire
3   |
4   | * questioni ← StackQ.pop() // charger et supprimer le premier élément de la pile
5   | * ask(questioni) // poser la question
6   | * responsei ← collect() // recueillir la réponse
7   | * PPO ← responsei // sauvegarder la réponse comme une donnée dans le profil
8   | * IHO ← responsei // sauvegarder l'historique de la réponse dans l'ontologie IHO
9   | * childrenQuestions ← loadChildren(questioni,responsei) // charger les questions enfants
   | correspondantes à la responsei
10  | si childrenQuestions.size() > 0 alors
11  | | * StackQ.push(childrenQuestions) // mettre les questions enfants en tête de la pile

```

---

### VI.3.2 Mode d'acquisition instantané

Ce mode d'acquisition consiste à poser toutes les questions parents. Les questions sont ordonnées en fonction des critères définis par les cliniciens. Les questions enfants apparaissent au fur à mesure en fonction du déroulement de l'interrogatoire. Ce mode d'acquisition est généralement utilisé par les infirmières pour avoir un accès aux questions plus rapide. Dans ce cas, nous avons choisi de modéliser la mise en œuvre du questionnaire comme une liste, ou plutôt, comme une liste de listes avec un comportement d'exécution adaptatif comme illustré par la figure VI.7.

Contrairement au mode séquentiel, le système ne fonctionne pas par itération. À l'étape 1 le moteur d'interrogatoire charge les 4 questions parents.

La question 1 (Q1) et la question 2 (Q2) ne possèdent pas de propriétés d'adaptation et conduisent directement à la question 3 (Q3) indépendamment de leurs réponses. À l'étape 2 la réponse de l'utilisateur à la Q3 déclenche l'appel à une autre question (question enfant). Cette question enfant (Q3.1) est placée dans la liste juste derrière Q3. À l'étape

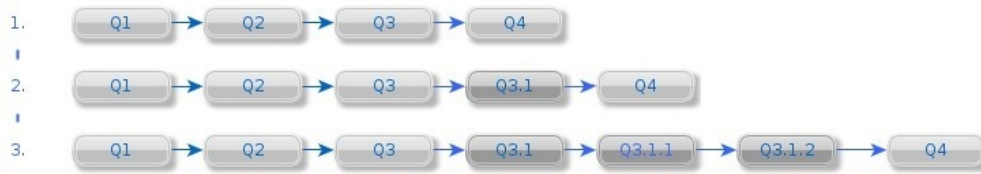


FIGURE VI.7 – Déroulement de l'interrogatoire en mode instantané à l'aide d'une liste

3, la réponse à la Q3.1 déclenche l'appel à deux questions enfants (Q3.1.1 et Q3.1.2). Ces questions enfants sont placées dans la liste juste derrière leur question parent (Q3.1), et dans l'ordre défini dans l'ontologie questionnaire.

En fonction des propriétés d'adaptation des questions, le processus d'ajout des nouvelles questions pourrait être répété jusqu'à ce que le moteur atteigne la fin de la liste (fin du questionnaire).

### VI.3.2.1 Création d'un nouvel d'un interrogatoire

Le lancement d'un nouvel interrogatoire en mode instantané consiste à créer une nouvelle instance du questionnaire. En partant du principe de questionnement décrit ci-avant. Les questions parents sont posées toutes en même temps sous forme d'une liste ordonnées suivant l'ordre défini par les cliniciens. L'algorithme 4 décrit le déroulement d'un nouvel interrogatoire en mode instantané.

---

#### Algorithme 4 : Nouvel interrogatoire

**Données :** **QO** : Questionnaire Ontology ; **IHO** : Interrogation History Ontology ; **PPO** : Patient Profile Ontology

```

1 * ListQ ← loadAllParents( QO ) // charger dans la liste les questions parents
2 * askAll(ListQ) // poser toutes les questions
  // Pour chaque question sélectionnée
3 pour chaque questioni ∈ ListQ faire
4   * responsei ← collect() ; // recueillir la réponse
5   * PPO ← responsei ; // sauvegarder la réponse en tant que donnée dans l'ontologie PPO
6   * IHO ← responsei ; // sauvegarder l'historique de la réponse dans l'ontologie IHO
7   * childrenQuestions ← loadChildren(questioni, responsei) // charger les questions enfants
    correspondantes à la responsei
8   si childrenQuestions.size() > 0 alors
9     * ListQ.add(questioni.order() + 1, childrenQuestions) // mettre les questions enfants
      après leur question parent dans la liste

```

---

À la validation de l'interrogatoire. L'instance de questionnaire est complétée avec les réponses de l'utilisateur puis stockée dans l'ontologie interrogatoire (IHO) sous un nouvel interrogatoire. Les données collectées sont ainsi stockées dans l'ontologie PPO qui représente le profil du patient.

### VI.3.2.2 Modifier un interrogatoire

Cette fonctionnalité permet aux utilisateurs de modifier les réponses aux questions déjà saisies en mode instantané. Cette fonctionnalité est importante car elle permet la rectification des erreurs de saisies, surtout de modifier la valeur des données collectées. L'algorithme 5 décrit le déroulement de la modification d'un interrogatoire en mode instantané :

---

#### Algorithme 5 : Modification d'un interrogatoire

---

**Données :** **QO** : Questionnaire Ontology ; **IHO** : Interrogation History Ontology ; **PPO** : Patient Profile Ontology

```

1 * ListQH ← loadHistory( IHO) // charger dans la liste l'historique de questions avec leurs
   réponses
2 * askAll(ListQH) // poser toutes les questions
   // Pour chaque question sélectionnée
3 pour chaque questionHistoryi ∈ ListQH faire
4   * responsei ← collect() ; // recueillir la réponse
5   * PPO ← responsei ; // sauvegarder la réponse en tant que donnée dans l'ontologie PPO
6   * IHO ← responsei ; // sauvegarder l'historique de la réponse dans l'ontologie IHO
7   * childrenQuestions ← loadChildren(questionHistoryi, responsei) // charger les questions
   enfants correspondantes à la responsei
8   si childrenQuestions.size() > 0 alors
9     * ListQH.add(questionHistoryi.order() + 1, childrenQuestions) // mettre les questions
   enfants après leur question parent dans la liste

```

---

L'algorithme consiste à charger dans un premier temps toutes les questions et éventuellement leurs réponses correspondant au questionnaire et au patient en question. Ces dernières sont proposées à l'utilisateur sous forme d'une liste ordonnée suivant l'ordre défini par les cliniciens.

À la validation de l'interrogatoire, ce dernier est enregistré dans l'ontologie de l'interrogatoires (IHO) sous une nouvelle version. Quant aux données collectées correspondantes aux réponses modifiées, elles sont modifiées dans le profil du patient.

### VI.3.2.3 Reprendre un interrogatoire

Cette fonctionnalité permet aux utilisateurs de reprendre des interrogatoires non terminés. Ceci est très utile car dans ce cas le moteur d'interrogatoire ne propose que les questions auxquelles le patient n'a pas répondu.

---

#### Algorithme 6 : Reprise d'un interrogatoire

---

**Données :** **QO** : Questionnaire Ontology ; **IHO** : Interrogation History Ontology ; **PPO** :

Patient Profile Ontology

```

1 * ListQ ← loadUnansweredParents( QO) // charger dans la liste les questions parents non
   répondues
2 * askAll(ListQ) // poser toutes les questions
   // Pour chaque question sélectionnée
3 pour chaque questioni ∈ ListQ faire
4   * responsei ← collect(); // recueillir la réponse
5   * PPO ← responsei; // sauvegarder la réponse en tant que donnée dans l'ontologie PPO
6   * IHO ← responsei; // sauvegarder l'historique de la réponse dans l'ontologie IHO
7   * childrenQuestions ← loadChildren(questioni, reponsei) // charger les questions enfants
   correspondantes à la reponsei
8   si childrenQuestions.size() > 0 alors
9     * ListQ.add(questioni.order() + 1, childrenQuestions) // mettre les questions enfants
   après leur question parent dans la liste

```

---

La reprise d'un interrogatoire en mode instantané est similaire au mode séquentiel. La différence réside dans le fait qu'on propose toutes les questions parents sans réponse au moment du chargement de l'interrogatoire.

## VI.4 Conclusion

Dans ce chapitre nous avons présenté notre contribution au niveau de l'acquisition de données pour les systèmes d'aide à la décision médicale. Nous avons présenté un outil générique de collecte de données à base d'ontologie et qui emploie un mécanisme de question/réponse. Cet outil permet la création des questionnaires d'une manière générique indépendamment de connaissances de domaine et du mode d'intervention du SADM. Il permet aussi à partir des modèles de questionnaires définis de collecter des données en déroulant des interrogatoires suivant deux modes : séquentiel et instantané. Nous avons par la suite présenté les modes de déroulements d'interrogatoires proposés à savoir le mode séquentiel, permettant de poser les questions les unes après les autres, et le mode instantané, permettant de poser toutes les questions en même temps. Enfin, pour chaque mode

nous avons présenté leur utilité dans le contexte médical.

# VII

## CONTRIBUTION DANS LE RAISONNEMENT À BASE DE CONNAISSANCES

Dans ce chapitre nous exposons notre contribution concernant l'intégration d'un raisonnement à base de connaissances guidé par une ontologie de domaine pour la construction d'un SADM à base de connaissances.

### Sommaire

---

<b>VII.1 Introduction</b> . . . . .	<b>83</b>
<b>VII.2 Raisonnement à base de règles</b> . . . . .	<b>84</b>
VII.2.1 Construction de la base de règles . . . . .	86
VII.2.2 Construction de la base de faits . . . . .	87
VII.2.3 Processus de raisonnement à base de règle . . . . .	88
<b>VII.3 Raisonnement à Partir des cas</b> . . . . .	<b>89</b>
VII.3.1 Représentation de cas . . . . .	90
VII.3.2 Processus de raisonnement . . . . .	92
VII.3.3 Modes d'exécution de raisonnement . . . . .	97
<b>VII.4 Conclusion</b> . . . . .	<b>99</b>

---

### VII.1 Introduction

Nous avons présenté dans le chapitre IV état de l'art sur les modes de raisonnement utilisés pour la construction des SADM à base de connaissances. Nous avons souligné que SADM à base de connaissances sont le plus appropriés pour la communauté médicale, étant donné que les modes de raisonnements employés dans ces derniers permettent de reproduire le raisonnement des experts du domaine (cliniciens) tout en exploitant leurs propres connaissances sous forme de règles ou sous forme de cas.

Dans l'approche de construction de SADM que nous proposons, nous intégrons le raisonnement à base de connaissances comme mode de raisonnement afin d'exploiter deux modes de connaissances :

- Les connaissances théoriques, représentées par les guides de bonnes pratiques, qui sont modélisées sous forme de règles ;
- Les connaissances pratiques, représentées par l'expérience des cliniciens, qui sont modélisées sous forme de cas.

Afin d'exploiter ces deux types de connaissances, nous proposons un raisonnement mixte à base de connaissances : le raisonnement à base de règles, qui permet essentiellement l'exploitation des guides de bonnes pratiques, et le raisonnement à base de cas, qui permet de capitaliser et d'exploiter l'expérience des cliniciens.

## VII.2 Raisonnement à base de règles

Dans l'état de l'art, nous avons montré que le raisonnement à base de règles (RàBR) est un mode de raisonnement à base de connaissances largement utilisé dans le domaine de l'intelligence artificielle pour reproduire les mécanismes cognitifs d'un expert. Il présente beaucoup d'avantages quant à sa facilité d'intégration et d'interprétation. En effet, le RàBR exploite les faits et les règles connues afin d'inférer de nouveaux faits, à leur tour utilisés éventuellement pour en inférer de nouveaux et ainsi de suite. Dans notre travail de thèse, nous intégrons le RàBR pour l'exploitation des guides de bonnes pratiques (GBP) sous forme de règles.

Les Guides de Bonnes Pratiques (GBP) sont des documents textuels médicaux qui décrivent les conduites cliniques les plus appropriées pour une pathologie donnée. Ils sont déterminés à partir des données les plus récentes de la littérature médicale et validées par un consensus d'experts (Georg, 2006). Les GBPs sont généralement élaborés par des organismes accrédités, comme la Haute Autorité de Santé (HAS)<sup>1</sup> pour la France, le National Institute for health and Clinical Excellence (NICE)<sup>2</sup> pour l'Angleterre, le National Guideline Clearinghouse (NGC)<sup>3</sup> et le National Heart, Lung, and Blood Institute (NIH)<sup>4</sup> pour les États-Unis (Georg, 2006). L'élaboration des GBPs est basée sur le paradigme EBM (Evidence-Based Medicine) ou la médecine fondée sur des preuves. L'EBM correspond à l'utilisation des meilleures données de la recherche scientifique dans la prise en charge personnalisée des patients.

Dans l'approche que nous proposons, les GBPs sont formalisés sous formes de règles par les ingénieurs des connaissances en collaboration avec les experts du domaine. Celles-ci sont ensuite interprétées par le moteur de raisonnement, qui exploite le profil du patient concerné, pour inférer des conclusions sous forme de nouveaux faits, d'alertes, d'indicateurs sur l'état de santé des patients et sur la qualité du suivi médical ou de recommandations.

Le raisonnement à base de règles peut être utilisé pour générer simplement de nouveaux

---

1. <http://www.has-sante.fr>

2. <http://www.nice.org.uk>

3. <http://www.guideline.gov>

4. <http://www.nhlbi.nih.gov>

faits, par exemple l'IMC (l'Indice de Masse Corporelle) est calculé à partir de la taille et du poids du patient selon la formule :  $IMC = Poids/Taille^2$ . L'IMC est enregistré dans le profil du patient comme une nouvelle donnée médicale. Cette donnée peut ensuite être utilisée par d'autres règles, par exemple, pour déterminer les risques d'accident cardio-vasculaire comme le montre la règle suivante : **SI (35 ≤ IMC ≤ 40) ALORS obésité sévère**, l'obésité étant considérée comme un facteur de risque.

Les règles peuvent être utilisées pour la génération des alertes quand une anomalie dans l'évolution d'état de santé du patient est détectée.

Les situations à risque, détectées à partir des données physiologiques et des antécédents médicaux du patient où des données inférées, sont matérialisées par des alertes.

Par exemple, pour un suivi des insuffisants cardiaques la règle suivante peut être exprimée comme suit :

**SI ((TA >130/80 mmHg OU Δ TA > ±10%) ET (FC > 70 bpm OU Δ FC > 20% FC)) ALORS alerte rouge**

Où TA : Tension artérielle et FC : Fréquence cardiaque. Alerte rouge : risque de décompensation cardiaque.

Dans cet exemple, le système alerte le corps médical pour éviter la décompensation en indiquant la sévérité de l'alerte.

Les règles peuvent être utilisées pour faire des calculs plus complexes comme pour déterminer le respect d'un protocole opératoire tel que le protocole ERAS (Enhanced Recovery After Surgery), appliqué pour la réhabilitation améliorée après une chirurgie colorectale, ou pour déterminer l'indicateur CCI (Comprehensive Complication Index) qui sert à mesurer la morbidité chirurgicale.

Dans notre approche et dans le but de concevoir un SADM ouvert, le raisonnement à base de règles doit être aussi bien générique que dynamique. Ceci nécessitera de relever les points suivants :

- La définition et l'intégration des règles doit être faite de manière simple, indépendamment de l'architecture du système et de la structure des données (faits).
- La structure des faits doit être évolutive et l'ajout d'un nouveau type de donnée ne doit engendrer aucune modification ni au niveau de l'architecture du système ni au niveau de la représentation des règles.
- L'exécution des règles par le moteur d'inférence doit être faite pour toute collecte de donnée sans interférer ni ralentir le processus de collecte.

Dans les sections suivantes nous détaillons l'approche de l'intégration d'un raisonnement à base de règles pour la construction d'un SADM à base de connaissances.



### VII.2.1 Construction de la base de règles

Les règles sont définies et stockées indépendamment de la base de connaissances. Ceci permet de rendre le système plus souple avec une définition des règles indépendante de la structure du profil de patient. Une règle est constituée des attributs suivants :

- **Nom** : identifiant pour la règle.
- **Description** : description précise du rôle de la règle et du ou des résultats attendus par cette dernière.
- **Activation** : cet attribut permet l'activation ou non de la règle. Il est utile quand on souhaite désactiver l'exécution d'une règle dans le système tout en la conservant dans la base.
- **Script** : cet attribut représente le script de la règle dont le moteur d'inférence se charge d'exécuter.
- **Déclencheurs** : cet attribut contient les éléments déclencheurs de la règle ou ce que nous appelons les éléments d'entrée de la règle. Ces éléments ce ne sont que les concepts définis dans l'ontologie de domaine DTOnto. Les éléments déclencheur permet au moteur d'inférence de déterminer si la règle est exécutable ou pas.

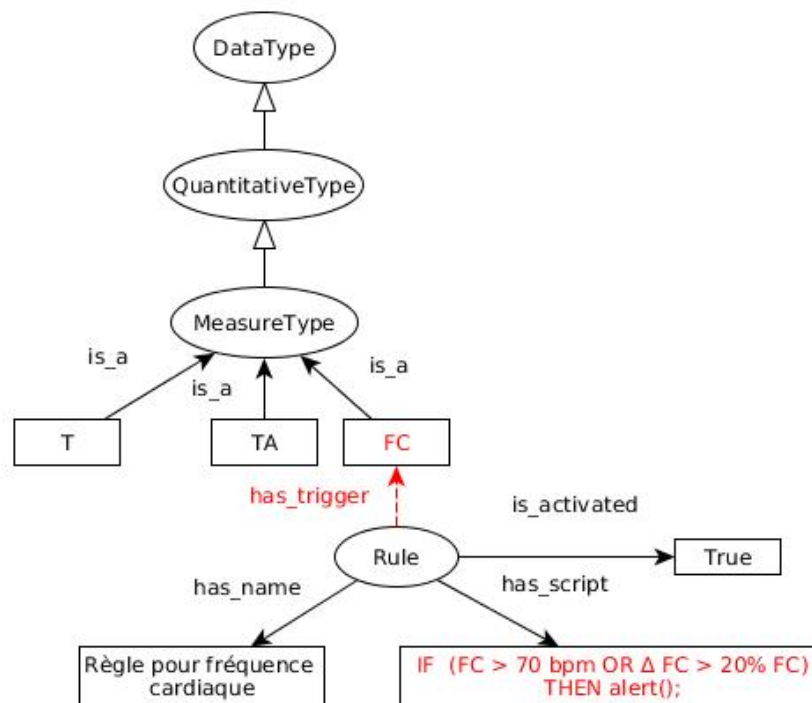


FIGURE VII.1 – Représentation d'une règle

La figure VII.1 illustre la représentation d'une règle pour la détection d'anomalie et l'inférence d'une alerte. La règle peut posséder un ou plusieurs éléments déclencheurs. Dans l'exemple de la figure VII.1, la règle possède un seul élément déclencheur qui est la fréquence cardiaque (FC) qui est définie comme un concept dans l'ontologie de domaine

(DTOnto). Ce qui revient à dire que la définition des règles est guidée par l'ontologie DTOnto dont les éléments déclencheurs des règles ne sont rien d'autres que les concepts de l'ontologie elle-même.

### VII.2.2 Construction de la base de faits

La base de faits est constituée de l'ontologie profil du patient (PPO) qui contient toutes les données collectées en relation avec le patient et qui contient ainsi les données inférées telle que les indicateurs, les alertes et les recommandations. C'est sur cette base que le moteur d'inférence exécute les règles pour produire des nouveaux faits qui sont ajoutés à la base ce qui modifiera l'état de la base de faits.

Dans l'exemple de la figure VII.2, la définition de la règle nécessite la définition du concept « fréquence cardiaque (FC) » dans l'ontologie DTOnto alors que l'exécution de la règle par le moteur d'inférence nécessitera la présence d'une valeur de FC dans l'ontologie du profil du patient PPO.

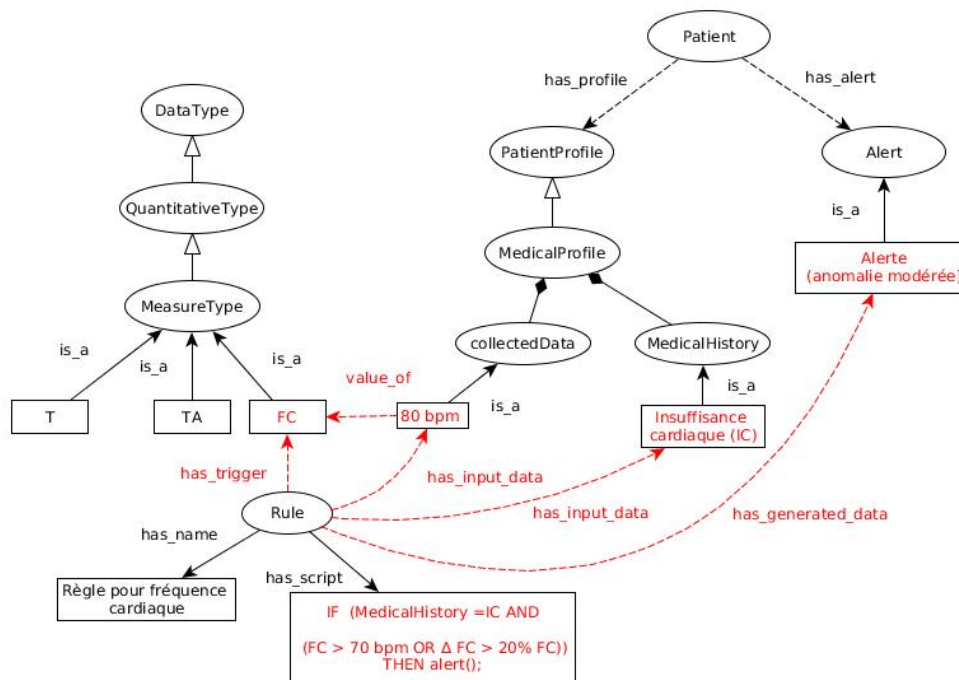


FIGURE VII.2 – Exemple de représentation de règle pour la détection d'anomalie pour un insuffisant cardiaque

Comme illustre la figure VII.2, l'exécution de la règle dépend des informations contenues dans le profil du patient et la règle est exécutée seulement si le patient est un insuffisant cardiaque et que la fréquence cardiaque est présente dans son profil. Ensuite, le fait inféré (l'alerte générée) est sauvegardé dans la base de faits en tant qu'anomalie modérée ce qui modifiera l'état de la base de faits.

### VII.2.3 Processus de raisonnement à base de règle

Le moteur d'inférence a pour but d'interpréter et d'exécuter les règles afin de générer des nouvelles données (faits).

Le processus d'inférence est décrit comme suit : dans un premier temps, on détermine les faits pertinents pour la base de règles, ensuite on détermine parmi les règles applicables la ou les règles qu'il convient de déclencher et, pour finir, on agrège éventuellement la partie conclusion des règles pour obtenir le résultat inféré à partir des faits.

Le processus d'inférence peut être décrit par la figure VII.3 :

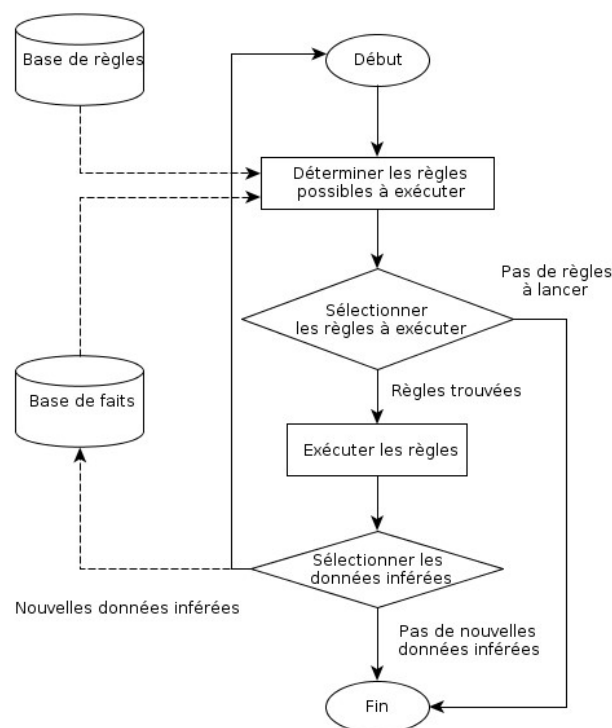


FIGURE VII.3 – Processus d'inférence

Dans l'approche que nous proposons ce processus d'inférence est lancé chaque fois qu'une donnée est introduite dans le système. Le moteur d'inférence vérifie s'il y a des règles qui peuvent être déclenchées. Si c'est le cas, il va les exécuter et sinon le processus arrive à sa fin. Après l'exécution des règles, si aucune donnée n'a été générée alors le processus s'arrête sinon les données générées sont enregistrées dans la base des faits (profil du patient) et le processus est réitéré.

