

Université Saad Dahlab Blida

N° D'ordre :



Faculté des sciences
Département d'informatique
Mémoire Présenté par :

DAHMANI Meriem

DERADJI Kahina Afaf

En vue d'obtenir le diplôme de master
Domaine : mathématique et informatique

Filière : Informatique
Spécialité : Informatique
Option : Ingénierie de logiciel

***Thème : Inférence dans les ontologies probabilistes : étude de cas pour la
prédiction du diagnostic médical.***

Soutenu le :

M : S. OUKID Président

M : J. CHERFA Examineur

M : Examineur

M : Fareh M. Promotrice

Promotion
2016/2017



Dédicaces

A :

Nos chers parents

Nos frères et sœurs

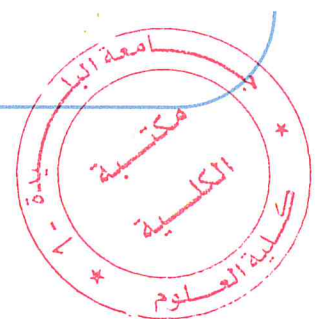
Toute Notre famille

Tous nos amis

On dédie ce mémoire

DAHMANI MERIEM

DERADJI KAHINA AFAF



Remerciement

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

*En second lieu, Nous tenons à remercier notre promotrice Mme : **FAREH.M** enseignante à l'université USDB, pour son aide et de nous avoir soutenues durant toute la période du travail.*

*Nous remercions le doctorant Mr **RIALI Ishak** pour nous avoir aidés et suivis durant la réalisation de ce projet*

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre travail en acceptant d'examiner et de l'enrichir par leurs propositions.

*Nous tenons à remercier également Dr **BENSALEM. S.M** médecin spécialiste en oncologie médicale au niveau de la clinique fatema al Azhar.*

Enfin, nous adressons nos plus sincères remerciements à nos familles : Nos parents, nos sœurs, nos frères, nos grands-parents et tous nos proches et amis, qui nous ont accompagnés, aidé, soutenu et encouragé tout au long de la réalisation de ce mémoire.

DAHMANI Meriem

DERADJI Kahina Afaf

Résumé :

Le domaine médical est basé sur les diagnostics ou les connaissances incertaines. Malgré le succès ind disputable des ontologies les ontologies classiques où ont été largement utilisés pour modéliser et faire des raisonnements avec les connaissances dans plusieurs domaines. Cependant, les ontologies classiques ne fournissent pas un support adéquat pour faire face à la connaissance incertaine. Dans l'autre côté, les ontologies probabilistes jouent un rôle important dans la représentation de l'incertitude.

Notre objectif est de concevoir un système de prédiction du diagnostic médical sur des connaissances incertaines en utilisant les ontologies probabilistes, nous appliquons dans ce mémoire la méthodologie UMP-ST qui sert à modéliser une ontologie probabiliste à partir d'une ontologie classique « Liver_Cancer » qui présente notre cas d'étude.

Mots clé : ontologie, ontologie probabiliste, OWL, PR-OWL, MEBN, réseaux bayésiens, incertitude.

Abstract:

The medical domain is based on uncertain diagnoses or uncertain knowledge. In spite of the undisputable success of the classical ontologies where they have been widely applied to represent and reason with the knowledge in several domains, as a main defect of the classical ontologies cannot represent and reason with uncertain knowledge. On the other hand, the probabilistic ontologies play an important role for dealing with uncertain knowledge.

Our goal is to design a system for predicting medical diagnosis on uncertain knowledge using probabilistic ontologies, We apply in this paper the UMP-ST methodology which is used to model a probabilistic ontology from a classical ontology "Liver_Cancer" which presents our case study.

Key words: **ontology, OWL, PR-OWL, MEBN, Bayesian networks, uncertainty, probabilistic ontology**

ملخص :

على الرغم من النجاح الذي لا جدال فيه الذي حققته الأنطولوجيا الكلاسيكية حيث تم تطبيقها على نطاق واسع لتمثيل والاستدلال مع المعرفة في العديد من المجالات، العيب الرئيسي في الأنطولوجيا الكلاسيكية أنها لا يمكنها أن تقوم بتمثيل والاستدلال مع المعرفة غير مؤكدة. ومن ناحية أخرى، تلعب الأنطولوجيات الاحتمالية دوراً هاماً في التعامل مع المعرفة غير المؤكدة. هدفنا هو تصميم نظام للتنبؤ التشخيص الطبي على معرفة غير مؤكدة باستخدام أنطولوجي الاحتمالية.