Afin de décentraliser l'exécution des règles pour ne pas alourdir le système, nous avons opté pour une architecture orientée service dont le moteur de raisonnement est conçu dans un web service distant. La figure VII.4 illustre une représentation simplifiée de l'architecture orientée service.

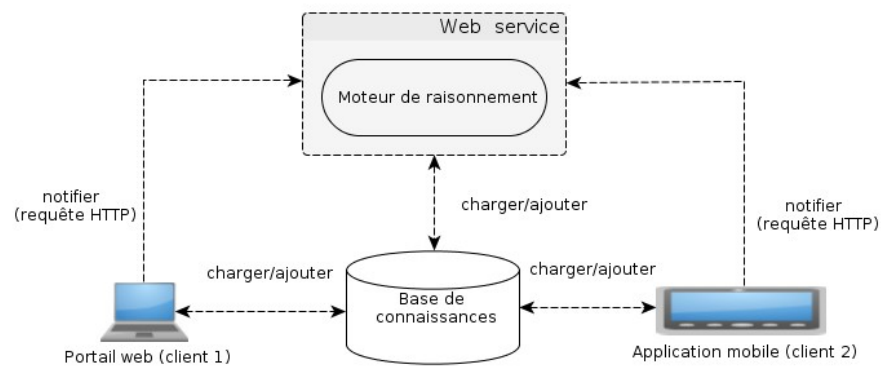


FIGURE VII.4 – Architecture orientée service pour la délocalisation du raisonneur.

Le moteur de raisonnement, qui intègre le moteur d'inférence, est considéré comme un web service distant tandis que le portail web et l'application mobile du système sont considérés comme des clients aux web service.

À chaque nouvelle donnée collectée par le système via le portail web (client 1) ou l'application mobile (client 2), le moteur est notifié par une requête HTTP afin que le processus d'inférence soit déclenché. Une fois l'exécution des règles terminée, les données inférées sont injectées dans la base de connaissances.

Le processus de raisonnement est exécuté de manière asynchrone pour ne pas bloquer les clients (portail web et application mobile) lors du processus de collecte de données.

### VII.3 Raisonnement à Partir des cas

Dans l'état de l'art, nous avons montré que le raisonnement à partir de cas est un mode de raisonnement à base de connaissances largement utilisé dans le domaine médical. Comme nous l'avons montré, l'un des inconvénients du RàPC est la structure des cas qui est rigide et qui ne peut pas évoluer. Ceci présente une énorme limitation dans l'évolution du système car la modification dans la structure du cas nécessitera souvent une intervention lourde dans le système. Étant donné que le but de nos travaux de recherches est la définition d'un SADM générique pour le suivi médical et prise en charge médicale, nous avons proposé une structure flexible de cas à base de concepts guidée par une ontologie de domaine.

Nous avons intégré le RàPC pour la construction de SADM à base de connaissances. Ce type de raisonnement est utilisé pour sa capacité de capitalisation de l'expérience clinique et d'apprentissage, il permet d'exploiter cette expérience pour la résolution des nouveaux cas cliniques.

### VII.3.1 Représentation de cas

Dans une technique de RàPC, les connaissances sur la façon d’accomplir les buts, sont représentées dans le système par un ensemble de cas sources. Dans notre approche un cas est principalement composé de trois parties, à savoir : le descripteur du problème, la solution et l’évaluation de la solution.

#### VII.3.1.1 Descripteur du problème

Cette partie contient la description de l’état de santé du patient au cours d’un suivi médical. Dans notre cas cette partie est représentée par le profil du patient  $P^i$ . La représentation du profil est faite à base d’une ontologie et elle est décrite dans le chapitre V. Le profil du patient  $P^i$  est utilisé pour calculer le degré de similarité entre le cas cible et les cas sources dans le processus de raisonnement.

#### VII.3.1.2 Solution

Cette partie contient la solution  $S^i$  qui a été proposée et validée par les professionnels de santé pour le profil  $P^i$ . Elle peut contenir plusieurs solutions à savoir une solution pour chaque épisode de soin.

Pour des raisons de généricité nous avons choisi de représenter la solution comme un acte médical (thérapie médicale). Ce dernier est modélisé par le concept ontologique MedicalAct et il est constitué comme suit :

- **Title** : le titre de l’acte médical.
- **Description** : la description de l’acte médical qui est représentée par un texte libre pour offrir plus d’expressivité.
- **Date** : la date de la l’acte médical.

L’acte médical est relié aux concepts de différentes ontologies comme suit :

- **Patient** : représente le patient concerné par l’acte médical et fait référence à l’ontologie Profil du patient (PPO).
- **MedicalStaff** : représente le professionnel de santé qui a effectué l’acte médical et fait référence à l’ontologie d’application du système. Ceci permet de garder une traçabilité afin de savoir qui a fait quoi, ce qui est important pour un suivi médical.
- **Tag** : représente la connaissance qui a été à l’origine de l’intervention du professionnel de santé pour poser l’acte médical. Un acte médical peut être relié à un ou plusieurs tags dont chaque tag est relié au concept correspondant dans DTOnto.

La figure VII.5 illustre une représentation simplifiée d’un cas médical pour un insuffisant cardiaque dans le cas d’une hypotension artérielle.

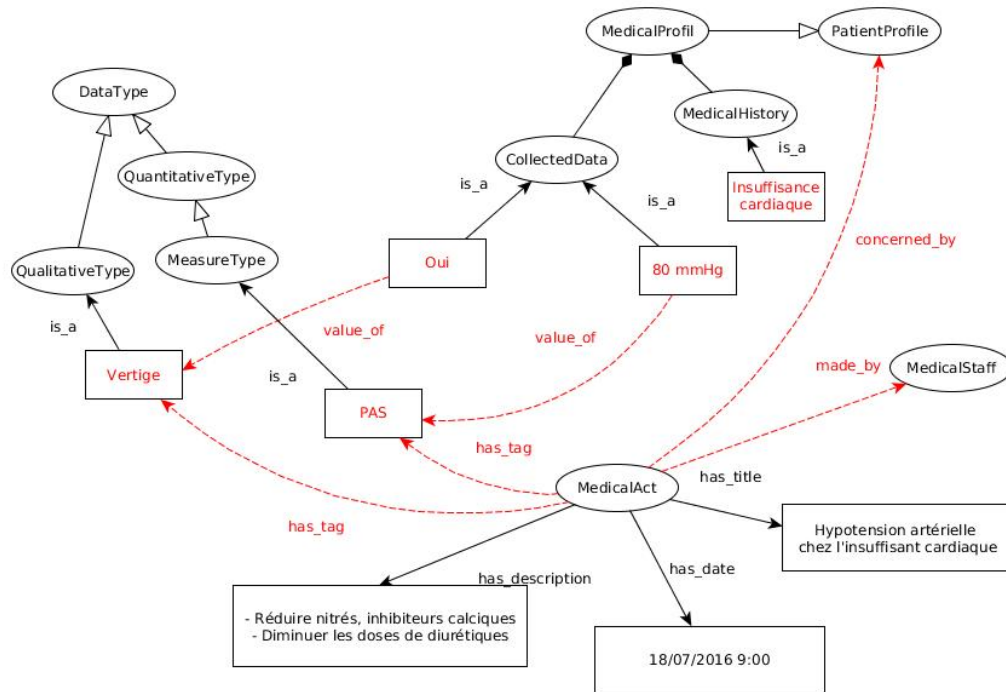


FIGURE VII.5 – Exemple d’une représentation simplifiée de deux parties problème-solution d’un cas

L’exemple illustré dans la figure VII.5 peut se traduire comme suit :

- Description du problème (profil du patient) :
  - Antécédents : insuffisance cardiaque.
  - Symptôme : vertiges.
  - Mesure physiologique : pression artérielle systolique (PAS) de 80 mmHg.
- Solution de cas (l’acte médical) :
  - Réduction nitrée et inhibiteurs calciques.
  - Diminution des doses de diurétiques.

Dans l’approche que nous proposons, la représentation de cas est guidée par l’ontologie du domaine DTOnto. Ceci permet d’offrir une genericité dans la représentation de la partie descripteur du problème (le profil du patient) où les données collectées font référence aux concepts appropriés dans DTOnto et dans la représentation des tags dans la partie solution, où eux aussi font référence aux concepts appropriés dans DTOnto.

### VII.3.1.3 Résultat et évaluation

Cette partie contient l’évaluation  $E^i$  de la solution proposée. Un score est attribué pour chaque solution proposée.

Chaque cas, de notre base de cas, modélise la liaison entre le profil du patient, les solutions proposées et l’évaluation des solutions. Un cas est dénoté par le triplet :  $Cas = (P^i, S^i, E^i)$ .

### VII.3.2 Processus de raisonnement

Le processus de raisonnement est constitué de plusieurs étapes tel qu'il a été proposé par (Aamodt and Plaza, 1994) :

1. **La phase de recherche d'un cas similaire** : permet de sélectionner les cas sources les plus similaires au cas cible sur la base d'une mesure de similarité des profils de patients.
2. **La phase de réutilisation** : permet d'adapter le(s) solution(s) par rapport au profil du patient en cours en utilisant les cas sélectionnés dans la phase de recherche et de récupérer le feedback des cliniciens (leurs évaluations).
3. **La phase d'apprentissage** : assure la mise à jour et l'évolution de la base de cas par des nouvelles connaissances (cas).

#### VII.3.2.1 Recherche des cas les plus similaires

Dans nous cette étape nous distinguons deux phases :

**Recherche des profils les plus similaires.** À chaque nouveau patient, un cas cible est construit avec comme problème le profil du patient. Pour déterminer le cas source le plus similaire au cas cible, la base des cas est parcourue pour comparer le profil en cours aux profils cliniques précédents. Ensuite les cas similaires sont renvoyés pour l'étape suivante. Soit  $PS = \{P^1, \dots, P^n\}$  l'ensemble des profils dans la base des cas. Le raisonnement consiste à sélectionner le profil  $P^{opt}$  qui vérifie :

$$P^{opt} = \frac{\sum_{i=1}^m w_i \times sim_i(A_j^*, A_j^i)}{\sum_{i=1}^m w_j} \geq \alpha \quad (VII.1)$$

Où  $A_j^*$  (resp.  $A_j^i$ ) est la valeur du  $j^{\text{ème}}$  attribut du profil  $P^*$  (resp.  $P^i$ ),  $sim_j$  est la mesure de similarité pour le  $j^{\text{ème}}$  attribut et  $w_j$  est le poids associé.  $\alpha$  est un seuil défini par les experts médicaux.

Notre choix s'est porté sur la similarité Manhattan pondérée permettant d'obtenir un score de similarité compris entre 0 et 1.

Ci-dessous, nous décrivons les mesures de similarité définies sur chacun des attributs de profil en fonction de leurs natures.

- La similarité entre deux attributs symboliques est une similarité booléenne qui ne peut prendre que deux valeurs : 0 ou 1. Par exemple la similarité par rapport à l'attribut sexe du patient. Cette similarité est calculée comme suit :

$$sim_{smb}(a_i, a_j) = \begin{cases} 1 & \text{si } a_i = a_j \\ 0 & \text{sinon} \end{cases} \quad (VII.2)$$

Ceci signifie que si deux patients ont le même sexe alors la similarité est égale à 1, sinon 0.

- La similarité entre deux séries de données symboliques est calculée comme suit :

$$sim_{ss}(A^i, A^*) = \frac{\sum_{k=1}^N sim_{smb}(a_k^*, a_k^i)}{N} \quad (\text{VII.3})$$

Où  $N$  est le nombre d'éléments dans la série et  $a_k^*$  (resp.  $a_k^i$ ) est le  $k^{\text{ème}}$  élément de série de donnée  $A^*$  (resp.  $A^i$ ).  $sim_{smb}$  est similarité qui vaut 0 ou 1.

Cette fonction de similarité est utilisée principalement dans le cas où les données de la série ne sont pas liées sémantiquement entre elles.

Par exemple dans un suivi opératoire certaines données sont collectées quotidiennement dans la phase postopératoire. Ces données peuvent prendre des valeurs non-numérique, par exemple « Oui » ou « Non », qui constituent une série de données où les éléments ne sont pas liés sémantiquement.

- La similarité entre deux attributs numériques, par exemple l'âge et la taille est calculée selon la formule utilisée dans (Gottlieb et al., 2013) :

$$sim_{num}(a_i, a_j) = 1 - \frac{|a_i - a_j|}{\max(a_i, a_j)} \quad (\text{VII.4})$$

Cette fonction de similarité est utilisée quand nous ne connaissons pas la valeur maximale que l'attribut peut prendre.

- La similarité entre deux évaluations (douleur, fatigue, sommeil, etc.) est considérée comme une similarité numérique sauf que la valeur maximale qui peut prendre une évaluation est connue. Cette similarité est calculée selon la formule suivante :

$$sim_{eva}(a_i, a_j) = 1 - \frac{|a_i - a_j|}{N} \quad (\text{VII.5})$$

Où l'attribut  $a_i$  et  $a_j \in E/E = \{0, 1, 2, \dots, N\}$  et  $N$  est la valeur maximale de l'évaluation.

- La similarité entre deux ensembles, en prenant en compte la similarité sémantique entre les éléments de l'ensemble, est calculée en utilisant la formule BMA (Best-Match Average) proposée par (Wu et al., 2013). Cette formule a été appliquée au calcul de similarités entre deux ensembles de terminologies de l'ontologie de gènes (Gene Ontology (GO)).

Nous utilisons cette formule pour calculer la similarité des antécédents médicaux entre deux patients. Étant donné que les antécédents médicaux (pathologies, traitements et interventions chirurgicales) sont représentés par des ensembles non ordonnés où ils n'ont pas forcément la même cardinalité.

Soit  $A^* = \{a_1^*, a_2^*, \dots, a_n^*\}$  et  $A^i = \{a_1^i, a_2^i, \dots, a_m^i\}$ . La similarité est donc calculée selon la formule suivante :



$$sim_{BMA}(A^*, A^i) = \frac{\sum_{k=1}^m \max_{1 \leq j \leq n} sim_{sem}(a_k^*, a_j^i) + \sum_{j=1}^n \max_{1 \leq k \leq m} sim_{sem}(a_j^*, a_k^i)}{m + n} \quad (VII.6)$$

Où  $m = Card(A^*)$  et  $n = Card(A^i)$  et  $sim_{sem}$  est la similarité sémantique entre deux concepts  $C_j$  et  $C_k$  de l'ontologie de domaine correspondants respectivement aux attributs  $a_j$  et  $a_k$ . La similarité sémantique entre deux concepts dépend de leur degré de proximité dans la taxonomie (ontologie de domaine). Nous utilisons la mesure de similarité sémantique décrite dans (Wu and Palmer, 1994) qui propose de calculer la similarité sémantique en s'appuyant sur la longueur du chemin entre deux concepts d'une même hiérarchie. Cette mesure est définie comme suit :

$$sim(c_i, c_j) = \frac{2 \times depth(LCS)}{depth(c_i) + depth(c_j)} \quad (VII.7)$$

Où LCS est le subsumant commun le plus spécifique,  $depth(LCS)$  est la longueur du chemin entre  $LCS$  et la racine de la hiérarchie,  $depth(C_i)$  est le nombre d'arcs entre ( $C_j$ ) et la racine en passant par LCS.

Cette formule nous permet de calculer la similarité entre deux pathologies ou interventions chirurgicales en utilisant l'ontologie CIM-10 et la similarité entre deux traitements en utilisation l'ontologie ATC. Par exemple, la similarité entre **Infarctus du myocarde** et **angine de poitrine** (pathologies cardiovasculaires) est plus importante que la similarité entre **Infarctus du myocarde** et **diabète non insulino-dépendant**.

- La similarité entre deux séries de données numériques telles que les données physiologiques (poids, température, tension artérielle, etc.) est calculée en combinant la formule de Bray-Curtis et normalisation de Pearson. La similarité est donnée comme suit :

$$sim(A^*, A^i) = \alpha \times sim_{Bray-Curtis}(A^*, A^i) + \beta \times sim_{Pearson}(A^*, A^i) \quad (VII.8)$$

Où  $\alpha, \beta \in [0, 1]$  et  $\alpha + \beta = 1$

La similarité de Bray-Curtis est donnée par la formule suivante :

$$Sim_{Bray-Curtis}(A^*, A^i) = \frac{2 \times \sum_{k=1}^n \min(a_k^*, a_k^i)}{\sum_{k=1}^n (a_k^* + a_k^i)} \quad (VII.9)$$

Et la similarité de Pearson normalisée telle que proposée par (Panchenko and Morozova, 2012) est donnée par la formule suivante :

$$sim_{Pearson}(A^*, A^i) = \frac{Corre_{Pearson}(A^*, A^i) - \theta}{\delta} \quad (\text{VII.10})$$

Où  $Corre_{Pearson}(A^*, A^i)$  noté  $r$  est la corrélation de Pearson,  $r \in [-1, 1]$ . Le coefficient de corrélation est donc compris entre  $-1$  et  $1$ . Une valeur proche de  $+1$  signifie une forte liaison linéaire entre les deux séries de données (les données variées dans le même sens). Une valeur proche de  $-1$  signifie également une forte liaison mais la relation linéaire entre les deux séries de données est décroissante (les données varient dans le sens contraire). Tandis qu'une valeur proche de  $0$  signifie une absence de relation linéaire entre les deux séries de données.

$\theta$  correspond au négatif total (i.e.  $\theta = -1$ ) et  $\delta$  correspond à l'étendu (i.e. La différence entre la plus grande et la plus petite valeur  $\delta = 1 - (-1) = 2$ ). Les valeurs de la similarité normalisée de Pearson sont donc comprises entre  $0$  et  $1$ .

La corrélation de Pearson est donnée par la formule suivante :

$$Corre_{Pearson}(A^*, A^i) = \frac{\sum_{k=1}^n (a_k^* - \bar{A}^*)(a_k^i - \bar{A}^i)}{\sqrt{\sum_{k=1}^n (a_k^* - \bar{A}^*)^2} \sqrt{\sum_{k=1}^n (a_k^i - \bar{A}^i)^2}} \quad (\text{VII.11})$$

Où  $\bar{A}^*$  correspond à la moyenne de l'échantillon et elle est calculée comme suit :

$$\bar{A}^* = \frac{1}{n} \sum_{k=1}^n a_k^* \quad (\text{VII.12})$$

Où  $n$  est le nombre d'éléments dans la série de données.

**Recherche les solutions les plus pertinentes.** Les solutions  $S^{opt}$  correspondantes aux profils les plus similaires  $P^{opt}$  sélectionnés à la suite de la phase précédente seront proposées comme des solutions potentielles pour le cas cible.

Pour assurer une meilleure précision des résultats, le moteur de cas doit calculer un score de pertinence de la solution  $S^{opt}$  selon la formule suivante :

$$Relevance(S^{opt}) = \frac{1}{N} \sum_{k=1}^N SV(tag_k) \quad (\text{VII.13})$$

Où  $N$  correspond au nombre de tags dans la solution et  $S^{opt}$  et  $SV(tag_k)$  correspond au score de similarité obtenu par rapport au tag (attribut) dans  $P^{opt}$ .

Le score de pertinence  $Relevance(S^{opt})$  est comparé à un seuil  $\gamma$  et le score de similarité  $SV(tag_k)$  est comparé au seuil  $\lambda$  comme suit :  $Relevance(S^{opt}) \geq \gamma$  et  $SV(tag_k) \geq \lambda$

Où les seuils de pertinences sont définis par les experts du domaine.

Enfin, seules les solutions retenues seront proposées à l'étape de réutilisation.

L'algorithme 7 décrit le processus de recommandation des solutions (actes médicaux) en fonction des similarités calculées.

---

**Algorithme 7 : Recommandations des actes médicaux**

---

**Entrées :** `profilesList` : profils patients les plus similaires ;

**Sorties :** `medicalActsList` : actes médicaux (solutions) à recommandés ;

```

1 pour chaque  $profile_i \in profilesList$  faire
2   pour chaque  $medicalAct_j \in profile_i.medicalActsList()$  faire
3     relevance=0;
4     pour chaque  $tag_k \in medicalAct_j.tagsList()$  faire
5       pour chaque  $sim_h \in profile_i.similaritiesList()$  faire
6         si  $sim_h.tag() == tag_k$  alors
7           //  $\lambda$  est le seuil de pertinence du tag
8           si  $sim_h.value() \geq \lambda$  alors
9             relevance+=sim_h.value();
10          sinon
11            relevance=0;
12          relevance = relevance/N; // N correspond au nombre de tags dans  $medicalAct_j$ 
13          //  $\gamma$  est le seuil de pertinence de l'acte médical
14          if relevance >=  $\gamma$  then
15            medicalActsList.add( $medicalAct_j$ );
16 return medicalActsList;

```

---

### VII.3.2.2 Étape de réutilisation

L'ensemble des solutions  $S^{opt}$  retenues dans l'étape précédentes sont proposées comme des solutions potentielles au cas cible.

Cependant, la réutilisation d'un cas pourrait entraîner une adaptation de la solution où les cliniciens peuvent modifier les solutions recommandées afin de les adapter au cas cible et ensuite les enregistrer dans la base. Enfin, cette étape permet aussi de récupérer le feedback des cliniciens (leurs évaluations).

### **VII.3.2.3 Étape de mémorisation**

Dans cette étape, le nouveau cas résolu est ajouté à la base dans l'espoir que sa solution puisse à son tour aider le même clinicien ou un autre clinicien dans la résolution de nouveaux cas similaires.

### **VII.3.3 Modes d'exécution de raisonnement**

Nous avons proposé deux modes pour l'exécution du processus de raisonnement à base de cas : l'exécution à la demande et l'exécution à la collecte de données. Ces deux modes concernent uniquement le calcul des similarités.

#### **VII.3.3.1 Exécution à la demande**

Dans ce mode, le calcul des similarités entre le cas cible et les cas sources est effectué à la demande du clinicien.

Dans ce mode, le clinicien définit les critères de recherche des cas similaires en sélectionnant l'ensemble des attributs pris en compte pour le calcul des similarités tout en définissant les seuils de sélection. Une fois le processus lancé, le moteur de cas parcourt la base de cas pour sélectionner les cas les plus similaires en fonction des critères définis. Il renvoie ensuite la liste des cas les plus similaires avec les similarités qui correspondent.

Le calcul des scores de similarité, dans ce mode d'exécution, est effectué seulement si le besoin se présente. Ceci évite la consommation constante des ressources due au chargement des données et au calcul des similarités.

Par ailleurs, ce mode d'exécution n'est pas adapté aux bases de cas volumineuses où le calcul des scores de similarité est gourmand en temps. Nous avons donc proposé un deuxième mode d'exécution permettant l'optimisation de calcul.

#### **VII.3.3.2 Exécution à la collecte de données**

Dans le but d'optimiser le calcul des scores de similarité, nous avons proposé un mode d'exécution en arrière-plan. Le calcul des similarités se fait en même temps que l'acquisition des données sans bloquer le processus d'acquisition de données.

Pour ce faire, nous avons opté pour une architecture orientée services où le moteur de cas est considéré comme un service distant. Le système notifie le moteur de cas à chaque collecte de données par une requête HTTP contenant l'identifiant du patient et le type de donnée concernés. Le moteur relance le calcul de similarités pour le patient concerné et ensuite enregistre les scores de similarités dans une matrice appelée matrice de similarités (voir figure VII.6).

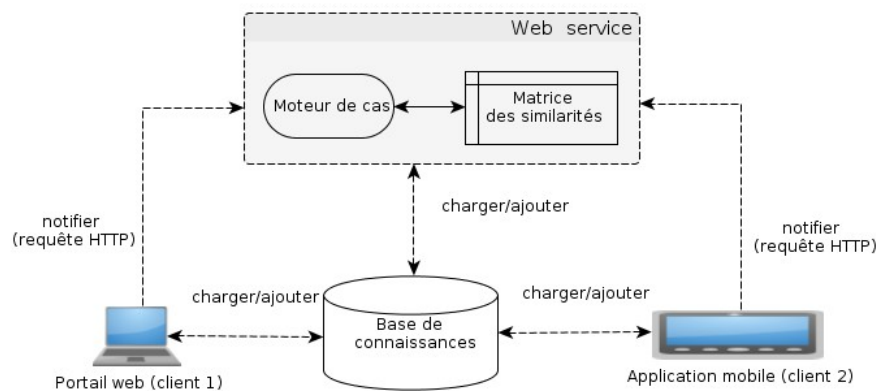


FIGURE VII.6 – Architecture orienté service pour le raisonnement à base de cas

Pour éviter la redondance dans le calcul et les calculs inutiles, les notifications venant du système sont enregistrées dans une file d'attente selon leur ordre d'arrivée. Une notification est caractérisée par l'identifiant du patient concerné et par le type de donnée qui a été mise à jour.

Le moteur vérifie constamment la file d'attente afin de vérifier s'il y a des nouvelles notifications non traitées. Si elles existent, le moteur prend celle qui se trouve en tête de file selon le mode FIFO (First In First out) et il recalcule les similarités qui concernent le type de données et le patient enregistrés. Ceci permet d'éviter de recalculer les similarités pour tous les types de données et pour tous les profils patients existant dans la base.

Une fois que ce calcul est terminé, le moteur met à jour la matrice des similarités par les nouvelles valeurs de score de similarité, puis il vérifie de nouveau la file d'attente. Si aucune notification n'est trouvée dans la file, le moteur se met en attente de nouvelles notifications.

L'algorithme 8 décrit la mise à jour de la matrice des similarités :

---

**Algorithme 8** : Mise à jour de la matrice des similarités
 

---

**Données** : **WL** : file d'attente des notifications ; **SM** : Matrice des similarités ;

```

1 tant que  $WL \neq \emptyset$  faire
2   *  $notification_i \leftarrow WL.getFirst()$  ; // charger et supprimer la première notification de la liste
    d'attente
3   *  $patient \leftarrow notification.patient()$  ; // charger l'identifiant du patient concerné
4   *  $dataType \leftarrow notification.dataType()$  ; // charger le type de donnée concerné
5   *  $similarityScore \leftarrow similarityComputing(patient, dataType)$  ; // calculer les nouveaux
    scores de similarité
6   *  $update(SM, similarityScore)$  ; // Mettre à jour la matrice avec les nouveaux scores de
    similarités
  
```

---

Dans ce mode d'exécution, l'étape de recherche de cas similaires consiste uniquement à charger la matrice des similarités contenant le score de toutes les similarités au lieu de faire le calcul au lancement de raisonnement comme dans le mode précédent. Ceci est très utile quand le système dispose d'une base de cas importante où le parcours de la base nécessiterait énormément de temps.

## VII.4 Conclusion

Dans ce chapitre, nous avons présenté notre contribution dans la définition d'un mode de raisonnement mixte à base de connaissance guidé par une ontologie de domaine.

Comme nous l'avons montré, l'intégration de ce mode de raisonnement dans un système d'aide à la décision médicale offre plus de généralité et de flexibilité par rapport à ce qui a été proposé dans l'état de l'art. Les connaissances utilisées pour le raisonnement sont enrichies de manière dynamique et l'ajout de nouvelles connaissances n'impactera ni la structure de données ni sur l'architecture du système. Par conséquent, la construction du moteur de raisonnement ne dépend ni du domaine d'application du système ni de son type d'intervention.

Par ailleurs, nous avons souligné que le mode de raisonnement proposé permet à la fois d'exploiter les sources théoriques, telles que les guides de bonnes pratiques à travers le raisonnement à base de règles, et les sources pratiques, telle que l'expérience clinique à travers le raisonnement à base de cas.



# III

## IMPLEMENTATION ET EXPÉRIMENTATION





# VIII

## IMPLÉMENTATION

Dans ce chapitre, nous présentons l'implémentation de nos travaux de recherche pour la construction d'un SADM ouvert dédié au suivi médical. Nous détaillons l'implémentation de ses différentes composantes permettant l'acquisition de données médicales, la représentation des connaissances et le raisonnement à base de connaissances.

### Sommaire

---

<b>VIII.1</b>	<b>Introduction</b>	<b>104</b>
<b>VIII.2</b>	<b>Plateforme E-care</b>	<b>104</b>
VIII.2.1	Architecture générale de la plateforme E-care	104
VIII.2.2	Rôles de la plateforme E-care	106
VIII.2.3	Langages et outils utilisés pour le développement de la plateforme	106
<b>VIII.3</b>	<b>Acquisition de connaissance spécifiques du domaine d'intervention</b>	<b>107</b>
VIII.3.1	Construction de l'ontologie DTonto	107
VIII.3.2	Création d'instances de l'ontologie DTonto via le portail web	107
<b>VIII.4</b>	<b>Profil du patient</b>	<b>109</b>
VIII.4.1	Construction de l'ontologie PPO	109
VIII.4.2	Création de compte patient	110
VIII.4.3	Renseignement des antécédents médicaux	111
<b>VIII.5</b>	<b>Outil d'acquisition de données</b>	<b>112</b>
VIII.5.1	Construction des ontologies	112
VIII.5.2	Création de questionnaires via le portail web	116
VIII.5.3	Module de collecte de données	120
VIII.5.4	Module de l'historique des interrogatoires	124
<b>VIII.6</b>	<b>Moteur de raisonnement à base de règles</b>	<b>125</b>
VIII.6.1	Gestion des règles d'inférence	125
VIII.6.2	Exécution des règles	126
VIII.6.3	Inférence des données à partir des règles	127
<b>VIII.7</b>	<b>Moteur de raisonnement à base de cas</b>	<b>129</b>
VIII.7.1	Gestion des thérapies médicales	129
VIII.7.2	Recherche des cas similaires	130
<b>VIII.8</b>	<b>Conclusion</b>	<b>133</b>

---

## VIII.1 Introduction

Nous avons présenté dans la partie précédente notre contribution concernant la construction d'un système d'aide à décision médicale (SADM) et nous avons décrit l'approche proposée pour que le SADM soit facilement adaptable et configurable en fonction de son domaine d'intervention. En effet, cette approche permet la construction d'un SADM ouvert à base de connaissances qui ne dépend ni de son domaine et ni de son mode d'intervention. Dans ce chapitre, nous présentons l'implémentation de notre approche pour la construction d'un système générique à base de connaissances dédié au suivi médical. Dans la continuité du projet E-care, décrit dans la section suivante, nous avons choisi d'enrichir la plateforme E-care en y intégrant notre approche. Cette plateforme sera ainsi en mesure d'organiser la collecte des données médicales auprès des patients en prenant en compte le contexte du patient, d'analyser les données collectées à l'aide d'un moteur de raisonnement à base de cas guidée par une ontologie de domaine et de fournir des résultats permettant d'aider les professionnels de santé dans le suivi des patients. Après avoir présenté l'architecture générale de la plateforme E-care et ses principaux rôles, dans la section suivante, nous présenterons l'acquisition de connaissances du domaine utilisées dans la plateforme, la création du profil du patient et son enrichissement. Nous décrirons ensuite l'outil d'acquisition de données médicales, l'implémentation du moteur de raisonnement à base de règles et la gestion des règles d'inférences avant d'aborder, pour finir ce chapitre, l'implémentation du moteur de raisonnement à base de cas et la recommandation des thérapies. La dernière section conclut le chapitre.

## VIII.2 Plateforme E-care

La plateforme E-care a été développée dans le cadre d'un projet collaboratif financé par les investissements d'avenir. Le projet a été porté par la société NEWEL en collaboration avec l'UTBM, le CHU de Strasbourg, l'UHA (Université de Haute-Alsace) et le CENTICH (Centre d'Expertise National des Technologies de l'Information et de la Communication pour l'autonomie). L'objectif de cette plateforme est d'assurer le suivi et la prise en charge des patients dans toutes leurs démarches de soin.

### VIII.2.1 Architecture générale de la plateforme E-care

La figure VIII.1 présente l'architecture générale de la plateforme E-care. Dans son ensemble, elle peut être sous-divisée en un certain nombre de composants indépendants. Ces derniers interagissent entre eux pour accomplir les rôles décrits ci-dessous. Nous détaillons dans ce qui suit chaque composant :

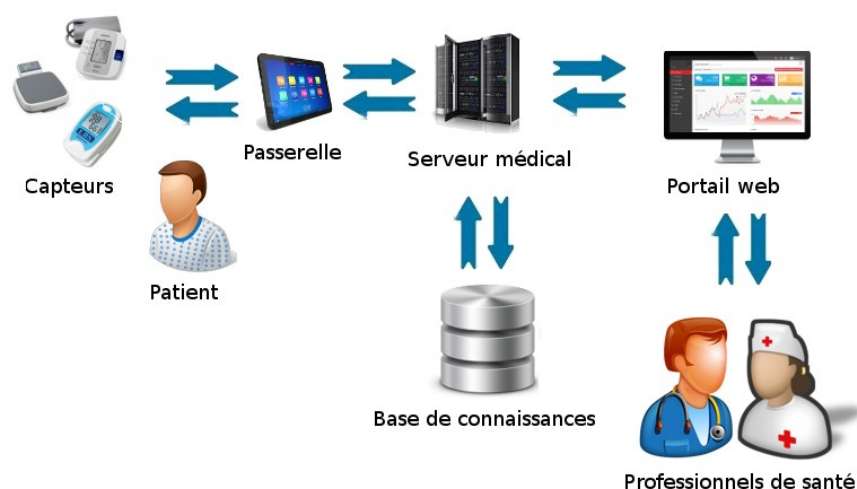


FIGURE VIII.1 – Architecture globale de la plateforme E-care

- **Base de connaissances** : un composant primordial pour tout système à base de connaissances. Elle permet le stockage de connaissances liées au domaine, les données utilisées pour le fonctionnement de la plateforme et toutes les données collectées sur les patients.
- **Portail web** : une application web dédiée aux professionnels de santé ainsi qu’aux patients. Il regroupe différents modules permettant aux professionnels de santé de suivre leurs patients d’une manière efficace et rapide. Il fournit également des interfaces permettant aux patients de consulter leurs données recueillies, etc.
- **Capteurs médicaux** : permettent la collecte automatique de certaines mesures physiologiques (température, tension artérielle, oxymétrie, fréquence cardiaque et glycémie). Ils permettent également de mesurer l’activité physique ou encore la qualité de sommeil du patient.
- **Passerelle** : un dispositif mobile (tablette ou smartphone) qui sert principalement à relier les capteurs médicaux et le serveur médical. Elle récupère les données des capteurs, permet leurs visualisations sous différentes formes, et leurs transfère au serveur. Elle dispose d’une application qui permet de fournir des fonctionnalités telles que la visualisation de mesures physiologiques sous forme de graphe ainsi que la collecte de données à l’aide de l’outil d’acquisition de données que nous détaillons dans la suite du chapitre.
- **Serveur médical** : il assure la communication entre la passerelle et le reste de la plateforme. Il récupère les données provenant de cette dernière et les injecte dans la base de connaissances. Il analyse ainsi constamment les données recueillies par le biais du moteur de raisonnement pour fournir des résultats à la plateforme.

### VIII.2.2 Rôles de la plateforme E-care

La plateforme E-care est considérée comme un système d'aide à la décision médicale à base de connaissances qui a essentiellement trois rôles principaux :

- **Collecte de données** : pour assurer le bon suivi médical, la plateforme collecte constamment des données médicales sur les patients (données physiologiques, test de laboratoire, etc.) et sur leur hygiène et qualité de vie (sommeil, activités physiques, habitudes alimentaires, etc.).
- **Raisonnement et traitement de données** : les données collectées sont ensuite traitées à l'aide d'un moteur de raisonnement à base de connaissances issues des experts du domaine (cliniciens) pour fournir des résultats.
- **Présentation des résultats des traitements** : une fois les données collectées traitées par le moteur de raisonnement, la plateforme est en mesure de fournir des résultats en fonction de son type d'intervention :
  - Proposer des actes médicaux (thérapies) personnalisés pour les patients afin d'optimiser et d'améliorer toute démarche de soin.
  - Alerter le corps médical si une anomalie est détectée dans l'évolution de l'état de santé du patient.
  - Fournir des indicateurs sur la qualité de soin et l'évolution de l'état de santé du patient.
  - Proposer des recommandations personnalisées pour l'amélioration de l'hygiène de vie des patients.

### VIII.2.3 Langages et outils utilisés pour le développement de la plateforme

Étant donné que le système utilise des ontologies, nous avons choisi l'outil Protégé<sup>1</sup>, réputé par sa puissance et ses multiples fonctionnalités, pour manipuler les connaissances et les modéliser en langage OWL. Nous avons utilisé Virtuoso comme serveur de base de données. Il permet de manipuler les ontologies où les instances des ontologies (données) sont stockées sous forme de triplets en langage RDF. Les triplets RDF sont interrogés à l'aide de langage de requête SPARQL depuis la plateforme. La partie portail et serveur de la plateforme ont été développés en utilisant la technologie Java/JEE sous le Framework Spring. Tandis que l'application mobile a été développée sous Android. Dans la suite du chapitre nous détaillons les modules développés et qui sont en relation avec l'approche présentée précédemment à savoir l'acquisition et la représentation des connaissances de domaine, l'acquisition des données médicales et le raisonnement à base de connaissances.

---

1. <http://protege.stanford.edu>

### VIII.3 Acquisition de connaissances spécifiques du domaine d'intervention

Nous avons implémenté un module d'acquisition des connaissances spécifiques du domaine d'intervention de la plateforme. Le développement de ce module nécessite en effet la construction de l'ontologie DTOnto et les interfaces web associées permettant la création et la mises à jour des instances de cette dernière.

#### VIII.3.1 Construction de l'ontologie DTOnto

Le diagramme UML, présenté dans partie contribution, a été traduit dans le langage OWL en utilisant l'outil Protégé où les classes qui composent l'ontologie ont été créées comme le montre la figure VIII.2.

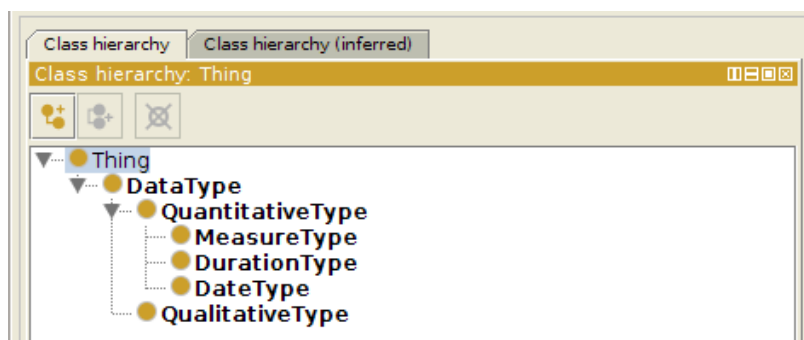


FIGURE VIII.2 – Création de classes de l'ontologie DTOnto sous Protégé

La relation d'héritage entre les différentes classes a été modélisée par la propriété OWL `< rdfs :subClassOf >` permettant ainsi la définition de liens hiérarchiques.

#### VIII.3.2 Création d'instances de l'ontologie DTOnto via le portail web

Nous avons développé des interfaces web, accessibles depuis le portail web de la plateforme, qui permettent aux experts médicaux d'ajouter et de mettre à jour les instances de l'ontologie DTOnto de manière simple sans à avoir recours à l'outil Protégé. Les connaissances spécifique du domaine d'intervention de la plateforme sont donc introduites dans la plateforme par le biais d'une interface web en précisant simplement leur nom et leur format. La figure VIII.3 montre l'ajout de l'instance « durée de l'opération » comme connaissance spécifique du domaine de la chirurgie où elle est de format « durée ».

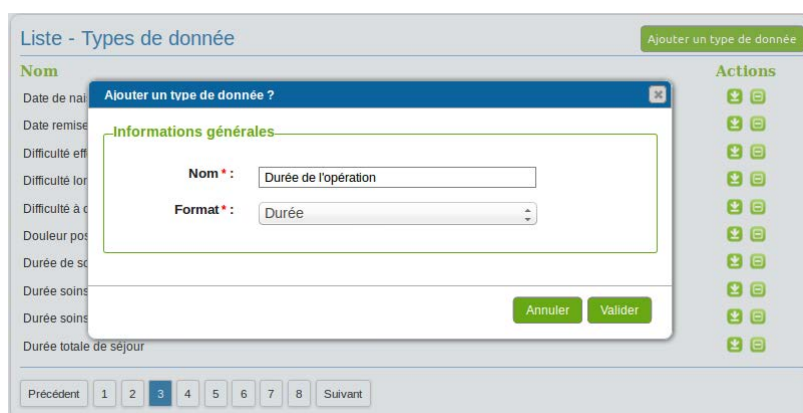


FIGURE VIII.3 – Ajout d’une connaissance spécifique du domaine (type de donnée) depuis le portail web

Les types de mesures sont considérés comme une spécification de types de données et ils sont caractérisés par un nom, un signe, une unité et un type de représentation graphique telle que le montre la figure VIII.4.

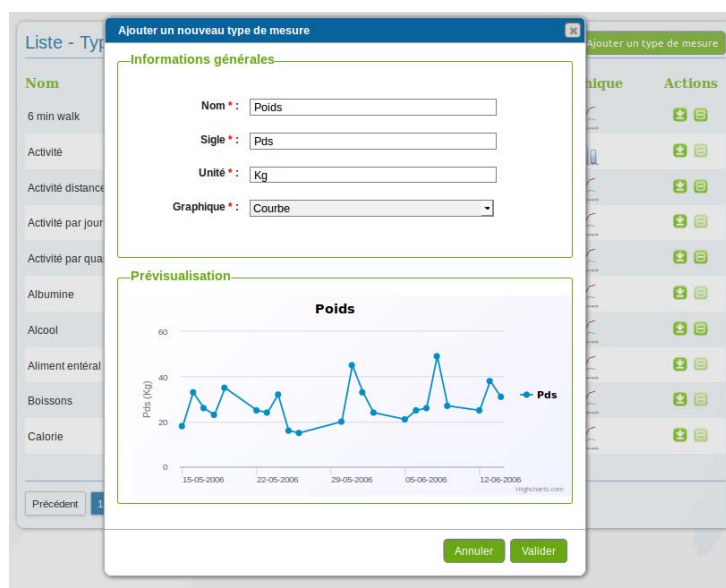


FIGURE VIII.4 – Ajout d’un type de mesure depuis le portail web

La figure VIII.4 montre la création d’une instance pour le type de mesure « Poids » où l’expert médical renseigne l’unité, le sigle et le type de graphique permettant de présenter les mesures poids collectées.

## VIII.4 Profil du patient

Dans cette section nous présentons la création de l'ontologie PPO (Patient Profile Ontology) et les étapes à suivre pour la constitution d'un profil à savoir la création d'un compte patient et le renseignement de l'historique médical.

### VIII.4.1 Construction de l'ontologie PPO

Le diagramme UML présentant le profil du patient illustré dans la partie contribution a été traduit dans le langage OWL en utilisant l'outil Protégé.

#### VIII.4.1.1 Création de classes

Cette étape consiste à créer les classes qui composent le profil du patient (voir figure VIII.5).

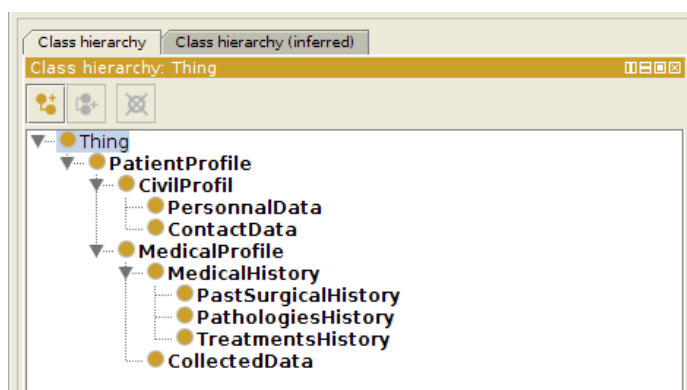


FIGURE VIII.5 – Création de classes pour l'ontologie profil patient

La relation d'héritage entre les différentes classes a été modélisée par la propriété OWL `<< rdfs :subClassOf >>` permettant la définition de liens hiérarchiques.

#### VIII.4.1.2 Création de propriétés

Une fois les classes créées, nous avons définis les propriétés d'objet qui relient les classes du profil entre elles et avec les autres classes des autres ontologies.



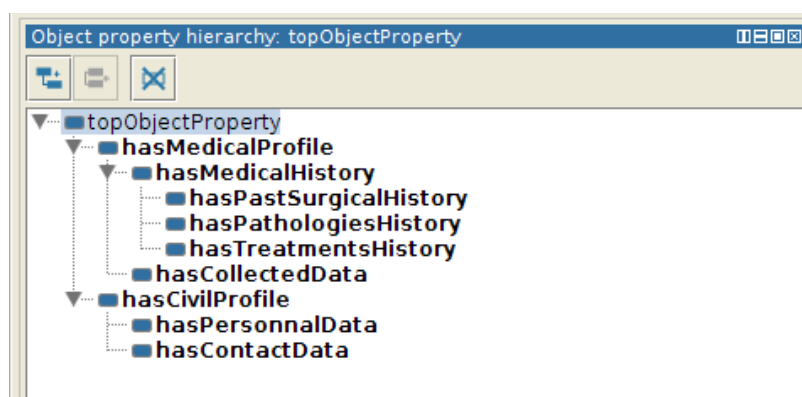


FIGURE VIII.6 – Propriétés d’objet de l’ontologie Profile patient

### VIII.4.2 Création de compte patient

La création de compte patient consiste à créer une instance de profil dans l’ontologie PPO où les données administratives du patient sont renseignées, des données telles que : les données personnelles, la liste des médecins suiveurs et les permissions de compte (voir figure VIII.7).

The screenshot shows a web form titled 'Création - Compte patient'. It has three tabs: 'Informations civiles' (selected), 'Médecins suiveurs', and 'Permissions'. The 'Informations civiles' tab is divided into two sections: 'Identité' and 'Coordonnées'.  
In the 'Identité' section, there are fields for: 'Sexe' (radio buttons for 'Homme' and 'Femme'), 'Nom' (text input: BENMIMOUNE), 'Prénom' (text input: Lamine), 'Date de naissance' (text input: 22/08/1959), 'Numéro de sécurité sociale' (text input), and 'NIP (IPP)' (text input: 555127). There is a placeholder for a profile picture.  
In the 'Coordonnées' section, there are fields for: 'Email', 'Téléphone fixe', 'Téléphone portable', 'Adresse postale' (text area), 'Code postal' (text input: 68200), and 'Ville' (text input: MULHOUSE).  
At the bottom, there is a progress indicator 'Étape 1/3' and two buttons: 'Annuler' and 'Étape suivante'.

FIGURE VIII.7 – Création de compte patient

### VIII.4.3 Renseignement des antécédents médicaux

Une fois que le compte du patient est créé, le médecin suiveur pourra renseigner l'historique médical du patient. Il est constitué des pathologies, des interventions chirurgicales et des traitements déjà pris ou en cours comme le montre la figure VIII.8.

Historique médical Exporter

Antécédents Traitements

Pathologies médicales

Ajouter une pathologie Filtrer par période

Type	Date début	Date fin	Actions
Diabète sucré insulino-dépendant, avec complications rénales			
Insuffisance rénale aiguë avec nécrose médullaire			

Précédent **1** Suivant

Interventions chirurgicales

Ajouter une intervention chirurgicale Filtrer par période

Type	Date	Etablissement	Service	Actions
intervention chirurgicale coeur ouvert		Hopitaux de strasbourg	Cardiologie	

Précédent **1** Suivant

FIGURE VIII.8 – Historique médical du patient

La liste des pathologies et les interventions chirurgicales sont définies dans l'ontologie CIM-10, tandis que la liste des traitements est définie dans l'ontologie ATC. Les antécédents médicaux sont sélectionnés à l'aide de la fonction d'auto-complétion qui permet de suggérer une liste de propositions en fonction de ce que le médecin cherche. Ceci permet de réaliser des recherches plus rapides et de mieux contrôler la saisie des données. La figure VIII.9 montre l'exemple de la recherche de la pathologie diabète et la liste des suggestions proposées.