نطبق في هذه الورقة منهجية UMP-ST التي تستخدم لنمذجة الأنطولوجيا الاحتمالية من الأنطولوجيا الكلاسيكية "Liver_Cancer". الذي يعرض دراستنا للحالة.

الكلمات الدالة: الأنطولوجيا ، الشبكات البيزية , OWL, MEBN, PR-OWL, الأنطولوجيا الاحتمالية

Sommaire

Introduction générale.....	1
Contexte	1
Problématique.....	1
Objectif	2

Chapitre 1 : Ontologie classique et ontologie probabiliste

1. Introduction	4
2. Ontologie Classique	4
2.1. Définition.....	4
2.2. Les composants d'une ontologie.....	4
2.3. Classification des ontologies	5
2.3.1 Selon l'objet de la conceptualisation.....	5
2.3.2 Selon le niveau de granularité (détail).....	6
2.3.3 Selon le niveau de formalisation de la représentation	6
2.3.4 Selon le niveau de la représentation des connaissances.....	7
2.4. Rôles des ontologies	8
2.5. Les langages des ontologies.....	8
3. L'incertitude dans les ontologies	10
4. Ontologie Probabiliste.....	10
4.1. Définition.....	10
4.2. Objectifs d'ontologie probabiliste.....	10
4.3. Les langages et les approches des ontologies probabilistes	11
5. Matrice comparative	13
6. Conclusion	15

Chapitre 2 : L'inférence dans les ontologies

1. Introduction	16
2. Logiques de Description	16
2.1. Définition.....	16
2.2. L'inférence dans la Logiques de Description.....	16
2.2.1. Inférences élémentaires sur la TBOX.....	17
2.2.2. Inférences élémentaires sur la ABOX	17
3. Les inférences dans les Ontologies probabilistes.....	18

3.1.	Les réseaux bayésiens	18
3.2.	Le rôle des réseaux bayésiens	19
3.3.	L'inférence dans les réseaux bayésiens	20
3.3.1.	Inférence exacte	20
3.3.2.	Inférence approchée.....	20
3.4.	Quelques domaines d'application.....	21
4.	Les ontologies dans le domaine médical	21
4.1.	L'ontologie en médecine.....	21
4.2.	La terminologie médicale	22
4.3.	Les ontologies médicales existantes	23
4.3.1.	Ontologie « Liver_Cancer »	23
5.	Conclusion	24

Chapitre 3 : La conception du système

1.	Introduction	25
2.	Motivation de l'étude de cas	25
3.	Architecture globale du système	25
4.	L'ontologie classique de l'étude de cas	27
4.1.	L'ontologie Liver_cancer existante.....	27
5.	La création de l'ontologie probabiliste	30
5.1.	Définition de la méthodologie UMP-ST	30
5.2.	Les exigences (case bleue)	31
5.2.1.	Les besoins (Goals)	31
5.3.1.	Les entités	32
5.3.2.	Les règles (Rules)	34
5.3.3.	Groupes (Groups).....	36
5.4.1.	Mapping	38
5.4.2.	Probabilités	42
6.	Résultat final de l'ontologie probabiliste	48
6.1.	La différence entre l'ontologie classique et l'ontologie probabiliste obtenue	48
7.	Conclusion	52

Chapitre 4 : La mise en œuvre du système

1.	Introduction	53
2.	Outils de développement de notre système	53
3.	La mise en œuvre du système	54
4.	La base de connaissance.....	55
5.	Les interfaces	56
5.1.	L'interface d'accueil	56
5.2.	L'interface de consultation	59
5.3.	L'interface de prédiction.....	62
6.	Conclusion	63
	Conclusion générale	64