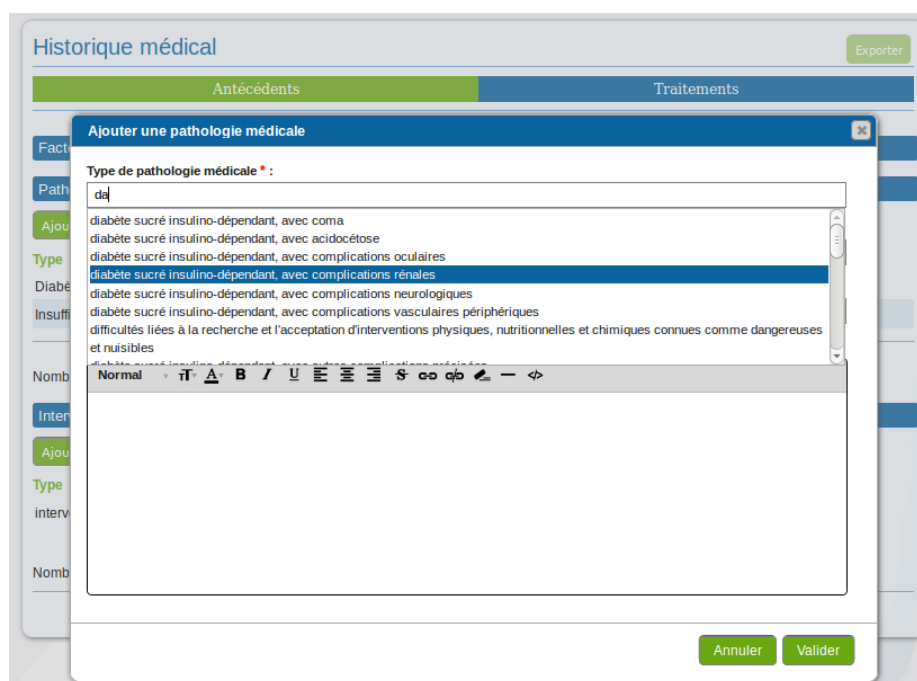


FIGURE VIII.9 – Chargement de la liste des pathologies depuis l'ontologie CIM-10.

## VIII.5 Outil d'acquisition de données

Nous présentons dans cette sections les différents modules qui composent l'outil d'acquisition des données. Nous présentons la construction des ontologies, le module de configuration des questionnaires, le module de collecte des données et le module d'historisation des interrogatoires.

### VIII.5.1 Construction des ontologies

Le développement des différents modules permettant l'acquisition des données nécessite la construction de deux ontologies décrites précédemment sous l'outil Protégé.

#### VIII.5.1.1 Création de l'ontologie Questionnaire

La construction de l'ontologie questionnaire a été faite en deux étapes avec la création des classes puis les créations des propriétés.

##### Création des classes

Les classes sont créées sous forme hiérarchique en utilisant la propriété « subClassOf » comme le montre la figure VIII.10.

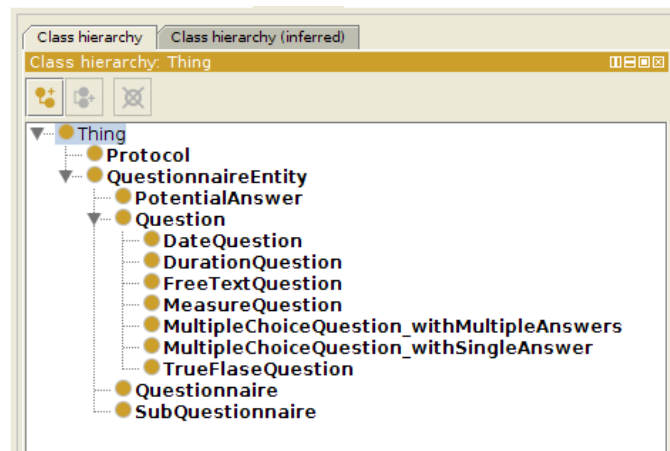


FIGURE VIII.10 – Création de classes de l'ontologie questionnaire sous Protégé.

### Création des propriétés

Une fois les classes créées, elles sont reliées entre elles et avec d'autres classes des autres ontologies par des propriétés d'objet comme le montre la figure VIII.11.

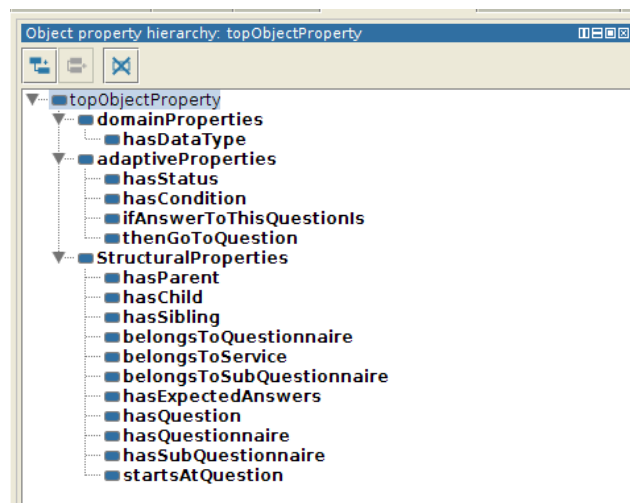


FIGURE VIII.11 – Création de propriétés de l'ontologie questionnaire sous Protégé

Nous avons regroupé les propriétés sous trois catégories :

- Les propriétés de domaine regroupent les propriétés qui relient les classes de l'ontologie questionnaires aux ontologies de domaine en l'occurrence la classe Question à la classe DataType de l'ontologie DTOnto.
- Les propriétés adaptatives regroupent les propriétés qui définissent le comportement adaptatif du questionnaire. Ces propriétés relient les classes du questionnaire avec d'autres ontologies. La propriété « hasStatus », par exemple, relie la classe Questionnaire à la classe « PatientStatus » dans l'ontologie du système afin d'adapter le questionnaire en fonction du statut du patient (contexte où le patient se retrouve).

- Les propriétés structurelles regroupent les propriétés qui définissent la structure du questionnaire. Ces propriétés relient uniquement les classes de l'ontologie questionnaire entre elles.

### VIII.5.1.2 Création de l'ontologie Interrogatoire

L'ontologie interrogatoire a été construite en deux étapes, les classes puis les propriétés.

#### Création des classes

Contrairement à l'ontologie questionnaire, les classes de l'ontologie interrogatoire ne sont pas créées dans une hiérarchie (voir figure VIII.12).

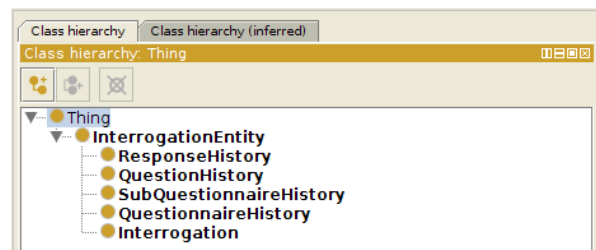


FIGURE VIII.12 – Création de classes de l'ontologie interrogatoire sous Protégé

#### Création des propriétés

La figure VIII.13 montre comment les classes créées sont reliées entre elles et avec d'autres classes des autres ontologies.

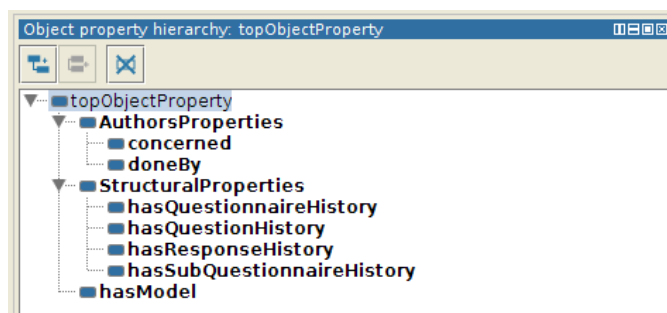


FIGURE VIII.13 – Création de propriété d'objet de l'ontologie interrogatoire sous Protégé

Nous avons regroupé les propriétés en deux groupes :

- Les propriétés d'auteurs qui relient l'interrogatoire à la personne qui a déroulé l'interrogatoire (professionnel de santé) et le patient interrogé.
- Les propriétés structurelles qui relient les classes entre elles.

Nous avons créé ainsi la propriété `hasModel` qui sert à relier les classes d'objets de l'ontologie interrogatoire à leurs modèles de classes dans l'ontologie questionnaire. Par exemple, la

classe `QuestionHistory` de `Interrogation Ontology (IO)` à comme modèle la classe `Question` de l'ontologie questionnaire. Cette propriété permet de créer un lien sémantique entre les deux ontologies et toute modification dans les attributs de la classe `Question` entraîne la modification des attributs de la classe `QuestionHistory`. Par exemple, si nous modifions l'ordre d'affichage pour une question dans l'ontologie `Questionnaire`, ceci entrainera une modification dans l'ordre de l'affichage de l'historique de la question correspondante.

### VIII.5.1.3 Construction de l'ontologie protocole de collecte

L'ontologie protocole sert à guider la collecte des données, elle a été construite en deux étapes, les classes puis les propriétés.

#### Création des classes

Les classes créées sont regroupées suivant les jours de la semaine, la fréquence et le moment de la journée (voir figure VIII.14).

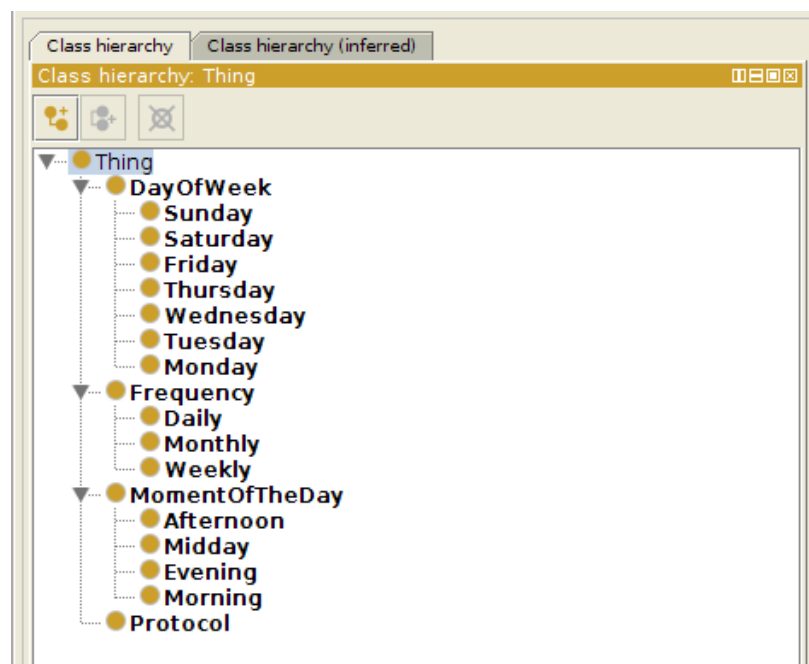


FIGURE VIII.14 – Création de classes de l'ontologie protocole sous Protégé

### Création des propriétés

La figure VIII.15 présente comment les classes créées sont reliées entre elles et avec d'autres classes des autres propriétés.

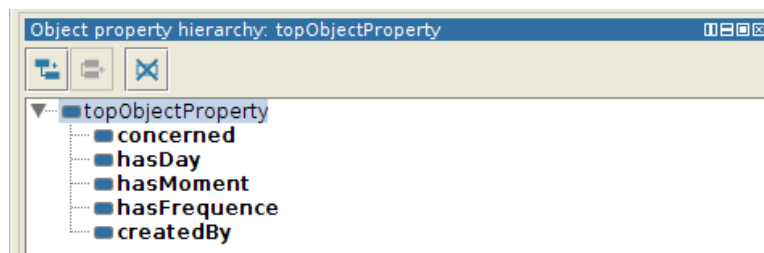


FIGURE VIII.15 – Création de propriétés d'objet de l'ontologie protocole sous Protégé

## VIII.5.2 Création de questionnaires via le portail web

Dans cette section nous décrivons les différentes interfaces associées à l'outil de l'acquisition des données médicales.

### VIII.5.2.1 Création de questionnaire

La création d'un questionnaire se fait depuis une interface web et elle se passe en trois étapes, décrites ci-dessous.

#### Informations générales

Cette étape consiste à renseigner des données générales sur le questionnaire, tel que le montre la figure VIII.16, avec le contexte médical (statut du patient) où le lieu d'utilisation du questionnaire, son titre, et une description pour décrire son rôle et son objectif. Une fréquence sur le nombre d'itération du questionnaire peut être définie dans cette étape si nous ne souhaitons pas l'attacher à un protocole de collecte.

The screenshot shows the 'Création de questionnaire' interface. At the top, there are three tabs: 'Informations générales' (highlighted in green), 'Sous-questionnaires', and 'Questions'. Below the tabs, the 'Informations générales' section contains the following fields:

- Statut \*: Domicile (dropdown menu)
- Titre \*: Hygiène de vie (text input)
- Fréquence \*: Plusieurs fois (dropdown menu)
- Description: Questionnaire sur l'hygiène de vie à poser quotidiennement. (text area)

At the bottom left, it says 'Étape 1/3'. At the bottom right, there are two buttons: 'Annuler' and 'Étape suivante'.

FIGURE VIII.16 – Étape 1 de la création d'un questionnaire via le portail web : renseignement des informations générales

### Sous questionnaires

Cette étape consiste à définir les sous-questionnaires qui servent à organiser les questionnaires en regroupant les questions sous différentes thématiques.

The screenshot shows the 'Création de questionnaire' interface, now on the 'Sous-questionnaires' tab. The 'Informations générales' tab is no longer highlighted.

The 'Nouveau sous-questionnaire' section contains:

- Nom \*: Qualité de sommeil (text input)
- Description: (empty text area)
- Ajouter (button)

Below this is the 'Liste des sous-questionnaires' section, which contains a table:

Nom	Description	Actions
Habitudes alimentaires	Repas et le respect de bonnes habitudes	⊖
Activité physique	Activité quotidienne : marche, sport, etc.	⊖

At the bottom left of the table, there are navigation buttons: 'Précédent', '1' (highlighted), and 'Suivant'. At the bottom right of the entire interface, there are three buttons: 'Annuler', 'Étape précédente', and 'Étape suivante'. At the bottom left, it says 'Étape 2/3'.

FIGURE VIII.17 – Étape 2 de la création d'un questionnaire via le portail web : création de sous-questionnaires

La figure VIII.17 montre la création de trois sous-questionnaires pour un questionnaire sur l'hygiène de vie. Les sous questionnaires créés servent à regrouper les questions en plusieurs thématiques à savoir : les habitudes alimentaires, l'activité physique, la qualité



de sommeil, etc.

## Questions

Cette étape consiste à créer les questions et à définir les relations sémantiques entre les questions et les comportements adaptatifs associés. La figure VIII.18 montre la création de deux questions sous le sous questionnaire Qualité de sommeil. La première question : « Hier, vous vous êtes endormi(e) facilement ? » a comme réponses potentielles « Oui /Non ». Alors que la deuxième question : « Votre sommeil était de bonne qualité ? » a comme réponses potentielles « Pas du tout, un peu et tout à fait » dont une valeur (1, 2 et 3) est assignée à chaque réponse potentielle.

FIGURE VIII.18 – Étape 3 de la création d'un questionnaire via le portail web : création des questions

Ces deux questions montrent le comportement adaptatif du questionnaire en reliant la deuxième question à une question parent (mère) (la première question) et en définissant une condition « si la réponse = Oui ». Ceci permet de poser la deuxième question « Votre sommeil était de bonne qualité ? » uniquement lorsque la réponse à la première question « Hier, vous vous êtes endormi(e) facilement ? » est « Oui ». Cet exemple illustre ainsi l'utilisation de l'ontologie de domaine DTonto où chaque question créée est reliée au concept correspondant afin de donner une signification à la question et à la donnée collectée par la suite : la deuxième question est reliée au concept « Sommeil de bonne qualité » de l'ontologie DTonto.

### VIII.5.2.2 Définition de protocole questionnaire

Un premier niveau de protocole est défini dans la première étape de la création du questionnaire (voir figure VIII.18) où le questionnaire peut avoir une fréquence d'utilisation. La création d'un protocole pour des questionnaires est faite indépendamment de la création des questionnaires. Le protocole relatif à un questionnaire permet d'organiser au mieux la collecte de données en fonction de son contexte d'utilisation. Ceci est très utile dans le cadre d'un suivi médical où les questionnaires sont posés de manière automatique. Dans ce cas, le protocole est défini par les médecins.

Nous avons implémenté la configuration des protocoles dans le portail web de la plateforme comme le montre la figure VIII.19.

Ajouter un nouveau protocole ?

**Informations générales**

Saisie Par : Medical ADMIN

Saisie le : 27/07/2016 11:24:59

Concerne \* : Hygiène de vie

**Occurrence**

Fréquence \* :

- Une seule fois
- Tous les jours
- Tous les 2 jours
- Lundi  Mardi  Mercredi  Jeudi  Vendredi  Samedi  Dimanche
- Toutes les semaines
- Tous les mois

Moment :

- Au lever
- Le matin
- Le midi
- L'après-midi
- Le soir
- Au coucher

Début \* : 27/07/2016

Fin \* : 27/10/2016

Annuler Valider

FIGURE VIII.19 – Configuration de protocole pour le questionnaire

Les experts médicaux peuvent rattacher un questionnaire à un protocole et définir la fréquence et les moments où il doit être utilisé. Ils peuvent ainsi définir une durée de validité pour une collecte de données. Ceci est avantageux dans le cadre d'un suivi médical à domicile où le médecin souhaite, par exemple, suivre l'évolution de l'état de santé et l'hygiène de vie de son patient pendant une période bien définie.

### VIII.5.3 Module de collecte de données

Dans cette section nous présentons le module de collecte de données. Ce module a pour but de gérer la collecte de donnée. Il permet de

- Poser les questions suivant le contexte du patient ;
- Adapter les questions suivant les réponses apportées ;
- Recueillir les données collectées et les stocker dans la base de données.

Nous avons implémenté le module de collecte de données sur le portail web ainsi que l'application mobile associée.

#### VIII.5.3.1 Collecte de données via le portail web

Le module de la collecte de donnée via le portail web de la plateforme est exclusivement utilisé par les professionnels de santé. Il peut être utilisé par les médecins dans le cas d'une consultation médicale. Les infirmières vont l'utiliser lorsqu'il s'agit d'une retranscrire de données extraites des dossiers médicaux ou pour renseigner des données qui ne sont pas demandées directement au patient, comme les données relevées lors de la phase peropératoire qui concernent le déroulement de l'anesthésie et l'opération chirurgicale. En fonction du contexte, c'est-à-dire en fonction du statut du patient, les questionnaires sont proposés à l'écran (voir la figure VIII.20).

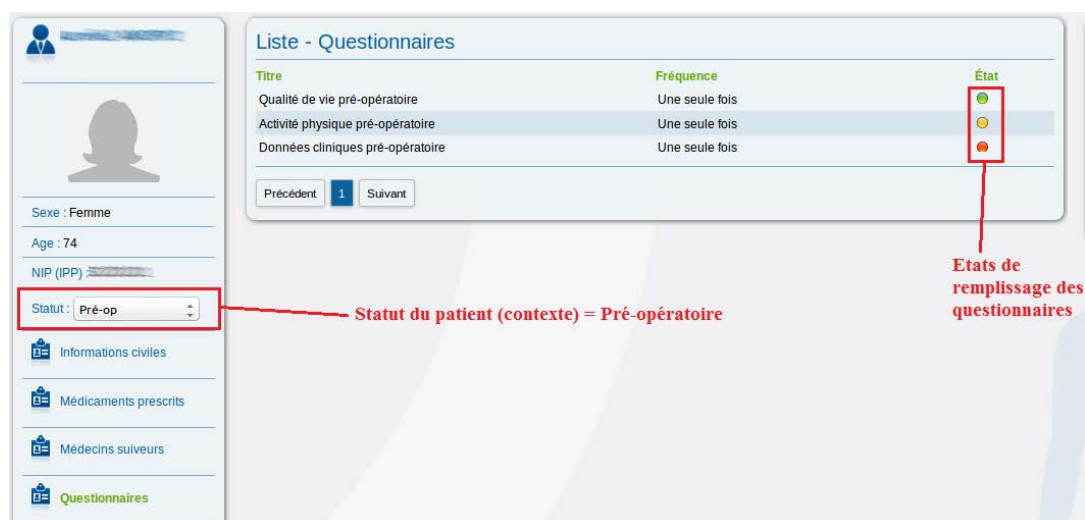


FIGURE VIII.20 – Affichage des questionnaires en fonction du statut du patient

Dans l'exemple illustré par la figure VIII.20, le patient se trouve en phase préopératoire, ce qui représente son contexte. Le module propose donc que des questionnaires en lien avec cette phase.

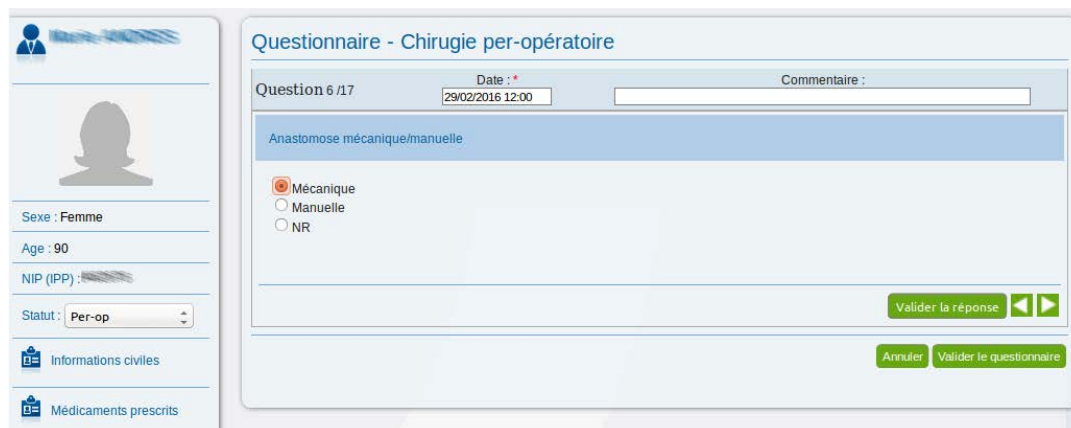
L'état de remplissage des questionnaires est aussi affiché afin de permettre au professionnel de santé de savoir si le questionnaire était rempli ou pas. Pour la collecte de données nous

avons implémenté les deux modes de collecte présentés précédemment, à savoir le mode de collecte séquentiel et le mode instantané.

### Mode de collecte de données séquentiel

Le mode de collecte de données séquentiel est utile pour guider la collecte en suivant un enchaînement dans l’affichage des questions. Ces dernières sont affichées les unes après les autres dans l’ordre défini par les experts médicaux. Généralement, ce mode de collecte est utilisé sur le portail web par les médecins pour l’élaboration d’un diagnostic, l’enchaînement logique des questions va les guider dans leur réflexion et leur raisonnement médical.

La figure VIII.21 montre un exemple de déroulement de la collecte des données en mode séquentiel depuis le portail web.



The screenshot displays a web interface for a questionnaire titled "Questionnaire - Chirurgie per-opératoire". On the left, a user profile sidebar shows a silhouette and details: "Sexe : Femme", "Age : 90", "NIP (IPP) : [redacted]", and "Statut : Per-op" with a dropdown arrow. Below this are links for "Informations civiles" and "Médicaments prescrits". The main content area shows "Question 6 /17" with a date field set to "29/02/2016 12:00" and a "Commentaire :" field. The question text is "Anastomose mécanique/manuelle". There are three radio button options: "Mécanique" (selected), "Manuelle", and "NR". At the bottom right of the question area are buttons for "Valider la réponse" (with left and right arrows) and "Annuler" and "Valider le questionnaire".

FIGURE VIII.21 – Interrogatoire en mode séquentiel via le portail web de la plateforme

### Mode de collecte de données instantané

Le mode de collecte de données instantané permet d’accéder à l’ensemble des questions. Il est généralement utilisé par les infirmières pour retranscrire des données recueillies soit auprès patients soit auprès des cliniciens. Ces données peuvent concerner, par exemple, le déroulement de l’anesthésie et la chirurgie dans le cadre d’un suivi opératoire (voir la figure VIII.22).

Question	Réponse(s)	Actions
Laprosopie	Non	✖
Laparotomie	Oui	✖
Laparoconversion	Oui	✖
Type op		✖
Type anastomose		✖
Anastomose mécanique/manuelle		✖
Distance anastomose/sphincter		✖
Stomie?		✖
Geste associé		✖
Contention élastique		✖

FIGURE VIII.22 – Interrogatoire en mode instantané via le portail web de la plateforme

### VIII.5.3.2 Collecte de données via l'application mobile

Nous avons implémenté un module de collecte de données sur l'application mobile de la plateforme E-care. Il permet la collecte de données depuis un dispositif mobile (tablette ou smartphone). La figure VIII.23 montre le chargement des questions en fonction du statut du patient (contexte) et l'état de remplissage des interrogatoires de chaque questionnaire.

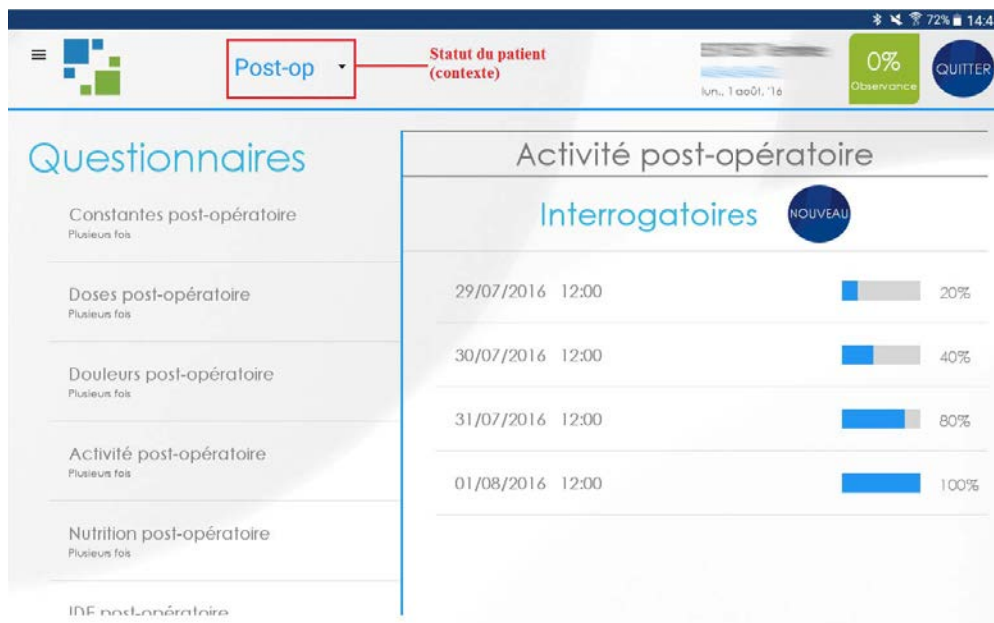


FIGURE VIII.23 – Chargement des questionnaires en fonction du statut du patient sur l'application mobile de la plateforme.

La collecte depuis l'application mobile est utilisée soit par les patients eux même soit par

les professionnels de santé selon les deux modes proposés précédemment.

### Mode de collecte de données séquentiel

Ce mode de collecte de données est utilisé sur l'application mobile par les patients dans le cadre d'un suivi à domicile où le patient dispose d'un dispositif mobile tablette ou smartphone (voir la figure VIII.24).

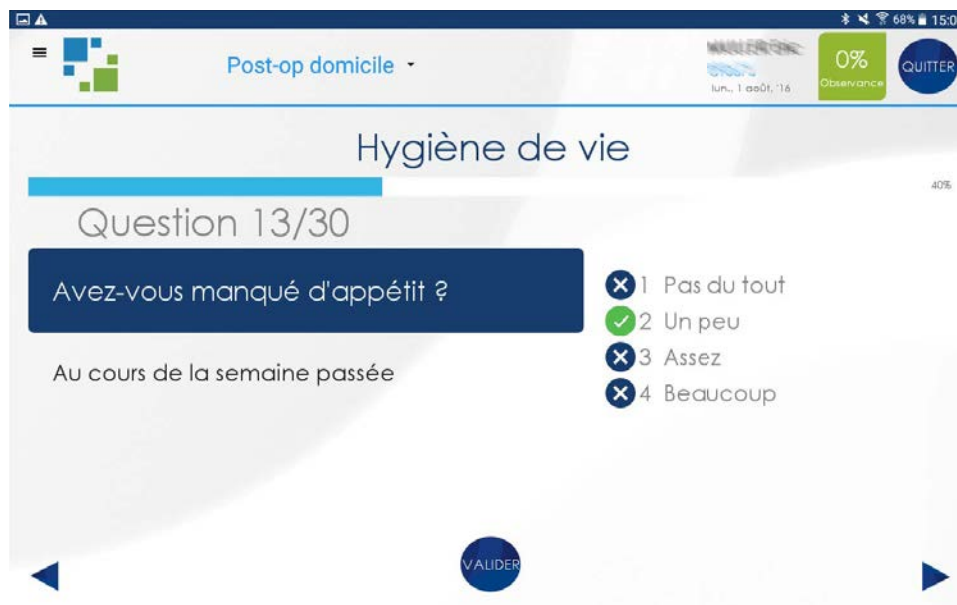


FIGURE VIII.24 – Interrogatoire en mode séquentiel via l'application mobile de la plateforme

### Mode de collecte de données instantané

Ce modèle de collecte de données est utilisé sur par les infirmières pour le recueil de données quotidiennes dans le cadre d'un suivi opératoire des patients où elles sont amenées à passer chez les patients du service et les interroger individuellement. L'avantage étant de présenter toutes les questions et les réponses sous forme de liste afin de permettre à l'infirmière d'avoir un aperçu exhaustif sur les données recueillies et les données qui reste à recueillir (voir la figure VIII.25).

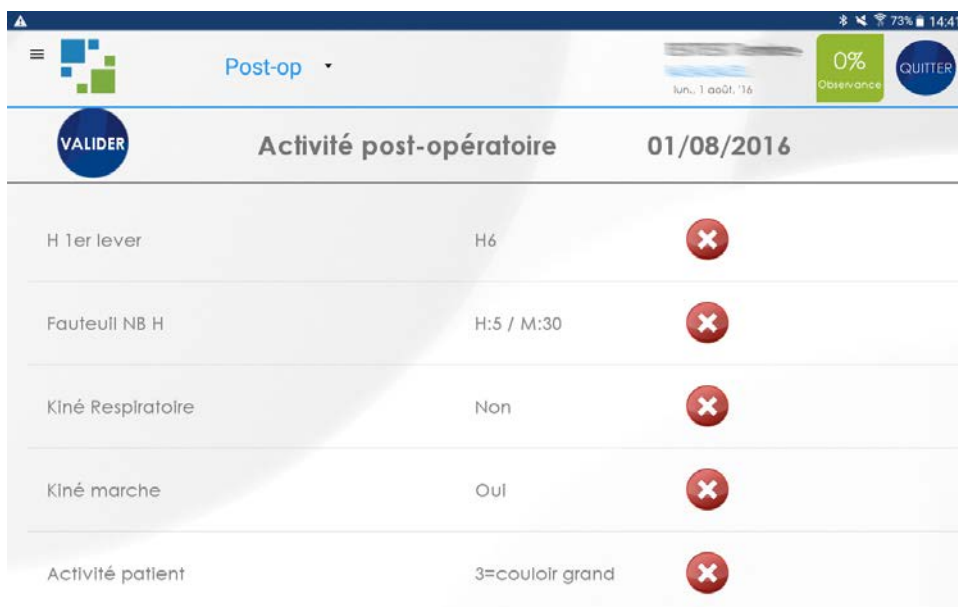


FIGURE VIII.25 – Interrogatoire en mode instantané via l’application mobile de la plateforme

### VIII.5.4 Module de l’historique des interrogatoires

Ce module est accessible depuis le portail web de la plateforme, il permet l’accès à tout l’historique des interrogatoires réalisés pour chaque patient comme le montre la figure VIII.26.

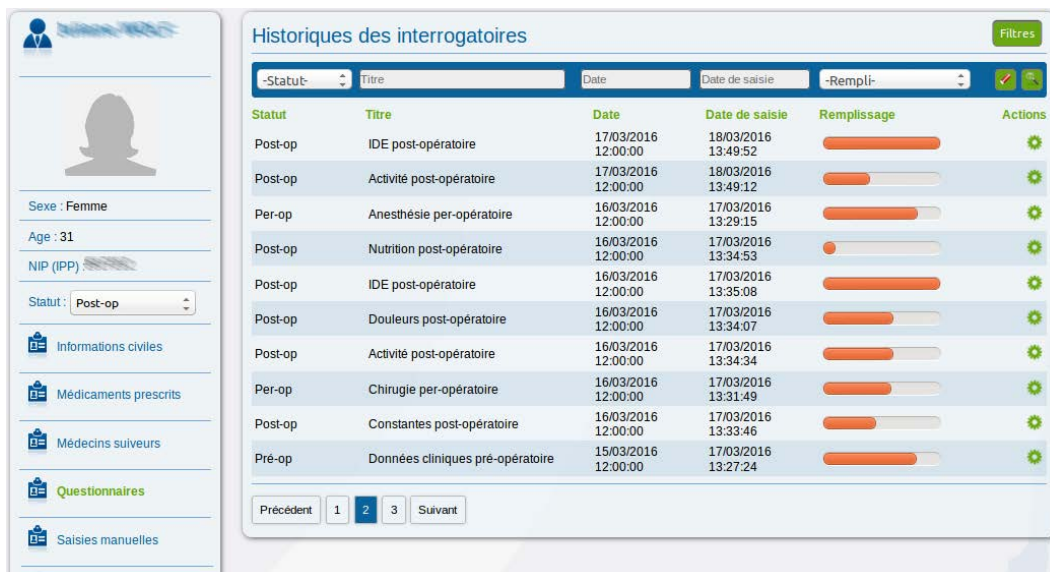


FIGURE VIII.26 – Liste de l’historiques des interrogatoires par patient



L'historique des interrogatoires regroupe les données collectées à travers les questionnaires c'est-à-dire les réponses aux questions posés, le taux de remplissage du questionnaire, la date de l'interrogatoire et le nom du professionnel de santé qui a déroulé l'interrogatoire. La figure VIII.27 montre l'affichage de l'historique d'un interrogatoire.

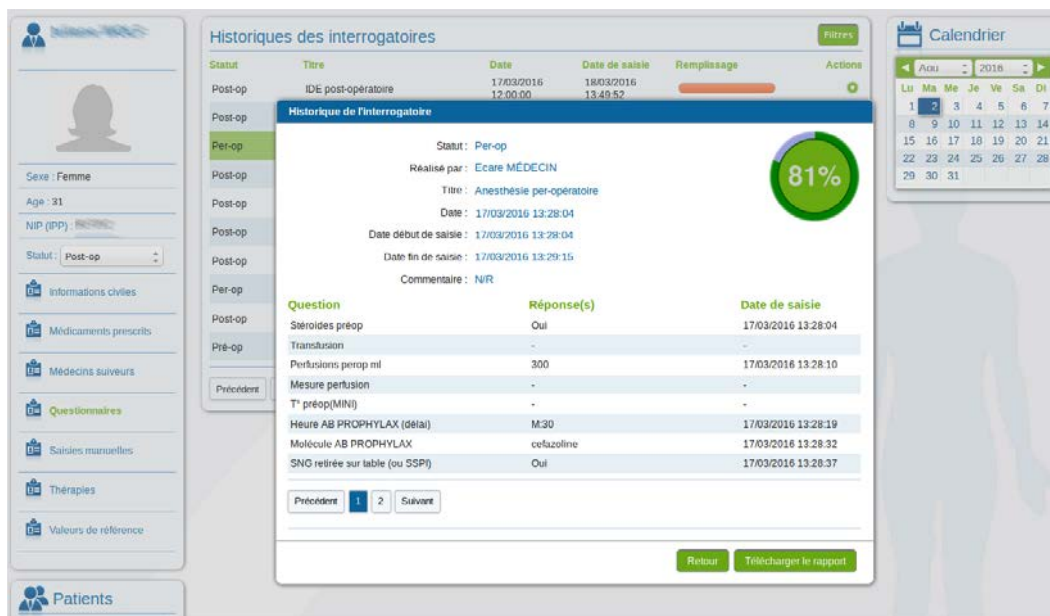


FIGURE VIII.27 – L'affichage d'un historique d'interrogatoire

## VIII.6 Moteur de raisonnement à base de règles

Dans cette section nous détaillons l'implémentation du moteur de raisonnement à base de règles et son utilisation pour l'exploitation des connaissances des experts médicaux sous forme de règles.

### VIII.6.1 Gestion des règles d'inférence

Pour rendre la plateforme plus générique, nous avons implémenté un module de gestion de règles d'inférence accessible depuis le portail web de la plateforme. Ce module permet la gestion des règles utilisées dans la plateforme. Il permet de coder les règles dans le portail à l'aide d'un éditeur de code comme le montre la figure VIII.28.

Nous avons choisi d'exprimer les règles d'inférence en langage JavaScript afin d'offrir plus d'expressivité par rapport au SWRL. Le SPARQL est ainsi utilisé comme langage de requête pour accéder aux données de la base de faits.

Une fois les règles écrites à l'aide de l'éditeur et validées, elles seront stockées dans la base de connaissances et ensuite chargées et interprétées par le moteur de raisonnement.



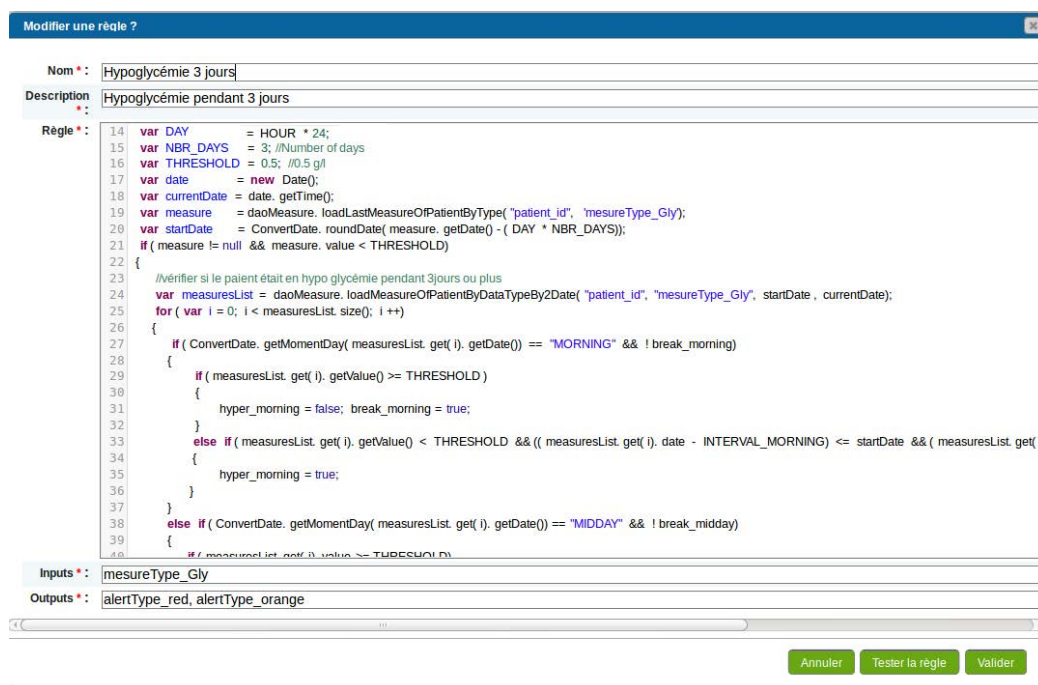


FIGURE VIII.28 – Éditeur de règle via le portail web

### VIII.6.2 Exécution des règles

L'exécution des règles se fait par le moteur d'inférence et ce pour chaque nouvelle donnée collectée. Nous avons implémenté le moteur d'inférence dans un web service distant afin de ne pas ralentir le processus de collecte de données.

À chaque donnée collectée par le portail web ou l'application mobile, le moteur d'inférence est notifié par le biais du web service via une requête HTTP dans un format JSON (JavaScript Object Notation).

```
{
  "patientId" : "Public1460389268059"
  "dataTypeId" : "mesureType_Pds",
  "dataId" : "mesure_1461939409545",
  "value" : "77.0"
}
```

FIGURE VIII.29 – Exemple d'une notification encapsulée dans un objet JSON

La figure VIII.29 illustre un exemple de notification envoyée vers le web service concernant la mesure de poids d'un patient donné.

Une fois la notification reçue, le moteur d'inférence procède à la sélection de l'ensemble des règles susceptibles d'être exécutées et ceci à partir de l'identifiant du déclencheur, ici

«*measureType\_Pds*» envoyé dans la notification. Par la suite, le moteur exécute les règles sélectionnées et enregistre les données inférées dans la base de faits.

Dans l'exemple illustré dans la figure VIII.29, si le système dispose d'une règle pour calculer l'IMC (sachant que pour calculer l'IMC le moteur aura besoin de la mesure du poids et de la taille du patient), cette dernière sera sélectionnée à la réception de la notification. Le moteur d'inférence charge ensuite la taille du patient depuis la base, étant donnée que la taille est considérée comme le deuxième déclencheur pour l'exécution de la règle, si la taille du patient existe dans la base dans ce cas le moteur exécute la règle et enregistre la donnée inférée (l'IMC) dans la base, sinon il se met en attente d'une nouvelle notification.

### VIII.6.3 Inférence des données à partir des règles

Les données inférées par le moteur d'inférence peuvent prendre trois natures différentes à savoir des indicateurs, des alertes et des recommandations.

#### VIII.6.3.1 Indicateurs

Les indicateurs générés par le moteur d'inférence sont affichés sur le portail web pour assister les professionnels de santé dans leur prise de décision. Plusieurs indicateurs peuvent être générés pour le suivi des patients, pour l'amélioration du service rendu aux patients et à la qualité de la pratique médicale.

La figure VIII.30, montre l'affichage des indicateurs médicaux sur le portail web de la plateforme.

The screenshot displays a medical dashboard with the following elements:

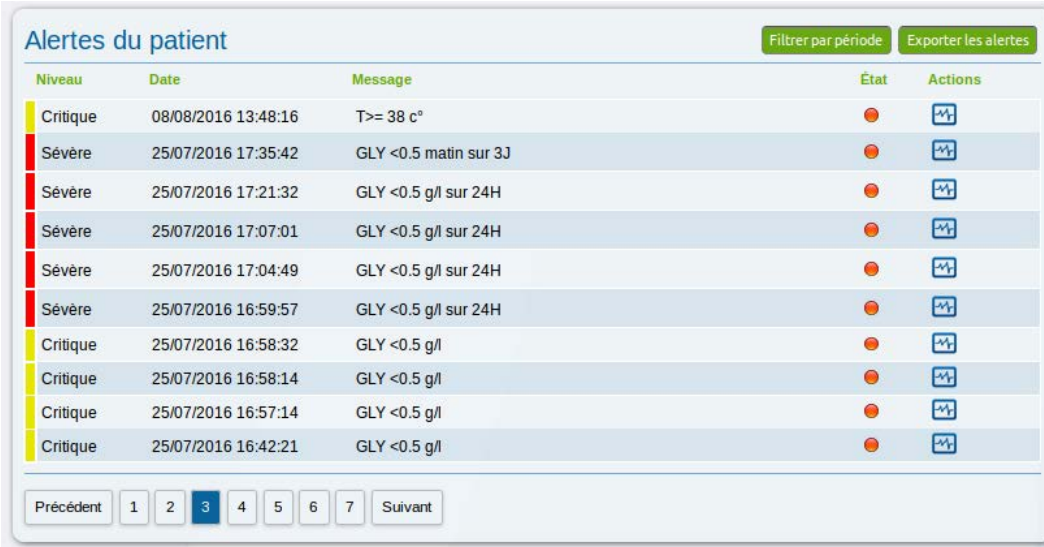
- Indicateurs (Indicators):** A sidebar on the left shows four indicators: "Jour : 13", "CCI : 26", "ByPass : 14", and "ERAS : 44,12 %". The ERAS indicator is highlighted with a red box.
- Mesures du patient (Patient Measurements):** A central table lists various measurements with columns for "Date" and "Type".
- Détails de calcul de score (Score Calculation Details):** A pop-up window on the right provides a breakdown of the ERAS score calculation. It lists ten parameters with their status and ALORS score:
  - Paramètre [1] manquant : Test UP&GO = NR ALORS score = 0
  - Paramètre [2] manquant : Test 6min walk = NR ALORS score = 0
  - Paramètre [3] manquant : Test vems = NR ALORS score = 0
  - Paramètre [4] manquant : Test grip = NR ALORS score = 0
  - Paramètre [5] manquant : Activité preop (bracelet) = NR ALORS score = 0
  - Paramètre [6] OK : Qualité de vie = 6.0 ALORS score = 1
  - Paramètre [7] OK : Préparation Colique = Oui ALORS score = 1
  - Paramètre [8] OK : Jéjune preop = 120 minutes ALORS score = 1
  - Paramètre [9] OK : Jus de pomme = Oui ALORS score = 1
  - Paramètre[10] Non OK : Délai antibioprophylaxie = 10 minutes ALORS score

FIGURE VIII.30 – Affichage des indicateurs et détail de calcul de score pour l'indicateur ERAS

Le professionnel de santé peut afficher le détail du calcul du score de l'indicateur afin de lui permettre de comprendre le déroulement de la règle. La figure montre le détail sur le calcul d'un indicateur utilisé dans le cadre du suivi opératoire.

### VIII.6.3.2 Alertes

Les alertes générées par le moteur d'inférence sont affichées sur le portail web de la plateforme comme le montre la figure VIII.31.



Niveau	Date	Message	État	Actions
Critique	08/08/2016 13:48:16	T >= 38 c°	●	[Icon]
Sévère	25/07/2016 17:35:42	GLY <0.5 matin sur 3J	●	[Icon]
Sévère	25/07/2016 17:21:32	GLY <0.5 g/l sur 24H	●	[Icon]
Sévère	25/07/2016 17:07:01	GLY <0.5 g/l sur 24H	●	[Icon]
Sévère	25/07/2016 17:04:49	GLY <0.5 g/l sur 24H	●	[Icon]
Sévère	25/07/2016 16:59:57	GLY <0.5 g/l sur 24H	●	[Icon]
Critique	25/07/2016 16:58:32	GLY <0.5 g/l	●	[Icon]
Critique	25/07/2016 16:58:14	GLY <0.5 g/l	●	[Icon]
Critique	25/07/2016 16:57:14	GLY <0.5 g/l	●	[Icon]
Critique	25/07/2016 16:42:21	GLY <0.5 g/l	●	[Icon]

Précédent 1 2 3 4 5 6 7 Suivant

FIGURE VIII.31 – Affichage des alertes sur le portail web

### VIII.6.3.3 Recommandations

Les recommandations générées par le moteur d'inférence sont affichées sur l'application mobile pour le patient concerné.

La figure VIII.32 montre la présentation des recommandations sur l'hygiène de vie sous forme d'une liste.