La liste des Figures

Figure 1 :	Aperçu des concepts de la théorie PROWL MEBN [20]	11
Figure 2 :	Un exemple d'ontologie représenté à l'aide du MEBN [20]	12
Figure 3 :	Un exemple d'ontologie représentée à l'aide de Bayes OWL [20]	13
Figure 4 :	Exemple d'un réseau bayésien simple [26].....	18
Figure 5 :	L'architecture globale du système.	26
Figure 6 :	L'ontologie Liver_Cancer existante [45].	28
Figure 7 :	L'ontologie Liver_Cancer après les modifications.	29
Figure 8 :	Cycle de modélisation de l'ontologie probabiliste (POMC) [46].....	31
Figure 9 :	Les entités.	33
Figure 10 :	Diagramme de classe	37
Figure 11 :	Les MFrag du domaine.	39
Figure 12 :	Le MFrag de Patient.	40
Figure 13 :	Le MFrag de RiskFactors.....	41
Figure 14 :	Le MFrag des Symptoms.	41
Figure 15 :	Le MFrag de Diagnosis.	42
Figure 16 :	LPD pour le nœud IsTouchedBy (patient).	43
Figure 17 :	Résultat SSBN pour le nœud IsTouchedBy (patient).	44
Figure 18 :	SSBN généré pour le premier scénario.	47

Figure 19 : SSBN généré pour le second scénario.....	48
Figure 20 : La hiérarchie des classes de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).....	49
Figure 21 : La hiérarchie des propriétés d'objet de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).	50
Figure 22 : La hiérarchie des propriétés des données de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).	51
Figure 23 : Base de connaissances.	55
Figure 24 : Interface d'accueil.....	56
Figure 25 : L'interface « Modifier patient ».	57
Figure 26 : L'interface « Ajouter patient ».	58
Figure 27 : L'interface « Ajouter Facteurs de risque ».	58
Figure 28 : L'interface « Ajouter Symptômes ».	59
Figure 29 : L'interface « Consultation Générale ».	60
Figure 30 : L'interface «RiskFactors du patient».....	61
Figure 31 : L'interface «Symptoms du patient».....	61
Figure 32 : L'interface « Prédication ».	62
Figure 33 : L'interface « pourcentage ».	62

La liste des tableaux

Tableau 1 : La matrice comparative des approches [20].....	14
Tableau 2 : La description d'architecture globale du système.	27
Tableau 3 : Les attributs de l'entité patient.....	33
Tableau 4 : Les attributs de l'entité RickFactors.	34
Tableau 5 : Les attributs de l'entité Symptoms.	34
Tableau 6 : Les relations entre entités.....	34

La liste des abréviations

OWL: Ontology Web Language.

PR-OWL: PProbabilistic Ontology Web Language.

XML: eXtensible Markup Language.

RDF: Resource Description Framework.

RDFS: Resource Description Framework Schema.

KIF: Knowledge Interchange Format.

OIL: Ontology Interchange Language and Ontology Inference Layer.

DAML+OIL : Agent Markup Language.

OWL DL: Ontology Web Language Description Logic.

DL : Description Logic.

RB: Réseau Bayésien. DAG: directed acyclic graph.

MEBN: Multy Entity Bayesian Network.

SSBN: Situation-Specific Bayesian Network.

RV: Random Variable.

LPD: Local Probability Distribution.

UMP-ST: Uncertainty Modeling Process for Semantic Technology.

UP: Unified process.

Introduction générale

Contexte

Une ontologie classique (traditionnelle) est «une spécification explicite et formelle d'une conceptualisation partagée». Elle est une représentation qui regroupe un ensemble de concepts et relations décrivant un domaine particulier.

L'exploitation des connaissances médicales, la complexité et les incertitudes liées aux maladies contribuent à de grandes variations dans la pratique clinique, à des erreurs cliniques et à un non respect des guidelines. Il est tout simplement irréaliste de croire que les soins des maladies peuvent continuer à être délivrés sans l'adoption des sciences de l'information. Parmi ces sciences, les ontologies, et tout particulièrement les ontologies probabilistes, sont devenues l'un des domaines prometteurs en raison de leurs grandes capacités de raisonnement qu'elles pourraient fournir aux outils informatiques pour aider à un diagnostic médical proche de la réalité et du raisonnement humain.