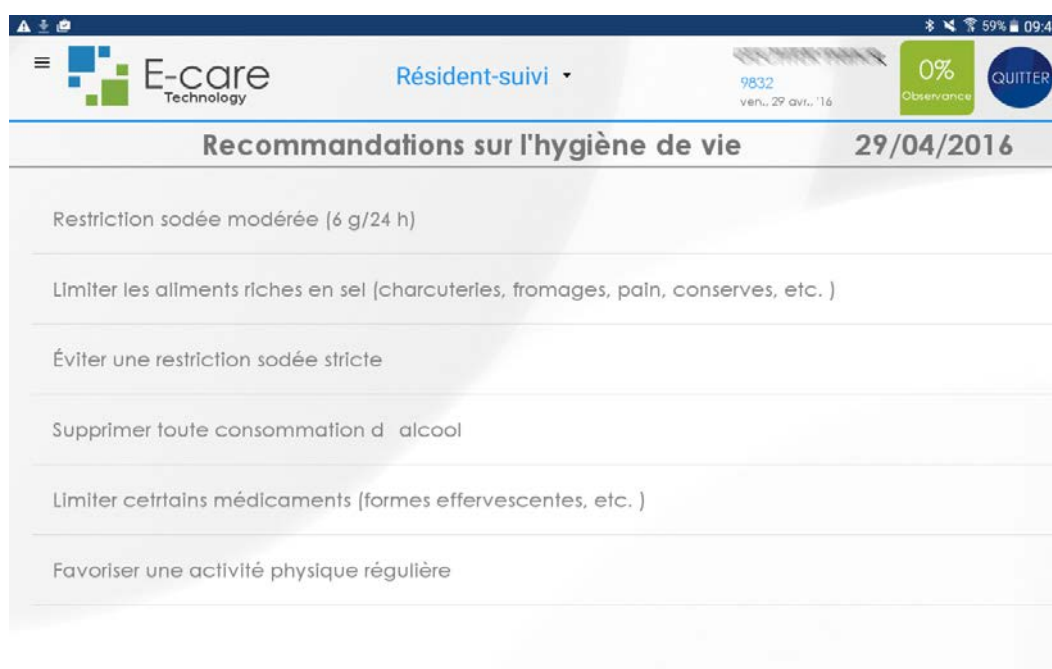


FIGURE VIII.32 – Affichage des recommandations sur l’application mobile de la plateforme

## VIII.7 Moteur de raisonnement à base de cas

Dans cette section nous présentons l’implémentation du moteur de raisonnement à base de cas et son utilisation pour l’exploitation de la connaissance pratique des experts médicaux, à savoir l’expérience clinique, pour la recommandation des actes médicaux (thérapies médicales).

### VIII.7.1 Gestion des thérapies médicales

Nous avons développé un module pour la gestion des actes médicaux pour la prise en charge thérapeutique. Ce module permet de gérer les thérapies médicales proposées pour chaque patient au cours de son suivi médical. Le module est accessible depuis le portail web de la plateforme et permet au professionnel de santé d’ajouter, de modifier et de supprimer des thérapies pour un patient donné (voir la figure VIII.33).

Afin que les thérapies soient exploitées par le moteur de raisonnement et pour faire des recommandations, ces dernières sont taguées par une liste de mots clés qui correspondent aux concepts représentés dans l’ontologie DTonto. Ceci permet de calculer le score de pertinence à partir des similarités tel qu’il a été expliqué dans la partie contribution.

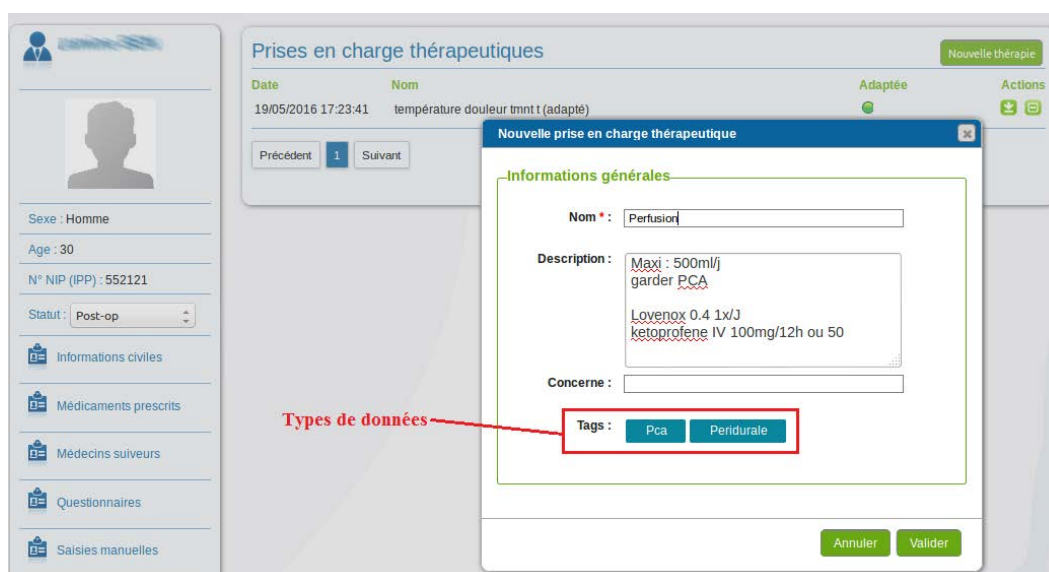


FIGURE VIII.33 – Ajout d'une nouvelle thérapie

## VIII.7.2 Recherche des cas similaires

Nous avons implémenté un module pour la recherche des cas les plus similaires et les recommandations afin d'expérimenter le moteur de raisonnement à base de cas. Ce module est accessible depuis le portail web de la plateforme et il permet de rechercher des profils de patients similaires au patient en cours et, par la suite, de recommander les thérapies les plus pertinentes.

### VIII.7.2.1 Recherche des profils

La recherche des profils les plus similaires constitue la première étape dans un raisonnement à base de cas où la sélection des profils est basée sur les scores des similarités calculés. Le médecin peut affiner les résultats de la recherche en définissant des critères de recherche, les plus pertinents en fonction du patient. Il sélectionne les types de données qui seront pris en compte dans le calcul de similarités global. Puis, pour chaque type de donnée, le médecin peut attribuer un poids désignant son degré d'importance et une similarité minimale à retenir. La figure VIII.34 montre la recherche des profils similaires et l'affichage du détail des similarités sous format de graphe.



FIGURE VIII.34 – Recherche des profils les plus similaires

### VIII.7.2.2 Recommandations des thérapies

Une fois les profils les plus similaires retrouvés, le moteur recommande la liste des thérapies les plus pertinentes associées à chaque profil. Les thérapies sont ordonnées selon leurs scores de pertinence, calculés en fonction des similarités, et les tags liés à chaque thérapie. Seules les thérapies pertinentes sont retenues et présentées à l'écran.

La figure VIII.35 montre la recommandation des thérapies associées à un profil similaire sélectionné.

The screenshot displays a medical software interface with the following components:

- Left Panel (Patient Profile):**
  - Sexe : Femme
  - Age : 60
  - NIP (IPP) : 1928290
  - Statut : Post-op
  - Informations civiles
  - Médicaments prescrits
  - Médecins suiveurs
  - Questionnaires
  - Saisies manuelles
  - Valeurs de référence
- Similarités Table:**

Classement	Patient	Similarité	Actions
1		82.343 %	
2		79.209 %	
3		78.951 %	
4		77.930 %	
5		77.628 %	
6		77.523 %	
7		76.561 %	
8		76.516 %	
9		75.542 %	
10		74.909 %	
- Thérapies recommandées Table (highlighted):**

Date	Titre	Pertinence	Évaluation	Actions
10/08/2016 15:35:25	Anti nausées	73.647 %	-	
10/08/2016 15:34:20	Antalgiques anti douleur	71.145 %	-	
- Détails de la similarité:** A radar chart showing similarity scores for 'Poids (2U)', 'Douleur (10U)', and 'Fréquence cardiaque (10U)'. The score for 'Poids' is 50.
- Criteria (Critères) Panel:**
  - Dosage hb: Poids 1, Sim(min): 0 %
  - Douleur: Poids 5, Sim(min): 0 %
  - Fréquence cardiaque: Poids 2, Sim(min): 0 %
  - Palier antalgique: Poids 1, Sim(min): 0 %
  - Buttons: Du J : 0 au J : 6, Appliquer
- Filtres Panel:**
  - Poids (82.6%)
  - Douleur (66.4%)
  - Pression artérielle systolique (80.0%)
  - Saturation en oxygène (82.2%)
  - Température (75.9%)
  - Fréquence cardiaque (78.7%)

FIGURE VIII.35 – Recommandation de thérapies

Toute thérapie recommandée peut être réutilisée comme solution pour le cas cible (patient en cours) si le médecin la juge pertinente. Sinon le médecin peut l'adapter en modifiant sa description. Le médecin peut évaluer ainsi la thérapie recommandée en attribuant une note pour juger son degré de pertinence par rapport au cas cible (voir la figure VIII.36).



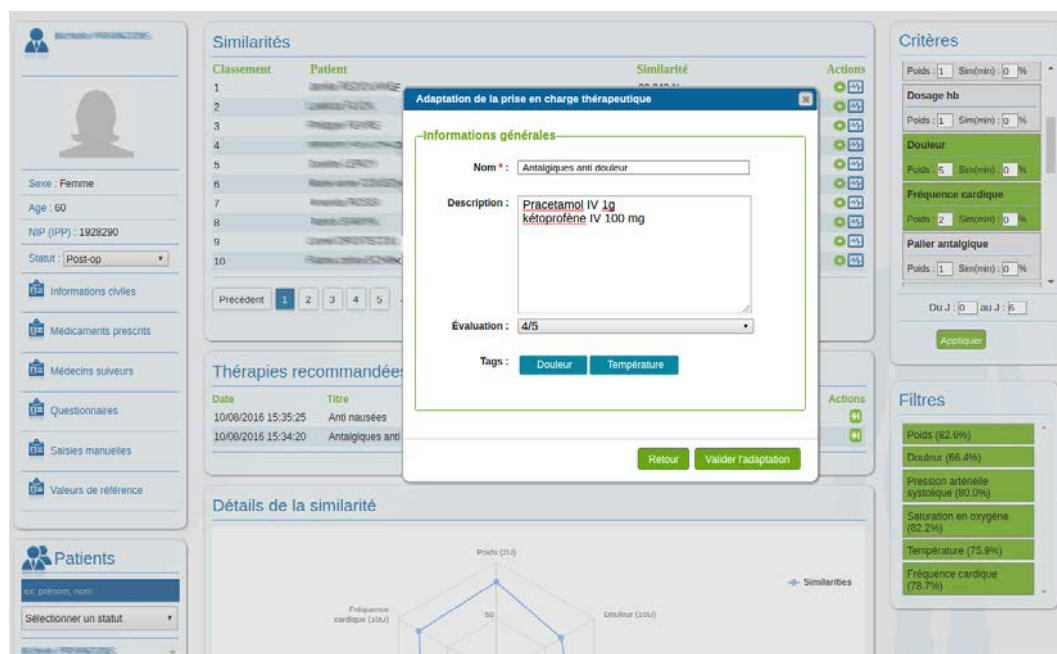


FIGURE VIII.36 – Adaptation d’une thérapie recommandée

## VIII.8 Conclusion

Nous avons présenté dans ce chapitre l’implémentation de nos travaux de recherche au sein de la plateforme E-care. Nous avons donné un aperçu de quelques composants, développés dans le cadre de cette thèse, qui concernent notre contribution. Nous avons implémenté principalement trois composants :

- Un module permettant aux experts médicaux d’introduire des connaissances médicales de manière simple depuis des interfaces web ;
- Un module permettant la création des profils de patients et le renseignement de leurs historiques médicaux depuis des interfaces web ;
- Un outil en ligne et une application mobile pour la collecte des données, basés sur un mécanisme de question/réponse basé sur l’utilisation des ontologies, a été implémenté ;
- Un moteur de raisonnement mixte à base de connaissances permet d’exploiter les guides de bonnes pratiques sous formes de règles d’inférence pour générer des indicateurs, des alertes et des recommandations. Il permet ainsi d’exploiter l’expérience clinique des médecins sous forme de cas pour faire des recommandations de thérapies.

Dans le prochain chapitre nous décrivons le déroulement de l’expérimentation de la plateforme E-care dans un environnement clinique et nous présentons ainsi les résultats de son évaluation.





# IX

## EXPÉRIMENTATION ET ÉVALUATION

Dans ce chapitre, nous allons présenter l'expérimentation de notre approche dans la plate-forme E-care dans le cadre de plusieurs projets de suivi médical.

### Sommaire

---

<b>IX.1 Introduction</b> . . . . .	<b>135</b>
<b>IX.2 Expérimentation</b> . . . . .	<b>136</b>
IX.2.1 Projet Repose . . . . .	136
IX.2.2 Projet Brave . . . . .	138
IX.2.3 Projet Chariot EHPAD . . . . .	140
<b>IX.3 Évaluation</b> . . . . .	<b>142</b>
IX.3.1 Collecte de données . . . . .	142
IX.3.2 Raisonnement à base de cas . . . . .	143
IX.3.3 Raisonnement à base de règles . . . . .	146
<b>IX.4 Conclusion</b> . . . . .	<b>146</b>

---

### IX.1 Introduction

Dans les chapitres précédents nous avons présenté la conception et l'implémentation de notre approche dans la plate-forme E-care dédiée au suivi médical. Nous avons mis l'accent sur le fait que notre approche est ainsi ouverte et indépendante du domaine ou du mode d'intervention défini. Cette approche permet à la plateforme E-care d'être utilisée pour le suivi de patients atteints de pathologies différentes et selon des modes d'interventions différents.

Dans ce chapitre nous présentons les projets de suivi médical dans lesquels nous avons expérimenté la plate-forme et nous mettons l'accent sur l'expérimentation des différents modules développés et qui sont en lien avec nos travaux de recherches.

Dans ce chapitre, nous allons présenter l'expérimentation de nos travaux de recherches et présenter le cadre dans lequel l'expérimentation a été menée dans les différents projets. Nous aborderons ensuite l'évaluation de notre approche.

## IX.2 Expérimentation

Nous avons expérimenté la plateforme dans plusieurs projets de suivi médical.

### IX.2.1 Projet Repose

Le projet Repose est un projet mené par l'entreprise Newel en collaboration avec le service de la Chirurgie Générale et Digestive l'hôpital Hautepierre en Alsace sous la direction du professeur ROHR. Le but du projet étant de proposer une plateforme dédiée à la réhabilitation améliorée permettant de réaliser un suivi médical de patients avant, pendant et après une opération chirurgicale dans le but d'améliorer le chemin clinique.

L'expérimentation a débuté, dans une première phase, en octobre 2015 pour une durée de 15 jours. Elle a permis de dresser un premier état des lieux, de tester les différents modules proposés, d'améliorer l'ergonomie, de détecter les failles et de valider les choix technologiques et médicaux. Suite à cette première phase, la plateforme a été améliorée et reconfigurée afin de répondre au besoin des chirurgiens.

Une deuxième phase a été lancée en Novembre 2015 et qui s'étale sur une période de deux années. Durant les six premiers mois de l'expérimentation plus de 270 patients ont été inclus avec :

- Une moyenne d'âge de 59 ans ;
- Une durée moyenne de l'hospitalisation de 10 jours ;

70 % des patients ont été opérés du cancer dont 34% au niveau foie et 55% au niveau de l'appareil digestif et 11% ont subi d'autres types d'opération chirurgicales.

#### IX.2.1.1 Définition de contexte du patient

Pour la définition du contexte des patients avant, pendant et après l'opération chirurgicale, quatre statuts ont été créés dynamiquement depuis le portail de plateforme :

1. **Statut Préopératoire** : désigne la phase avant l'opération où des données cliniques et des données sur la qualité de vie sont collectées auprès des patients.
2. **Statut Peropératoire** : désigne la phase pendant l'opération chirurgicale où des données sur l'anesthésie et le déroulement de la chirurgie sont collectées.
3. **Statut Postopératoire** : désigne la phase après l'opération chirurgicale où des données physiologiques, sur la nutrition, la douleur, la qualité du sommeil et l'activité physique sont collectées auprès des patients.
4. **Statut Domicile** : désigne la phase où le patient rentre chez lui à son domicile.

Les statuts créés permettent d'organiser la collecte de données auprès du patient suivant des étapes chronologiques permettent ainsi de regrouper les patients selon la phase opératoire où ils se trouvent.

### IX.2.1.2 Définition des connaissances du domaine

Nous avons créé, en collaboration avec les chirurgiens, 186 concepts pour représenter les connaissances du domaine de la chirurgie digestive et générale pour le projet Repose. Ces concepts ont été introduit dynamiquement depuis le portail web de la plateforme. Le tableau suivant présente le nombre de concepts créés en fonction de leurs types.

Concept du domaine	Nombre	Nombre (%)
Type de données mesures	42	22.58%
Type de données durées	7	3.76%
Type de données dates	5	2.68%
Type de données qualitatives	132	70.96%

TABLE IX.1: Liste des instances créés dans DTOnto en fonction de leurs types (projet REPOSE)

### IX.2.1.3 Création de questions

Les concepts du domaine créés ont permis la définition de 186 questions regroupées sous 14 questionnaires. Le tableau IX.2 présente le nombre de questions créées en fonction de leurs types.

Type de question	Nombre	Nombre (%)
Mesures	42	22.58%
Texte libre	24	12.90%
Durée	7	3.76%
Date	5	2.68%
Liste (Plusieurs éléments à sélectionner)	4	2.15%
Liste (Un seul élément à sélectionner)	101	54.30%
Booléenne	3	1.61%

TABLE IX.2: Liste des questions créés en fonction de leurs types (projet REPOSE)

Les questionnaires créés ont été rattachés aux différents statuts du patient créés permettant une collecte de données contextuelle pour bien organiser le travail quotidien des infirmières. Les questionnaires sont proposés suivants le statut du patient.

### IX.2.1.4 Définition des règles d'inférence

En collaboration avec les chirurgiens, des règles d'inférences ont été implémentées permettant la génération des scores sous forme d'indicateurs afin d'aider les chirurgiens dans leur prise de décision vise à vis le suivi médical des patients post-opératoire.

**Compliance ERAS.** La réhabilitation améliorée après chirurgie, ou Enhanced Recovery after Surgery (ERAS), est un programme médical qui vise à réduire les complications postopératoires, la durée d'hospitalisation et les coûts qui en découlent.

À partir du guide de bonnes pratiques proposé dans (Gustafsson et al., 2012) et en collaboration avec les chirurgiens, nous avons implémenté des règles permettant de calculer le score de compliance ERAS.

Le score obtenu est sauvegardé en tant qu'indicateur dans la base de faits pour le patient concerné. Ensuite, l'indicateur inféré est affiché sur le portail web de la plateforme afin de permettre aux chirurgiens d'estimer la qualité du suivi médical et de prendre nouvelles mesures pour la prise en charge du patient concerné.

Le calcul de la compliance a été testé sur les patients inclus dans le système et les résultats obtenus ont été évalués et validés par les chirurgiens.

**CCI.** L'indice global Complication après chirurgie, ou The Comprehensive Complication Index (CCI), est un indice permettant d'évaluer le degré des complications dans la phase postopératoire. À partir du guide de bonnes pratiques proposé dans (Slankamenac K, 2013) et en collaboration avec les chirurgiens, nous avons implémenté la règle permettant de calculer le score du CCI. Le score obtenu est sauvegardé en tant qu'indicateur dans la base de fait pour le patient concerné. Ensuite, l'indicateur inféré est affiché sur le portail web de la plateforme afin de permettre aux chirurgiens d'estimer le degré des complications et de prendre mesures pour la prise en charge du patient concerné.

Le calcul du score de CCI a été testé sur les patients inclus dans le système et les résultats obtenus ont été évalués et validés par les chirurgiens.

**BYPASS 3J.** Le BYPASS 3J c'est un indicateur permettant aux chirurgiens de déterminer si le patient opéré du bypass gastrique peut sortir au bout du 3<sup>ème</sup> jour après son opération ou pas. Pour déterminer ça, nous avons défini et implémenté en collaboration avec les chirurgiens la règle bypass qui permet de calculer un score à partir de certaines données collectées. Le score obtenu est représenté comme un indicateur et affiché dans le portail web de la plateforme afin de permettre aux chirurgiens de déterminer si le patient est apte de sortir ou pas.

Le calcul du score de bypass 3J a été testé sur les patients opérés du bypass gastrique inclus dans le système et les résultats obtenus ont été évalués et validés par les chirurgiens.

## IX.2.2 Projet Brave

Le projet Brave est un projet mené par l'entreprise Newel en collaboration avec le service de la Chirurgie Viscérale du Centre Hospitalier Universitaire Vaudois de Lausanne en Suisse sous la direction du professeur DEMARTINES. Le but du projet étant de proposer une plateforme dédiée à la réhabilitation améliorée afin de permettre de réaliser un suivi médical de patients avant et après une chirurgie viscérale.

L'expérimentation a débuté en mois de Juin 2016 qui s'étale sur une période d'une année

où durant les deux premiers mois plus de 20 patients ont été inclus.

### IX.2.2.1 Définition de contexte de patient

Pour la définition du contexte des patients avant et après la chirurgie viscérale. Nous avons défini en collaboration avec les chirurgiens 8 statuts patients afin de modéliser les étapes chronologiques auxquels le patient peut se trouver au cours de son suivi médical. Ces statuts ont été créés dynamiquement depuis le portail de plateforme :

1. **Statut Préopératoire (J-15)** : désigne l'étape à 15 jours avant l'opération chirurgicale où le profil du patient est créé dans la plateforme et les données administratives sont recueillies.
2. **Statut Préopératoire domicile** : désigne la phase où le patient est rentré chez lui. Aucune donnée n'est collectée durant cette phase.
3. **Statut Préopératoire (J-1)** : désigne la veille de l'opération chirurgicale où des données sur la qualité de vie sont collectées auprès du patient.
4. **Statut Postopératoire** : désigne la phase après l'opération chirurgicale où des données physiologiques, douleurs, fatigue, qualité du sommeil et l'activité sont collectées auprès des patients.
5. **Statut Sortie** : désigne la sortie de patient de l'hôpital où des données sur les complications et sur l'état de santé général sont collectées.
6. **Statut Postopératoire Domicile** : désigne la phase où le patient rentre chez lui.
7. **Statut Postopératoire (J+30)** : désigne la dernière étape à 30 jours à peu près après l'intervention chirurgicale, où le patient revient pour un contrôle médical chez les chirurgiens. Durant ce contrôle médical des données sur les complications sont collectées.
8. **Statut Terminé** : désigne la fin du suivi médical.

Les statuts créés permettent d'organiser la collecte de données auprès du patient suivant les étapes chronologiques définies par les chirurgiens. Ils permettent ainsi de regrouper les patients selon l'étape opératoire où ils se trouvent.

### IX.2.2.2 Définition des connaissances du domaine

Nous avons défini en collaboration avec les chirurgiens 155 concepts pour représenter les connaissances du domaine de la chirurgie viscérale pour le projet Brave. Ces concepts ont été introduits dynamiquement depuis le portail web de la plateforme. Le tableau IX.3 présente le nombre d'instances de concepts créés en fonction de leurs types.

Concept du domaine	Nombre	Nombre (%)
Type de données mesures	14	9.03%
Type de données durées	0	0%
Type de données dates	4	2.58%
Type de données qualitatives	137	88.38%

TABLE IX.3: Liste des instances créées dans DTOnto en fonction de leurs types (projet BRAVE)

### IX.2.2.3 Création de questions

Les concepts du domaine définis ont permis la création de 155 questions regroupés sous 8 questionnaires réparties sur 5 statuts. Le tableau IX.4 présente le nombre de questions créées en fonction de leurs types.

Type de question	Nombre	Nombre (%)
Mesures	14	9.03%
Texte libre	8	5.16%
Durée	0	0%
Date	4	2.58%
Liste (Plusieurs éléments à sélectionner)	0	0%
Liste (Un seul élément à sélectionner)	129	83.22%
Booléenne	0	0%

TABLE IX.4: Liste des questions créées en fonction de leurs types (projet BRAVE)

### IX.2.2.4 Définition de règles d'inférence

La seule règle définie jusqu'à présent est la règle du CCI présentée précédemment.