Les ontologies probabilistes sont utilisées dans le but de décrire en détail les connaissances sur un domaine et l'incertitude associée à ces connaissances d'une manière raisonnée, structurée et partageable, idéalement dans un format qui peut être lu et traité par un ordinateur. Ils élargissent également les possibilités d'ontologies standards en introduisant l'exigence d'une représentation adéquate des régularités statistiques et les preuves incertaines sur les entités dans un domaine d'application.

A ce jour, tous les langages développés pour la représentation des ontologies probabilistes se basent sur les réseaux bayésiens. Les réseaux bayésiens sont des techniques très performantes en matière d'aide à la décision. Ils sont issues d'un mariage entre la théorie bayésienne des probabilités conditionnelles et la théorie des graphes. Ces techniques ont prouvé leur efficacité dans plusieurs domaines, tels que le domaine d'aide à la décision et le domaine de modélisation de systèmes complexes.

Problématique

Dans un domaine médical les données ou bien les connaissances d'une maladie sont incertaines et des fois complexe, d'où les ontologies classiques sont incapables d'intégrer et

Chapitre 1 :

Ontologie

classique et

ontologie

probabiliste

1. Introduction

Les ontologies classiques permettent de représenter les connaissances. Elles sont des représentations qui regroupent un ensemble des concepts et des relations décrivant un domaine particulier. Un des principaux défauts de l'ontologie est l'incapacité de représenter et de raisonner sur l'incertitude, pour cela l'utilisation des ontologies probabilistes est nécessaire.

2. Ontologie Classique

2.1. Définition

Dans le cadre de l'intelligence artificielle, Neeches et ses collègues [1] furent les premiers à proposer une définition à savoir : «une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire».

En 1993, Gruber [2] propose la définition suivante : «spécification explicite d'une conceptualisation» qui est jusqu'à présent la définition la plus citée dans la littérature en intelligence artificielle.

2.2. Les composants d'une ontologie

Les connaissances dans les ontologies sont principalement formalisées en utilisant cinq types de composants :

- **Les concepts (ou classes)** : Un concept est un constituant de la pensée (un principe, une idée, une notion abstraite) sémantiquement évaluable et communicable. L'ensemble des propriétés d'un concept constitue sa compréhension ou son intention et l'ensemble des êtres qu'il englobe son extension.
- **Les relations (ou propriétés)** : Elles représentent des interactions entre les concepts, elles permettent de construire des représentations complexes de la connaissance du domaine, elles établissent des liens sémantiques binaires, organisables hiérarchiquement.

- **Les fonctions** : Elles présentent des cas particuliers de relations dans lesquelles le n ème élément de la relation est unique pour les $n-1$ éléments précédents. Formellement, les fonctions sont définies telles que : $F : c_1 * c_2 * \dots * c_{n-1} * c_n$.

- **Les axiomes (ou règles)** : Les axiomes sont des expressions qui sont toujours vraies. Ils ont pour but de définir dans un langage logique la description des concepts et des relations permettant de représenter leur sémantique. Ils représentent les intentions des concepts et des relations du domaine et, de manière générale, les connaissances n'ayant pas un caractère strictement terminologique.

- **Les instances (ou individus)** : Elles constituent la définition extensionnelle de l'ontologie, elles sont utilisées pour représenter des éléments dans un domaine donné [3].

2.3. Classification des ontologies

Les ontologies peuvent être subdivisées en plusieurs dimensions, qui sont :

2.3.1 Selon l'objet de la conceptualisation

Dépendamment de leur objet de conceptualisation, les ontologies sont classifiées de la manière suivante :

- **Ontologie de représentation de connaissances** : Elle modélise les représentations primitives utilisées pour la formalisation des connaissances sous un paradigme donné.
- **Ontologie supérieure ou de Haut niveau** : Cette ontologie est une ontologie générale. Son sujet est l'étude des catégories des choses qui existent dans le monde, qui sont les concepts de haute abstraction tels que : les entités, les événements, les états, les processus, les actions, le temps, l'espace, les relations et les propriétés.
- **Ontologie Générique/ méta-ontologies** : Egalement appelée noyau ontologique qui modélise des connaissances moins abstraites que celles véhiculées par l'ontologie de haut niveau mais assez générales néanmoins pour être réutilisées à travers différents domaines
- **Ontologie du Domaine** : Cette ontologie exprime des conceptualisations spécifiques à un domaine, elle est réutilisable pour plusieurs applications de ce domaine. Elle

fournit les concepts et les relations permettant de couvrir les vocabulaires, activités et théories de ces domaines.