## IX.2.3 Projet Chariot EHPAD

Le projet Chariot EHPAD est mené par l'entreprise Newel en collaboration avec le CENTICH (Centre d'Expertise National des Technologies de l'Information et de la Communication) et les maisons de retraites EHPAD (établissement d'hébergement pour personnes âgées dépendantes) Les Noisetiers, L'Orée des bois et Picasso dans le Maine et Loire. Le but de ce projet est d'assurer un suivi médical pour des résidents de ces EHPAD. Une première expérimentation a débuté en avril 2016 pour une durée de 6 mois, de ce fait 228 résidents polypathologiques ont été inclus avec une moyenne d'âge de 90 ans. Ils souffrent de plusieurs pathologies chroniques telles que : le diabète, l'hypertension, l'insuffisance cardiaque, etc. Les résidents sont répartis sur les trois EHPAD : Les noisetiers (90 résidents), L'Orée des Bois (94 résidents) et Picasso (44 résidents).

### IX.2.3.1 Définition du contexte de résident

En collaboration avec les professionnels de santé des EHPAD nous avons défini 3 statuts pour modéliser le contexte où peut se trouver un résident. Ces statuts ont été créés dynamiquement depuis le portail de la plateforme :

1. **Statut Présent** : désigne que le résident est présent dans la maison et qu'il peut être suivi par la plateforme.
2. **Statut Absent** : désigne que le résident n'est pas présent dans la maison et par conséquent, il ne peut pas être suivi par la plateforme.
3. **Statut Mort** : désigne que le résident est décédé. Par conséquent, fin du suivi médical.

Ces statuts permettent de mieux regrouper les résidents et d'organiser la collecte des données.

### IX.2.3.2 Définition des connaissances de domaine

Nous travaillons actuellement sur la définition des concepts pour la représentation des connaissances du domaine en collaboration avec les experts médicaux des centres EHPAD. Dans une première étape nous avons défini 38 concepts présentés dans le tableau IX.5.

Concept du domaine	Nombre	Nombre (%)
Type de données mesures	32	84.21%
Type de données durées	0	0%
Type de données dates	2	5.26%
Type de données qualitatives	4	10.52%

TABLE IX.5: Liste des instances créées dans DTOnto en fonction de leurs types (projet EHPAD)

Dans la première phase d'expérimentation, ces concepts de domaine ont permis de représenter les types des mesures physiologiques et les types des données administratives des résidents.

### IX.2.3.3 Création des questions

Dans la première phase d'expérimentation, la collecte des données de type mesures est faite principalement à l'aide des capteurs médicaux permettant de mesurer les données physiologiques des résidents telles que : l'activité physique, la qualité de sommeil, la glycémie, la tension artérielle, le rythme cardiaque, la température et l'oxymétrie. Pour chaque type de mesure une question a été créée permettant la collecte en cas de problème au niveau du capteur. La définition des questions sur la qualité de vie et l'hygiène de vie des résidents est en cours. Elle dépend des connaissances de domaine et de la définition de nouveaux concepts du domaine.



### IX.2.3.4 Définition de règles d'inférence

Nous avons défini en collaboration avec des médecins, des règles permettant de détecter si le résident se trouve dans une situation d'hyper ou d'hypoglycémie. Les données inférées par ces règles sont matérialisées par des alertes servant principalement à alerter les soignants des EHPAD afin qu'ils adaptent le traitement aux résidents concernés.

Ces règles sont exécutées à chaque collecte de données de type glycémie. Les données inférées sont enregistrées comme alertes dans la base.

Ces règles permettent de traiter les cas suivants :

- Hyperglycémie critique (alerte orange) si glycémie  $>4$  g/l sur 24H en continue.
- Hyperglycémie sévère (alerte rouge) si glycémie  $>4$  g/l pendant 5 jours soit :
  - Tous les matins,
  - Tous les midis,
  - Tous les soirs.
- Hypoglycémie (alerte orange) si glycémie  $<0,5$ g/l sur 24H en continue.
- Hypoglycémie (alerte rouge) si glycémie  $<0,5$ g/l pendant 3 jours soit :
  - Tous les matins,
  - Soit tous les midis,
  - Soit tous les soirs.

Les règles de glycémie implémentées sont en phase d'expérimentation sur les résidents diabétiques des 3 EHPAD.

## IX.3 Évaluation

Dans cette section nous allons présenter les résultats suite à l'évaluation de la plateforme E-care.

### IX.3.1 Collecte de données

Afin d'évaluer l'outil de collecte de données, nous avons réalisé une enquête de satisfaction auprès des utilisateurs de l'outil (les cliniciens) dans le cadre des projets *Repose* et *Brave*. Les cliniciens ont été interrogés par rapport à plusieurs points tel que la facilité de l'utilisation, l'ergonomie, la rapidité dans l'accès aux données, l'organisation de la collecte et l'amélioration du travail quotidien. Le déploiement de l'outil était précédé d'une phase d'évaluation de l'ergonomie et des choix technologiques, les résultats de l'enquête ont logiquement confirmé la satisfaction des soignants de manière générale quant à l'utilisation d'un tel outil. La collecte de données médicales ainsi réalisée a révélé plusieurs avantages par rapport aux méthodes classiques (papier ou fichier Excel). L'outil a permis une meilleure organisation de la collecte en fonction du contexte des patients tout en minimisant le temps et l'effort.

## IX.3.2 Raisonnement à base de cas

### IX.3.2.1 Évaluation de deux modes de calcul des similarités

Nous avons évalué les deux modes de calcul des similarités proposés dans la partie contribution à savoir le mode à la demande, où le calcul des similarités se fait à la demande des médecins, et le mode à la collecte de données, où le calcul des similarités se fait en arrière-plan à chaque nouvelle donnée collectée.

L'évaluation a été faite dans le cadre du projet Repose sur 270 cas, sur un serveur Linux disposant de 16 GO de mémoire vive (RAM) et un CPU de 8 cœurs. L'évaluation consiste donc à calculer le temps d'exécution (en millisecondes) de l'étape de la recherche des cas en fonction du nombre des critères sélectionnés pour la recherche des cas et selon deux modes d'exécution proposés. Les résultats obtenus sont illustrés dans la figure IX.1 :

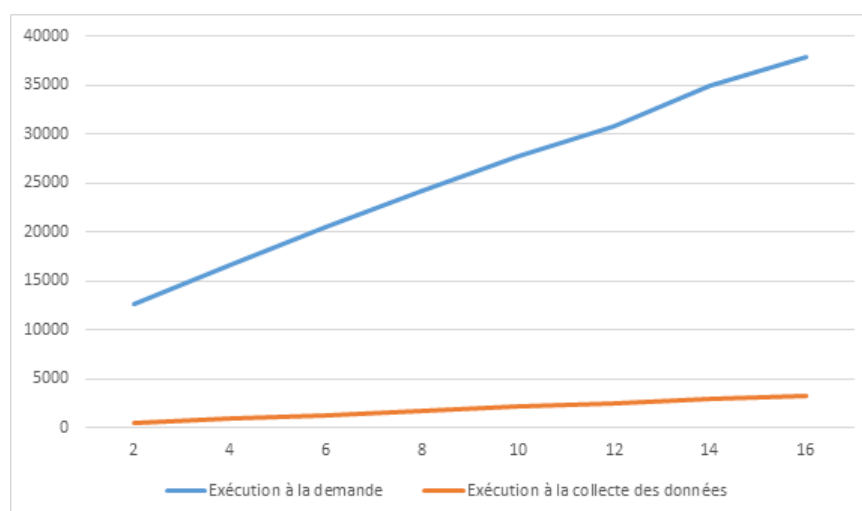


FIGURE IX.1 – Le temps de la recherche des cas (Millisecondes) en fonction du nombre de critères sélectionnés selon les deux modes d'exécutions

La figure montre que le temps de calcul des similarités augmente proportionnellement en fonction du nombre de critères (type de données) pris en compte dans le calcul, où plus on sélectionne de critères plus le temps d'exécution est important.

Nous constatons ainsi que le premier mode de calcul (exécution à la demande) est beaucoup plus coûteux en termes de temps, ceci est dû au temps de chargement des données et au temps mis pour le calcul des similarités locales et de la similarité globale. Tandis que le deuxième mode (exécution à la collecte des données) est moins coûteux étant donné que l'étape de la recherche des cas dans ce mode consiste uniquement à charger les scores des similarités déjà calculés depuis la matrices des similarités. Seul la similarité globale est calculée.

### IX.3.2.2 Évaluation expérimentale des similarités

L'évaluation expérimentale a pour but de juger de la capacité de détection du système à retrouver des cas similaires pour résoudre un nouveau cas de patient.

L'évaluation a été faite dans le cadre du projet REPOSE où la base de cas utilisée est constituée de 270 patients qui ont subi différentes chirurgies de l'appareil digestive.

Cette évaluation a consisté à tester la capacité du système à rechercher des cas similaires en fonction des critères de recherche sélectionnés. Pour ce faire, nous avons pris le cas cible suivant :

- Sexe : Femme
- Âge : 83 ans
- Type Opération effectué : Colorectale
- Durée de Séjour à l'hôpital : 9 jours

Ensuite, sur l'ensemble des 186 critères de recherche proposés par le système et à la demande des chirurgiens, nous avons sélectionné uniquement les critères suivants :

- **Activité physique** : le nombre de pas effectué par jour, cette donnée est collectée depuis un capteur connecté (podomètre).
- **Type de mobilité préopératoire** : le type de mobilité du patient avant l'opération, trois valeurs sont possibles :
  - Sédentaire : correspond à une activité quasi inexistante du patient.
  - Modérée : correspond à une activité modérée du patient.
  - Intense : correspond à une activité intense du patient.

Cette donnée n'est collectée qu'une seule fois dans la phase préopératoire à l'aide d'un questionnaire de fréquence unique.

- **Type de mobilité post-opératoire** : le type de mobilité du patient après l'opération, quatre valeurs sont possibles
  - Lit : le patient n'a pas bougé de son lit.
  - Fauteuil : le patient s'est mis sur un fauteuil.
  - Petit couloir : le patient a marché dans le petit couloir du service.
  - Grand couloir : le patient a marché dans le grand couloir du service.

Cette donnée est collectée tous les jours à l'aide d'un questionnaire de fréquence multiple dans la phase postopératoire.

- **Évaluation de la douleur** : le degré de douleur ressenti par le patient après l'opération. Cette donnée est collectée tous les jours en phase postopératoire à l'aide d'un questionnaire de fréquence multiple. La valeur de la douleur vaut de 0 à 10 où 0 correspond à « Absence de douleur » et 10 correspond à « Douleur maximale ».
- **Type d'opération** : le type d'opération subit par le patient. Nous avons fixé le seuil de similarité locale à 100% afin de sélectionner que les patients qui ont subi le même type d'opération que le cas cible à savoir une opération au niveau du colon.

Le seuil de similarité globale est fixé à 50%, les poids attribués aux attributs sont fixés à 1. Le tableau ci-après montre les résultats des cas les plus similaires retrouvés par le système.

Cas sources	Activité	Mobilité préop	Mobilité postop	Douleur	Similarité
Cas 1	49.6%	100%	55.6%	42.5%	69.53%
Cas 2	39.9%	100%	33.3%	57.5%	66.13%
Cas 3	43.3%	100%	0%	71.9%	63.03%
Cas 4	50.3%	100%	0%	54.4%	60.93%
Cas 5	50%	0%	55.6%	66.9%	54.49%

TABLE IX.6: Liste des cas les plus similaires retrouvés

Dans un second temps nous avons sélectionné un nouvel critère à savoir ‘durée de séjour’ (DS) qui correspond au nombre de jours d’hospitalisation. Le poids attribué à ce dernier est égale à 1. La durée de séjour respective pour chaque cas est comme suit : cas 1 : (7 jours), cas 2 : (15 jours), cas 3 : (6 jours), cas 4 : (4 jours) et cas 5 : (11 jours), Nous avons relancé la recherche des cas les plus similaires, en gardant les autres critères. Les résultats sont montrés dans le tableau IX.7 :

Cas sources	Activité	Mobilité préop	Mobilité postop	Douleur	Durée Séjour	Similarité
Cas 2	39.9%	100%	33.3%	57.5%	77.77%	61.69%
Cas 1	49.6%	100%	55.6%	42.5%	60%	61.54%
Cas 3	43.3%	100%	0%	71.9%	66.66%	56.37%
Cas 5	50%	0%	55.6%	66.9%	77.77%	54.49%
Cas 4	50.3%	100%	0%	54.4%	44.44%	49.82%

TABLE IX.7: Liste des cas les plus similaires retrouvés après l’intégration de la durée de séjour

Comme montre le tableau ci-avant, la prise en compte du critère « durée de séjour (DS) » a impacté le calcul du score des similarités globales et, par conséquent, le classement des cas sélectionnés. Le cas 1 a été déclassé devant le cas 2 car sa durée de séjour est moins proche ( $|15-9| = 4$  jours) que celle du cas 2 ( $|7-9| = 2$  jours) par rapport au cas cible. Le cas 4 n’a pas été retenu, étant donné que le seuil minimal pour la sélection des cas est à 50%.

### IX.3.2.3 Évaluation des similarités en milieu hospitalier

L’évaluation en milieu hospitalier a pour but de juger la pertinence des cas similaires retrouvés par le système. L’étude a été menée sur une dizaine de patients récemment opérés dans le service de Chirurgie Générale et Digestive de l’hôpital Hautepierre en Alsace. La base de cas utilisée est constituée de 270 cas sources. Ces patients ont été choisis de manière à représenter l’ensemble des types de patients rencontrés. Les résultats de chaque cas de patient testé ont été analysés a posteriori avec les experts médicaux. Trois classes de résultats sont discriminées :

- Très cohérent : si le cas du patient similaire est réellement proche du nouveau cas ;
- Peu cohérent : s'il y a des points de ressemblance malgré des différences ;
- Pas cohérent : si le patient retrouvé lui paraît éloigné du patient à traiter.

En moyenne 5 cas similaires ont été retrouvés par le système pour chaque cas cible, ce qui fait une moyenne globale de 50 cas similaires retrouvés pour les 10 cas cibles testés. Les résultats de l'évaluation est illustré par la figure IX.2.

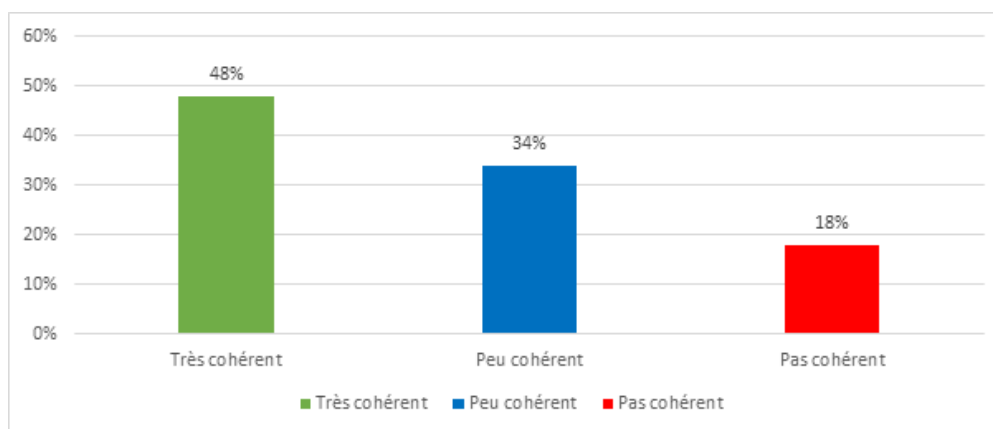


FIGURE IX.2 – Évaluation des résultats de similarités en milieu hospitalier

Le graphique ci-avant montre que les résultats retournés par le système ont été jugés plutôt cohérents, avec 48% « Très cohérent » et 34% « cohérent », contre 18% pour un résultat jugé « Pas cohérent ».

Le résultat jugé « Pas cohérent » est principalement lié à la non prise en compte de certaines valeurs dans le calcul de la similarité locale de l'activité physique. Ceci s'explique par l'écart existant entre la période d'hospitalisation (durée de séjour) du cas cible et celle de certains cas sources.

### IX.3.3 Raisonnement à base de règles

La capacité du système à inférer des données à partir des règles d'inférence a été évaluée dans le cadre des trois projets de suivi médical cités précédemment. Les données inférées à savoir les indicateurs (CCI, ERAS et PYPASS 3J) ont été validées dans le cadre du projet REPOSE à 100% par les chirurgien-experts. Les alertes générées pour la détection d'hyperglycémie et hypoglycémie ont été validées dans le cadre du projet EHPAD à 100% par les experts médicaux.

## IX.4 Conclusion

Nous avons exposé dans ce chapitre les différentes expérimentations menées dans le cadre de plusieurs projets de suivi médical pour la validation de l'approche proposée au sein de

la plateforme E-care.

Les résultats de ces expérimentations ont montré l'efficacité de l'approche proposée. L'adaptation de la plateforme par rapport au domaine et au mode d'intervention de chacune de ces expérimentations se limitait à sa configuration. De plus, l'approche proposée a suscité l'intérêt du personnel médical impliqués dans les différentes expérimentations aussi bien par rapport à l'organisation de la collecte des données, qui tient compte du contexte du patient, que par rapport à l'exploitation des connaissances médicales qui apporte aux professionnels de santé une assistance pour une meilleure prise de décision.

La deuxième partie de ce chapitre était consacrée à l'évaluation de la plate-forme où nous avons évalué l'outil proposé pour collecte de données médicales et le raisonnement à base de cas et à base de règles dans un milieu hospitalier et avec des données réelles.



# IV

## CONCLUSION GÉNÉRALE





## CONCLUSION GÉNÉRALE

### X.1 Synthèse des contributions

Les travaux présentés dans ce manuscrit s'inscrivent dans le contexte de la construction des systèmes d'aide à la décision médicale à base de connaissances. C'est un des domaines émergents de l'intelligence artificielle appliquée au domaine médical avec de nombreux enjeux multidisciplinaires tels que l'acquisition des données médicales, la représentation des connaissances médicales ou encore le raisonnement à base de connaissances.

L'objet de cette thèse étant la proposition, l'implémentation et l'expérimentation d'un système d'aide à la décision médicale ouvert et générique à base de connaissances et de l'utilisation des ontologies. Nous nous sommes particulièrement intéressés à la représentation des connaissances des experts du domaine, à l'acquisition des données médicales et à l'intégration de modes de raisonnements à base de connaissances guidés par une ontologie de domaine.

Nous avons axé notre revue de l'état de l'art selon les trois problématiques traitées dans la thèse en donnant un aperçu des différentes approches qui ont été proposées pour chacune des problématiques.

Dans cette direction de recherche, nous avons proposé plusieurs contributions que nous décrivons brièvement ci-dessous :

- Représentation des connaissances : nous avons proposé une représentation sémantique des connaissances à l'aide des ontologies sur deux niveaux :
  - Le premier niveau concerne la modélisation des connaissances médicales utilisées pour l'élaboration du raisonnement. Deux ontologies de domaine, basées sur des classifications existantes dans la littérature (ATC et CIM-10), ont été construites afin d'offrir une représentation à la fois exhaustive et détaillée sur les antécédents médicaux. Une autre ontologie de domaine (DTOnto) a été ainsi construite pour la représentation des connaissances spécifiques du mode d'intervention du système. L'intérêt d'une telle représentation est d'assurer le contrôle du vocabulaire des termes et la facilité du partage des connaissances.

- Le deuxième niveau concerne la modélisation de profil patient. Nous nous sommes reposés sur une approche de représentation développée dans un travail précédent au sein de notre équipe dans (Benyahia, 2015) et nous l’avons étendue pour qu’elle soit plus générique. Pour cela, nous avons développé l’ontologie Patient Profil Ontology (PPO) qui est guidée par les ontologies de domaine (ATC, CIM-10 et DTOnto) où les antécédents médicaux présents dans le profil sont issus des ontologies ATC et CIM-10, tandis que les données collectées présents dans le profil font référence à des concepts définis dans DTOnto.
- Acquisition de données médicales : nous avons proposé une approche d’acquisition permettant l’automatisation et l’organisation du recueil des données médicales. Cette approche se base sur l’utilisation des ontologies et sur un mécanisme de questions/réponses. Les ontologies permettent la définition et la structuration de questionnaires contextuels, adaptatifs et évolutifs. Ces questionnaires sont sensibles au contexte du patient interrogé et adaptés aux réponses apportées. Leurs structures peuvent évoluer chaque fois que le besoin se présente et se limite à une simple configuration. La création des questions est guidée par l’ontologie du domaine DTOnto qui permet de donner une signification aux questions créées et aux données collectées. Les ontologies ont été ainsi utilisées pour la modélisation des interrogatoires médicaux afin de permettre la structuration et l’historisation des réponses (données collectées).

Nous avons proposé deux modes d’acquisitions de données, le mode séquentiel où les questions sont posées les unes après les autres dans un ordre défini par les experts médicaux, et le mode instantané où les questions sont posées toutes à la fois sous forme de liste. Nous avons montré pour chacun des modes son intérêt et son apport dans l’organisation de la collecte de données en fonction du mode d’intervention du système.
- Raisonnement mixte à base de connaissances : nous avons proposé un mode de raisonnement à base de connaissances guidé par une ontologie de domaine. Ce mode de raisonnement permet l’exploitation des deux sources de connaissances dont l’une complète l’autre, à savoir : les guides de bonnes pratiques et l’expérience des cliniciens dans la définition des thérapies. Les guides de bonnes pratiques sont modélisés sous forme de règles et exploités par un moteur à base de règles et L’expérience des cliniciens est modélisée sous forme de cas et exploitée par un moteur à base de cas.

## X.2 Perspectives

Les différentes expérimentations et évaluations menées ont montré l’efficacité de nos contributions vis-à-vis des approches proposées dans l’état de l’art concernant la représentation des connaissances, l’acquisitions des données, et le raisonnement à base de connaissances. Le système implémenté a montré son intérêt dans le milieu médical.

Ce manuscrit ouvre de nombreuses perspectives que nous synthétisons dans ce qui suit.

À court terme, nous souhaitons améliorer nos contributions selon trois aspects :

- Définition d'un cadre d'évaluation pour la recommandation des thérapies : pour cela il faudrait disposer d'abord d'un grand nombre de cas et d'un nombre important de thérapies définies dans le système ;
- Expérimentation du système dans le cadre d'un suivi à domicile pour les patients atteints de maladies chroniques. Ce cadre va nous permettre d'expérimenter la collecte de données et la recommandation des conseils thérapeutiques et voir comment les patients vont se comporter avec un tel système.

À long terme, nous souhaitons :

- L'intégration des techniques de fouilles de données afin d'optimiser la recherche des cas similaires ;
- La proposition de nouvelles interfaces permettant la définition des règles d'inférence de manière simplifiée. Dans notre système, cette définition reste difficile car les règles sont exprimées dans un langage non maîtrisé par les experts médicaux.



# BIBLIOGRAPHIE

- Agnar Aamodt and Enric Plaza. Case-based reasoning : Foundational issues, methodological variations, and system approaches. **AI Commun.**, 7(1) :39–59, March 1994. ISSN 0921-7126.
- Blackford Middleton Tonya Hongsermeier Vipul Kashyap Sean M. Thomas Dean F. Sittig Adam Wright, David W. Bates. Creating and sharing clinical decision support content with web 2.0 : Issues and examples. **Journal of Biomedical Informatics**, 42 :334–346, 2009.
- M. Alipour, Aghdam. Ontology-driven generic questionnaire design. thesis for the degree of master of science in computer science. Master’s thesis, University of Guelph, August 2014.
- M. H. Alsun. **Indexation guidée par les connaissances en imagerie médicale**. PhD thesis, Télécom Bretagne, 2012.
- Alain Mille Amélie Cordier, Bruno Masclet. Étendre les possibilités du raisonnement à partir de cas grâce aux traces. In **7ème atelier de Raisonnement à Partir de Cas**. 2009.
- I. Autonomy. Autonomy technology overview. 2009.
- Shahina Begum. A case-based reasoning system for the diagnostic of individual sensitivity to stress in psychophysiology. Master’s thesis, Mälardalen University, 2009.
- A. A. Benyahia, A. Hajjam, V. Hilaire, and M. Hajjam. e-care : Ontological architecture for telemonitoring and alerts detection. In **2012 IEEE 24th International Conference on Tools with Artificial Intelligence**, volume 2, pages 13–17, Nov 2012. doi : 10.1109/ICTAI.2012.183.