- **Ontologie de Tâches** : L'ontologie de tâches fournit un vocabulaire systématisé des termes employés pour résoudre des problèmes liés aux tâches qui peuvent être ou non du même domaine.
- **Ontologie d'application** : C'est l'ontologie la plus spécifique, elle contient des concepts dépendants d'un domaine et d'une tâche particulière, elle est spécifique et non réutilisable.
- **Ontologie de méthodes** : Ce type d'ontologie modélise les définitions des concepts et des relations pertinentes pour le processus de raisonnement afin d'effectuer une tâche spécifique [4].

2.3.2 Selon le niveau de granularité (détail)

Par rapport au niveau de détail utilisé lors de la conceptualisation de l'ontologie en fonction de l'objectif opérationnel envisagé pour l'ontologie, deux catégories au moins peuvent être identifiées [4] :

- **Granularité fine** : ce niveau correspond à des ontologies très détaillées, elles possèdent ainsi un vocabulaire plus riche capable d'assurer une description détaillée des concepts pertinents d'un domaine ou d'une tâche.
- **Granularité large** : ce niveau correspond à des vocabulaires moins détaillés.

2.3.3 Selon le niveau de formalisation de la représentation

Par rapport au niveau de formalisation de la représentation des connaissances de l'ontologie, on peut distinguer quatre sortes d'ontologies:

- **Informelle** : l'ontologie est exprimée en langage naturel (sémantique ouverte). Cela peut permettre de rendre plus compréhensible l'ontologie pour l'utilisateur mais peut rendre plus difficile la vérification de l'absence de redondances ou de contradictions.

- **Semi-informelle** : l'ontologie est exprimée dans une forme restreinte et structurée du langage naturel, cela permet d'augmenter la clarté de l'ontologie tout en réduisant l'ambiguïté.
- **Semi-formelle** : l'ontologie est exprimée dans un langage artificiel défini formellement.
- **Formelle** : l'ontologie est exprimée dans un langage artificiel disposant d'une sémantique formelle ainsi que des théorèmes permettant de prouver ses propriétés, l'intérêt d'une ontologie formelle est la possibilité d'effectuer des vérifications telles que la complétude, la non-redondance, la consistance, la cohérence, etc.

2.3.4 Selon le niveau de la représentation des connaissances

L'appellation de niveaux de représentation des connaissances ainsi que leur interprétation peuvent varier selon les auteurs. Nous décrivons les typologies de Mizoguchi [5].

Mizoguchi propose une typologie sur trois niveaux très axée sur le processus d'ingénierie ontologique :

- **Niveau 1 (ou niveau conceptuel)** : Il est spécifié en faisant abstraction de toute contrainte informatique et sert donc de support à l'acquisition des connaissances.,
- **Niveau 2 (ou niveau formel)** : Il est formalisé au moyen d'un langage de représentation interprétable par une machine ; en plus des définitions du niveau 1, les définitions formelles sont ajoutées pour empêcher des interprétations inattendues des concepts.
- **Niveau 3 (ou niveau opérationnel)** : Il est encodé au moyen d'un langage de programmation dont la spécification obtenue est exécutable.

2.4. Rôles des ontologies

Plusieurs chercheurs se sont intéressés à la finalité des exploitations des ontologies. Nous précisons dans ce qui suit les rôles et l'intérêt des ontologies au sein des systèmes à base de connaissances (SBC) et du Web Sémantique :

- **Les connaissances du domaine d'un SBC** : En particulier, elles servent de squelette à la représentation des connaissances du domaine dans la mesure où elles décrivent les objets, leurs propriétés et la façon dont ils peuvent se combiner pour constituer des connaissances du domaine complètes.

- **La communication** : Les ontologies peuvent intervenir dans la communication entre personnes, organisations et logiciels [7]. En effet, les ontologies servent par exemple, à créer au sein d'un groupe ou d'une organisation un « vocabulaire conceptuel commun ». Dans ce cas, on est plutôt dans le cadre d'une ontologie informelle.

- **L'interopérabilité** : L'interopérabilité est une spécialisation de la communication qui permet de répertorier les concepts que des applications peuvent s'échanger même si elles sont distantes et développées sur des bases différentes.

2.5. Les langages des ontologies

❖ KIF

Est un langage basé sur les prédicats du premier ordre avec des extensions pour représenter des définitions et des méta-connaissances. Tant que la logique du premier ordre est un langage de bas niveau pour l'expression d'ontologies, l'outil Ontolingua [9] permet aux utilisateurs de construire des ontologies KIF à un niveau plus élevé de la description par l'importation des définitions des ontologies prédéfinies.

❖ KL-ONE

Est un langage basé sur la logique de description [10]. Il est une formalisation de représentation de la connaissance à base de cadres (« frame-based »). Ce système maintient la définition des concepts par un simple nommage, et l'indication de la correspondance des concepts dans une hiérarchie de généralisation/spécialisation. De nouveaux termes peuvent être définis par des opérations de conjonction des concepts. De nouveaux rôles peuvent être introduits pour représenter les relations qui peuvent exister entre des individus dans le domaine modélisé. Les définitions des concepts peuvent inclure des restrictions sur les

valeurs possibles, sur les nombres de valeurs, ou sur le type de valeurs qu'un rôle peut avoir pour un concept.

❖ RDF et RDF Schéma

« Resource Description Framework » est un formalisme graphique pour représenter des méta-données. Il est basé sur la notion de triplet (sujet, prédicat, objet). RDF utilise la syntaxe XML, mais il ne donne aucune signification spécifique pour le vocabulaire comme sous classe de, ou le type. Les primitives de modélisation offertes par RDF sont très basiques. RDF Schéma [11] est un langage qui étend RDF avec un vocabulaire de termes et les relations entre ces termes. RDFS est reconnu comme un langage d'ontologie qui définit :

Des classes et des propriétés. Les sous-classes, les super-classes, les sous-propriétés, et les super-propriétés. Le domaine de définition et le domaine image des propriétés.

❖ DAML + OIL

Beaucoup de travaux ont été faits dans le domaine de la représentation des connaissances parmi lesquels on peut citer les plus importants : SHOE, OntoBroker [23], OIL [24], et encore DAML + OIL [25] qui a remplacé DAML – ONT. DAML + OIL est un langage construit sur des normes précédentes du W3C telles que RDF et RDF Schéma, et étend ces langages avec des primitives de modélisation plus riches. DAML+OIL a été conçu à partir du langage d'ontologie DAML-ONT [26] en vue de combiner plusieurs composants du langage OIL [24]. OIL « Ontology Inference Language » est une représentation basée sur le Web, et une couche d'inférence pour des ontologies

❖ OWL

Développé par le groupe de travail sur le Web Sémantique du W3C, OWL peut être utilisé pour représenter explicitement les sens des termes des vocabulaires et les relations entre ces termes. OWL vise également à rendre les ressources sur le Web aisément accessibles aux processus automatisés [16], d'une part en les structurant d'une façon compréhensible et standardisée, et d'autre part en leur ajoutant des méta-informations. Pour cela, OWL a des moyens plus puissants pour exprimer la signification et la sémantique que XML, RDF, et RDF-S [11]. De plus, OWL tient compte de l'aspect diffus des sources de connaissances et permet à l'information d'être recueillie à partir de sources distribuées, notamment en permettant la mise en relation des ontologies et l'importation des informations provenant explicitement d'autres ontologies [17].

3. L'incertitude dans les ontologies

L'approche probabiliste est un domaine qui essaie de trouver des mécanismes efficaces pour modéliser le raisonnement, tenant compte de l'incertitude de certaines connaissances. Dans ce domaine, les modèles graphiques probabilistes permettent de fournir un outil compact et expressif pour modéliser l'incertitude et la complexité. Ceux-ci joignent dans la même représentation la théorie des probabilités et la théorie des graphes [18].