- A. A. Benyahia, A. Hajjam, V. Hilaire, M. Hajjam, and E. Andres. E-care telemonitoring system : Extend the platform. In **Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on**, pages 1–4, July 2013. doi : 10.1109/IISA.2013.6623725.
- A. Ahmed Benyahia. **Etude d’une méthodologie pour la construction d’un système de t élé-surveillance médicale**. PhD thesis, Université de Technologie Belfort-Montbéliard, 2015.
- Petr Berka. Nest : A compositional approach to rule-based and case-based reasoning. **Adv. in Artif. Intell.**, 2011 :4 :1–4 :15, January 2011. ISSN 1687-7470. doi : 10.1155/2011/374250. URL <http://dx.doi.org/10.1155/2011/374250>.
- M. Bouamrane, A. Rector, and M. Hurrell. Gathering precise patient medical history with an ontology-driven adaptive questionnaire. In **Computer-Based Medical Systems, 2008. CBMS ’08. 21st IEEE International Symposium on**, pages 539–541, June 2008a. doi : 10.1109/CBMS.2008.24.

- Matt-Mouley Bouamrane, Alan Rector, and Martin Hurrell. Ontology-driven adaptive medical information collection system. In **Foundations of Intelligent Systems**, volume 4994 of **Lecture Notes in Computer Science**, pages 574–584. Springer Berlin Heidelberg, 2008b. ISBN 978-3-540-68122-9. doi : 10.1007/978-3-540-68123-6\_62. URL [http://dx.doi.org/10.1007/978-3-540-68123-6\\_62](http://dx.doi.org/10.1007/978-3-540-68123-6_62).
- Matt-Mouley Bouamrane, Alan Rector, and Martin Hurrell. Using ontologies for an intelligent patient modelling, adaptation and management system. In **On the Move to Meaningful Internet Systems : OTM 2008**, volume 5332 of **Lecture Notes in Computer Science**, pages 1458–1470. Springer Berlin Heidelberg, 2008c. ISBN 978-3-540-88872-7. doi : 10.1007/978-3-540-88873-4\_36. URL [http://dx.doi.org/10.1007/978-3-540-88873-4\\_36](http://dx.doi.org/10.1007/978-3-540-88873-4_36).
- Matt-Mouley Bouamrane, Alan Rector, and Martin Hurrell. Using owl ontologies for adaptive patient information modelling and preoperative clinical decision support. **Knowledge and Information Systems**, 29(2) :405–418, 2011. ISSN 0219-3116. doi : 10.1007/s10115-010-0351-7. URL <http://dx.doi.org/10.1007/s10115-010-0351-7>.
- Aussenac-Gilles N. et Charlet J. Bourigault, D. Construction de ressources terminologiques ou ontologiques à partir de textes un cadre unificateur pour trois études de cas. **Revue d’Intelligence Artificielle**, page 87–110, 2004.
- M. M. Cabrera and E. O. Edey. Integration of rule based expert systems and case based reasoning in an acute bacterial meningitis clinical decision support system. **International Journal of Computer Science and Information Security**, (2) :112–118, 2010.
- B. Chakraborty, S. I. Srinivas, P. Sood, V. Nabhi, and D. Ghosh. Case based reasoning methodology for diagnosis of swine flu. In **GCC Conference and Exhibition (GCC), 2011 IEEE**, pages 132–135, Feb 2011.
- Adrien Coulet. Ontologies et alignement d’ontologies : notes de cours. 19, Novembre 2007. Dernière visite novembre 2015.
- Catarina DA-SILVA. **Découverte de correspondances sémantiques entre ressources hétérogènes dans un environnement coopératif**. PhD thesis, Université Claude Bernard, Lyon 1, 2007.
- Vladimir Dimitrieski, Gajo Petrović, Aleksandar Kovačević, Ivan Luković, and Hamido Fujita. **A Survey on Ontologies and Ontology Alignment Approaches in Healthcare**, pages 373–385. Springer International Publishing, Cham, 2016. ISBN 978-3-319-42007-3. doi : 10.1007/978-3-319-42007-3\_32. URL [http://dx.doi.org/10.1007/978-3-319-42007-3\\_32](http://dx.doi.org/10.1007/978-3-319-42007-3_32).
- D. Dinevski, O. Šušteršič, T. Šarenac, U. Bele, and U. Rajkovič. **Clinical Decision Support Systems**. INTECH Open Access Publisher, 2011. ISBN 9789533073545. URL <https://books.google.fr/books?id=f4emoAEACAAJ>.

- Ba-Duy Dinh. **Accès à l'information biomédicale : vers une approche d'indexation et de recherche d'information conceptuelle basée sur la fusion de ressources terminologiques.** PhD thesis, Université Paul Sabatier, 2012.
- FCMI FHIMSS Tonya J. La Lande Eta S, Berner EdD. **Clinical Decision Support Systems Book Subtitle**, chapter Overview of Clinical Decision Support Systems, pages 3–22. Health Informatics. Springer New York, 2007.
- J. Lieber F. Le Ber and A. Napoli. Les systèmes à base de connaissances. **Encyclopédie de l'informatique et des systèmes d'information**, pages 1197–1208, 2006.
- K. Farooq, A. Hussain, S. Leslie, C. Eckl, and W. Slack. Ontology-driven cardiovascular decision support system. In **Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on**, pages 283–286, May 2011.
- C. Franco, J. Demongeot, Y. Fouquet, C. Villemazet, and N. Vuillerme. Perspectives in home telehealthcare system : Daily routine nycthemeral rhythm monitoring from location data. In **Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on**, pages 611–617, Feb 2010a. doi : 10.1109/CISIS.2010.192.
- C. Franco, J. Demongeot, C. Villemazet, and N. Vuillerme. Behavioral telemonitoring of the elderly at home : Detection of nycthemeral rhythms drifts from location data. In **Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on**, pages 759–766, April 2010b. doi : 10.1109/WAINA.2010.81.
- Mille Frederic. **Systèmes de détection des interactions médicamenteuses : points faibles & propositions d'améliorations.** PhD thesis, Université pierre et marie curie, Décembre 2008.
- Gersende Georg. **Analyse Informatique de Guides de Bonnes Pratiques Cliniques.** Informatique médicale, Université paris 6, Septembre 2006.
- Mary K. Goldstein, Robert W. Coleman, Samson W. Tu, Ravi D. Shankar, Martin J. O'Connor, Mark A. Musen, Susana B. Martins, Philip W. Lavori, Michael G. Shlipak, Eugene Odone, Aneel A. Advani, Parisa Gholami, and Brian B. Hoffman. Translating research into practice : organizational issues in implementing automated decision support for hypertension in three medical centers. **Journal of the American Medical Informatics Association**, 11 (5) :368 – 376, 2004. ISSN 1067-5027. doi : <http://dx.doi.org/10.1197/jamia.M1534>. URL <http://www.sciencedirect.com/science/article/pii/S1067502704000957>.
- Assaf Gottlieb, Gideon Stein, Eytan Ruppim, Russ Altman, and Roded Sharan. A method for inferring medical diagnoses from patient similarities. **BMC Medicine**, 11(1) :194, 2013. ISSN 1741-7015. doi : 10.1186/1741-7015-11-194. URL <http://www.biomedcentral.com/1741-7015/11/194>.
- Thomas R. Gruber. A translation approach to portable ontology specifications. **Knowledge Acquisition**, 5(2) :199–220, June 1993.



- Nicola Guarino. Understanding, building and using ontologies. **International Journal of Human-Computer Studies**, 46 :293–310, February 1997.
- Mahdi Gueffaz. **ScaleSem : Model Checking et Web Sémantique**. PhD thesis, Université de Bourgogne, Décembre 2012.
- Dong-ling Xu et Jian-bo Yang Guilan kong. Clinical decision support systems : a review on knowledge representation and inference under uncertainties. **International Journal of Computational Intelligence Systems**, 1(2) :159–167, May 2008.
- U.O. Gustafsson, M.J. Scott, W. Schwenk, N. Demartines, D. Roulin, N. Francis, C.E. McNaught, J. MacFie, A.S. Liberman, M. Soop, A. Hill, R.H. Kennedy, D.N. Lobo, K. Fearon, and O. Ljungqvist. Guidelines for perioperative care in elective colonic surgery : Enhanced recovery after surgery (eras®) society recommendations. **Clinical Nutrition**, 31(6) : 783 – 800, 2012. ISSN 0261-5614. doi : <http://dx.doi.org/10.1016/j.clnu.2012.08.013>. URL <http://www.sciencedirect.com/science/article/pii/S026156141200180X>.
- Asunción Gómez-Pérez. Ontological engineering : A state of the art, 1999.
- S. H. Ha and Z. Zhang. Collaborative intelligence for intelligent diagnosis systems in hospital environment. In **Knowledge Discovery and Data Mining, 2010. WKDD '10. Third International Conference on**, pages 532–535, Jan 2010. doi : 10.1109/WKDD.2010.128.
- Jean-Paul Haton and Marie-Christine Haton. Systèmes à bases de connaissances. **Techniques de l'ingénieur Gestion de contenus numériques**, base documentaire : TIB311DUO. (ref. article : h3740), 2016. URL <http://www.techniques-ingenieur.fr/base-documentaire/technologies-de-l-information-th9/gestion-de-contenus-numeriques-42311210/systemes-a-bases-de-connaissances-h3740/>. fre.
- Abdullah Uz Tansel Hsien-Tseng Wang. Composite ontology-based medical diagnosis decision support system framework. In **Communications of the IIMA**, volume 13, 2013.
- A. S. Hussein, W. M. Omar, X. Li, and M. Ati. Efficient chronic disease diagnosis prediction and recommendation system. In **Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on**, pages 209–214, Dec 2012. doi : 10.1109/IECBES.2012.6498117.
- Sébastien Dubois et al Isabel Estevez. Le raisonnement à partir de cas est-il utilisable pour l'aide à la conception inventive. In **14ème atelier de Raisonnement à Partir de Cas**. Besançon (France)., 30 – 31 Mars 2006.
- U. Ansari et D. Sharma J. Soni. Intelligent and effective heart disease prediction system using weighted associative classifiers. In **International Journal on Computer Science and Engineering (IJCSE)**, 2011.
- Tommi S. Jaakkola and Michael I. Jordan. Variational probabilistic inference and the QMR-DT network. **Journal Of Artificial Intelligence Research**, 10 :291–322, 1999. URL <http://arxiv.org/abs/1105.5462>.

- RENAUD-SALIS Jean louis, LAGOUARD Philippe, and DARMONI Stefan. **Étude des systèmes d'aide à la décision médicale**, 2010.
- Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. **CoRR**, cmp-lg/9709008, 1997. URL <http://arxiv.org/abs/cmp-lg/9709008>.
- S. Darmoni. J.L. Renaud-Salis, P. Lagouarde. **Étude des systèmes d'aide à la décision médicale**. Livrable 2., 12 Juillet 2010.
- R. J. Calantone K. K. Boyer, J. R. Olson and E. C. Jackson. Print versus electronic surveys : a comparison of two data collection methodologies. **Journal of Operations Management**, 20 (4) :357–373, 2002.
- Xavier Lacot. Introduction à owl, un langage xml d'ontologies web, Juin 2005. URL [http://lacot.org/public/introduction\\_a\\_owl.pdf](http://lacot.org/public/introduction_a_owl.pdf).
- M. K. Goldstein et al. Lai, S. Insights from testing the accuracy of recommendations from an automated decision support system for primary hypertension : Athena dss. In **MEDINFO**, volume CD : 1706. 2004.
- Pascal Lando. Conception et développement d'applications informatiques utilisant des ontologies : application aux eiah. In **actes des premières Rencontres jeunes chercheurs en EIAH (RJC-EIAH'06)**, Évry, France, 11–12 Mai 2006.
- Claudia Leacock and Martin Chodorow. Combining Local Context and WordNet Similarity for Word Sense Identification. In Christiane Fellbaum, editor, **WordNet : An electronic lexical database.**, chapter 13, pages 265–283. MIT Press, 1998. ISBN 978-0262061971.
- Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. **IEEE Intelligent Systems**, 16(2) :72–79, March 2001.
- Elvire Mervoyer. **L'interrogatoire**. Université Médicale Virtuelle Francophone, 2009. URL <http://campus.cerimes.fr/semiologie-cardiologique/enseignement/interrogatoire/site/html/cours.pdf>.
- A. Minutolo, G. Sannino, M. Esposito, and G. De Pietro. A rule-based mhealth system for cardiac monitoring. In **Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on**, pages 144–149, Nov 2010. doi : 10.1109/IECBES.2010.5742217.
- Riichiro Mizoguchi and Jacqueline Bourdeau. Using ontological engineering to overcome common ai-ed problems. **International Journal of Artificial Intelligence and Education**, 11 : 107–121, 2000.
- Hala Najmeddine, Frédéric Suard, Arnaud Jay, Philippe Marechal, and Marié Sylvain. Mesures de similarité pour l'aide à l'analyse des données énergétiques de bâtiments. In **Reconnaissance des Formes et Intelligence Artificielle**, Lyon, France., Janvier 2012.

- M. Reddy M. Xenou N. Oliver D. Johnston C. Toumazou P. Pesl, P. Herrero and P. Georgiou. An advanced bolus calculator for type 1 diabetes : System architecture and usability results'. **Biomedical and Health Informatics, IEEE Journal**, PP(99), August 2015.
- G. Kulkarni A. Rao J. Britto P. Ramnarayan, A. Tomlinson. **A novel diagnostic aid (ISABEL) : development and preliminary evaluation of clinical performance**, volume 107 of **Studies in Health Technology and Informatics**, pages 1091 – 1095. 2004.
- Alexander Panchenko and Olga Morozova. A study of hybrid similarity measures for semantic relation extraction. In **Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data**, HYBRID '12, pages 10–18, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=2388632>. 2388634.
- Sallie-Anne Pearson, Annette Moxey, Jane Robertson, Isla Hains, Margaret Williamson, James Reeve, and David Newby. Do computerised clinical decision support systems for prescribing change practice ? a systematic review of the literature (1990-2007). **BMC Health Services Research**, 9(1) :154, 2009. ISSN 1472-6963. doi : 10.1186/1472-6963-9-154. URL <http://dx.doi.org/10.1186/1472-6963-9-154>.
- Margot Phaneuf. La collecte des données base de toute intervention infirmière. novembre 2012. URL <http://www.prendresoin.org/?p=940>.
- Jim Prentzas and Ioannis Hatzilygeroudis. **Integrating Hybrid Rule-Based with Case-Based Reasoning**, pages 336–349. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002. ISBN 978-3-540-46119-7. doi : 10.1007/3-540-46119-1\_25.
- Valéry Psyché, Olavo Mendes, and Jacqueline Bourdeau. Apport de l'ingénierie ontologique aux environnements de formation à distance. **Revue des Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF)**, 10 :89–126, 2003.
- R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. **IEEE Transactions on Systems, Man, and Cybernetics**, 19(1) :17–30, Jan 1989. ISSN 0018-9472. doi : 10.1109/21.24528.
- AM Mohan. Rao. Simulation of a virtual physician for diagnosis : Novel applications of genetic code for interlinks in data bank. In **Internet and citizen card technologies**, pages 107– 111. Paper presented at the IPE (Institute of Public Enterprises, Hyderabad) National conferences on Medical Informatics, 2000a.
- MM. Rao. Cyber doc : Windows for medical diagnosis and data. In **Papers presented at the IPE National conferences on Medical Informatics**. 2000b.
- Alan Rector and Jeremy Rogers. **Ontological and Practical Issues in Using a Description Logic to Represent Medical Concept Systems : Experience from GALEN**, pages 197– 231. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-38412-0. doi : 10.1007/11837787\_9. URL [http://dx.doi.org/10.1007/11837787\\_9](http://dx.doi.org/10.1007/11837787_9).

- Pole P. A. Rector A. L., Rogers J. E. The galen high level ontology. **MIE 96**, MIE 96 :174–178, 1996.
- Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In **Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'95**, pages 448–453, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8, 978-1-558-60363-9. URL <http://dl.acm.org/citation.cfm?id=1625855.1625914>.
- Rifat Shahriyar, Md. Faizul Bari, Gourab Kundu, Sheikh Iqbal Ahamed, and Md. Mostofa Akbar. **Electronic Healthcare : Second International ICST Conference, eHealth 2009, Istanbul, Turkey, September 23-15, 2009, Revised Selected Papers**, chapter Intelligent Mobile Health Monitoring System (IMHMS), pages 5–12. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-11745-9. doi : 10.1007/978-3-642-11745-9\_2. URL [http://dx.doi.org/10.1007/978-3-642-11745-9\\_2](http://dx.doi.org/10.1007/978-3-642-11745-9_2).
- P.C Sherimon, Vinu P.V, R. Krishnan, and Y Saad. Ontology driven analysis and prediction of patient risk in diabetes. **Canadian Journal of Pure and Applied Sciences**, 8(3) :3043–3050, October 2014a.
- P.C. Sherimon, P.V. Vinu, Reshmy Krishnan, Youssef Takroni, Yousuf AlKaabi, and Yousuf Al-Fars. Adaptive questionnaire ontology in gathering patient medical history in diabetes domain. In Tutut Herawan, Mustafa Mat Deris, and Jemal Abawajy, editors, **Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)**, volume 285 of **Lecture Notes in Electrical Engineering**, pages 453–460. Springer Singapore, 2014b. ISBN 978-981-4585-17-0. doi : 10.1007/978-981-4585-18-7\_51. URL [http://dx.doi.org/10.1007/978-981-4585-18-7\\_51](http://dx.doi.org/10.1007/978-981-4585-18-7_51).
- Vinu P.V. Krishnan R. Takroni Y Sherimon, P.C. Ontology based system architecture to predict the risk of hypertension in related diseases. **International Journal of Information Processing and Management**, 4(4) :44 – 50, 2013.
- E. H. Shortliffe and J. J. Cimino. *Biomedical informatics : computer applications in health care and biomedicine*. 2006.
- Edward H. Shortliffe. **Computer-Based Medical Consultations : MYCIN**. New York, 1976.
- Barkun J Puhan MA Clavien PA. Slankamenac K, Graf R. The comprehensive complication index a novel continuous scale to measure surgical morbidity. **Annals of Surgery**, 258(1) :1–7, 2013.
- John F. Sowa. Top-level ontological categories. **International Journal of Human-Computer Studies**, 43(5–6) :669–685, 1995.
- S. Andrew Spooner. **Mathematical Foundations of Decision Support Systems**, pages 23–43. Springer New York, New York, NY, 2007. ISBN 978-0-387-38319-4. doi : 10.1007/978-0-387-38319-4\_2. URL [http://dx.doi.org/10.1007/978-0-387-38319-4\\_2](http://dx.doi.org/10.1007/978-0-387-38319-4_2).

- Rudi Studer, V.Richard Benjamins, and Dieter Fensel. Knowledge engineering : Principles and methods. **Data & Knowledge Engineering**, 25(1–2) :161–197, 1998.
- Karin A. Thursky and Michael Mahemoff. User-centered design techniques for a computerised antibiotic decision support system in an intensive care unit. **International Journal of Medical Informatics**, 76(10) :760 – 768, 2007. ISSN 1386-5056. doi : <http://dx.doi.org/10.1016/j.ijmedinf.2006.07.011>. URL <http://www.sciencedirect.com/science/article/pii/S1386505606002012>.
- Karin A. Thursky, Kirsty L. Buising, Narin Bak, Lachlan Macgregor, Alan C. Street, C. Raina Macintyre, Jeffrey J. Presneill, John F. Cade, and Graham V. Brown. Reduction of broad-spectrum antibiotic use with computerized decision support in an intensive care unit. **International Journal for Quality in Health Care**, 18(3) :224–231, 2006. ISSN 1353-4505. doi : [10.1093/intqhc/mzi095](https://doi.org/10.1093/intqhc/mzi095). URL <http://intqhc.oxfordjournals.org/content/18/3/224>.
- Mike Uschold and Michael Gruninger. Ontologies : Principles, methods and applications. **KNOWLEDGE ENGINEERING REVIEW**, 11 :93–136, 1996.
- Hsien-Tseng Wang and A.U. Tansel. Medcase : A template medical case store for case-based reasoning in medical decision support. In **Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on**, pages 962–967, Aug 2013.
- Ian Watson. **Applying Case-Based Reasoning : Techniques for Enterprise systems**. Number 1 in The Morgan Kaufmann Series in Artificial Intelligence. 1997.
- Ian Watson. Is cbr a technology or a methodology ? In **Tasks and Methods in Applied Artificial Intelligence**, volume 1416 of **Lecture Notes in Computer Science**, pages 525–534. Springer Berlin Heidelberg, 1998.
- Xiaomei Wu, Erli Pang, Kui Lin, and Zhen-Ming Pei. Improving the measurement of semantic similarity between gene ontology terms and gene products : insights from an edge- and ic-based hybrid method. **PloS one**, 8(5) :e66745, January 2013. ISSN 1932-6203. doi : [10.1371/journal.pone.0066745](https://doi.org/10.1371/journal.pone.0066745). URL <http://dx.plos.org/10.1371/journal.pone.0066745>.
- Zhibiao Wu and Martha Palmer. Verbs semantics and lexical selection. In **Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics**, ACL '94, pages 133–138, Stroudsburg, PA, USA, 1994. Association for Computational Linguistics. doi : [10.3115/981732.981751](https://doi.org/10.3115/981732.981751). URL <http://dx.doi.org/10.3115/981732.981751>.
- A. Strauss Z. El Balaa and P. Uziel. Fm-ultranet : a decision support system using case-based reasoning, applied to ultrasonography. In **Workshop on Case-Based Reasoning in the Health Sciences**, Trondheim, Norway, June 2003.
- T. Ralph Z. El Balaa. Case-based decision support and experience management for ultrasonography. In **Experience Management. GWEM'03**, ., German Workshop, 2003.



## Résumé :

Afin d'accompagner les professionnels de santé dans leur démarche clinique, plusieurs systèmes de suivi et de prise en charge médicale ont été construits et déployés dans le milieu hospitalier. Ces systèmes permettent principalement de collecter des données médicales sur les patients, de les analyser et de présenter les résultats de différentes manières. Ils représentent un appui et une aide aux professionnels de santé dans leur prise de décision par rapport à l'évolution de l'état de santé des patients suivis. L'utilisation de tels systèmes nécessite systématiquement une adaptation à la fois au domaine médical concerné et au mode d'intervention. Il est nécessaire, dans un milieu hospitalier, que ces systèmes puissent s'adapter et évoluer d'une manière simple, en limitant toute maintenance corrective ou évolutive. Ils doivent être en mesure de prendre en compte dynamiquement des connaissances théoriques et empiriques du domaine issues des experts médicaux.

Afin de répondre à ces exigences, nous avons proposé une approche pour la construction d'un système d'aide à la décision médicale capable de s'adapter au domaine médical concerné et au mode d'intervention approprié pour assister les professionnels de santé dans leur démarche clinique. Cette approche permet notamment l'organisation de la collecte des données médicales, en tenant compte du contexte du patient, la représentation des connaissances du domaine à base d'ontologies ainsi que leur exploitation associée aux guides de bonnes pratiques et à l'expérience clinique.

Dans la continuité des travaux précédemment réalisés au sein de notre équipe de recherche, nous avons choisi d'enrichir, avec notre approche, la plateforme E-care qui est dédiée au suivi et à la détection précoce de toute anomalie de l'état de patients atteints de maladies chroniques. Nous avons pu ainsi adapter facilement la plateforme E-care aux différentes expérimentations qui ont été menées notamment dans des EPHAD de la Mutualité Française en Anjou-Mayenne, au CHU de Haute-pierre et au CHUV à Lausanne.

Les résultats de ces expérimentations ont montré l'efficacité de l'approche proposée. L'adaptation de la plateforme par rapport au domaine et au mode d'intervention de chacune de ces expérimentations se limite à de la simple configuration. De plus, l'approche proposée a suscité l'intérêt du personnel médical par rapport à l'organisation de la collecte des données, qui tient compte du contexte du patient, et par rapport à l'exploitation des connaissances médicales qui apporte aux professionnels de santé une assistance pour une meilleure prise de décision.

**Mots-clés :** Système d'aide à la décision médicale, Ontologie, Acquisition de données, Raisonnement à base de cas, Raisonnement à base de règle, Suivi médical

The logo for SPIM (École doctorale SPIM) features a blue horizontal bar on the left, followed by the letters 'S', 'P', 'I', and 'M' in a large, white, sans-serif font.