4. Ontologie Probabiliste

4.1. Définition

Une ontologie probabiliste est une représentation de la connaissance formelle explicite qui exprime des connaissances sur un domaine d'application. Ceci comprend :

- Types d'entités qui existent dans le domaine;
- Propriétés de ces entités;
- Les relations entre les entités;
- Les processus et les événements qui se passent avec ces entités;
- Régularités statistiques qui caractérisent le domaine;
- Connaissances peu concluantes, ambiguës, incomplètes et peu fiables liés aux entités du domaine ;
- Incertitude sur toutes les formes de connaissance;

D'où, le terme entité se réfère à tout concept (réel ou fictif, concret ou abstrait) qui peut être décrit et raisonner selon le domaine d'application [19].

4.2. Objectifs d'ontologie probabiliste

Les ontologies probabilistes sont utilisées dans le but de décrire en détail les connaissances sur un domaine et l'incertitude associée à ces connaissances d'une manière raisonnée, structurée et partageable, idéalement dans un format qui peut être lu et traité par un ordinateur. Ils élargissent également les possibilités d'ontologies standards en introduisant l'exigence d'une représentation adéquate des régularités statistiques et les preuves incertaines sur les entités dans un domaine d'application [19].

4.3. Les langages et les approches des ontologies probabilistes

❖ PR-OWL

PR-OWL est une extension qui permet à une ontologie OWL de représenter des modèles probabilistes bayésiens complexes d'une manière suffisamment flexible pour être utilisé par divers outils probabilistes bayésiens (exemple : Netica, Hugin, Quiddity*Suite, JavaBayes, etc.) [8].

La figure-1- ci-dessous décrit un aperçu de concepts de la théorie PR-OWL MEBN [20].

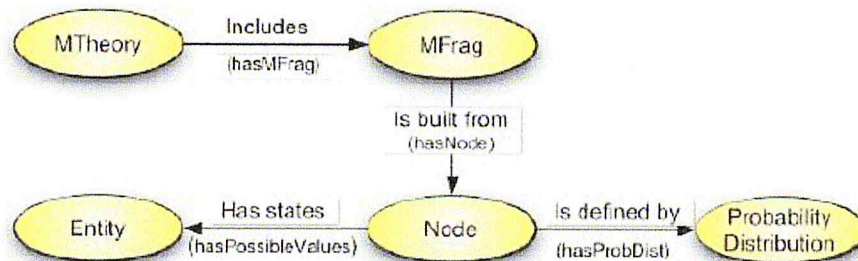


Figure 1 : Aperçu des concepts de la théorie PROWL MEBN [20]

❖ MEBN

MEBN (Multy Entity Bayesian Network) est un système logique qui intègre la logique du premier ordre (First Order Logic (FOL)) avec la théorie des probabilités bayésiennes, il étend les réseaux bayésiens ordinaires pour permettre la représentation des modèles graphiques avec des sous-structures répétées. La connaissance est codée comme une collection de fragments de réseaux bayésiens (MFrag) qui peuvent être instanciés et combinés pour former des réseaux bayésiens très complexes spécifiques à chaque situation. Une théorie MEBN (MTheory) représente implicitement une distribution de probabilité conjointe sur les chiffres éventuellement non bornées d'hypothèses et utilise l'apprentissage bayésien pour affiner une base de connaissances quand les observations s'accumulent. MEBN fournit une base logique pour la collection émergente des langages à base de probabilités très expressives [8].

Par exemple, La figure -2- ci-dessous a été prise d'un modèle MEBN sur le domaine de Star Trek représente la connaissance du degré du danger [20].

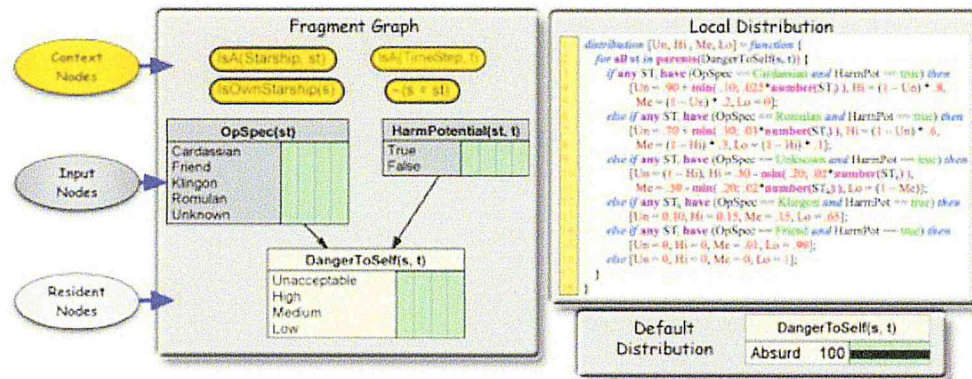


Figure 2 : Un exemple d'ontologie représenté à l'aide du MEBN [20]

❖ BayesOWL

Est un Framework probabiliste qui augmente et complète OWL pour représenter et raisonner avec l'incertitude basée sur les réseaux bayésiens (BN). Le modèle BayesOWL de base comprend un ensemble de règles de traduction structurelle pour convertir une ontologie OWL en un graphe acyclique dirigé (DAG) d'un BN et un mécanisme qui utilise des informations probabilistes disponibles pour construire une table de probabilité conditionnelle (CPT) pour chaque nœud du DAG. Pour aider à comprendre l'approche [20].

Exemple, la figure -3- ci-dessous donne un exemple de diagramme de gestion de l'incertitude en utilisant le cadre BayesOWL [20].

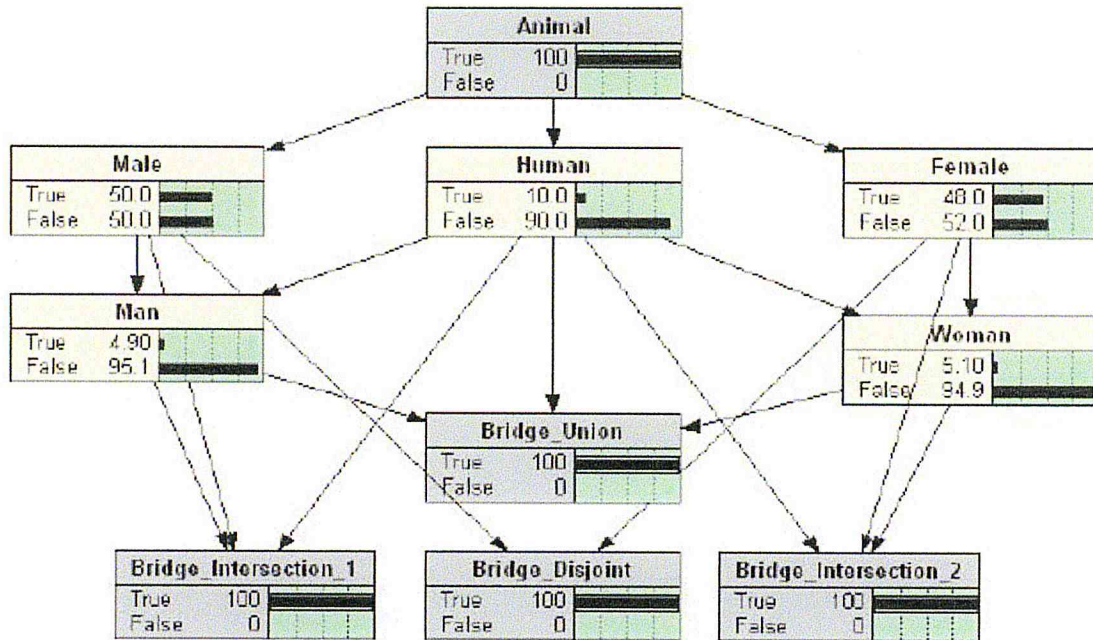


Figure 3 : Un exemple d'ontologie représentée à l'aide de Bayes OWL [20]

❖ Dempster-Shafer Theory

La théorie est venue à l'attention des chercheurs d'intelligence artificielle (AI) au début des années 1980, quand ils essayaient d'adapter la théorie des probabilités aux systèmes experts. Les degrés de croyance Dempster-Shafer ressemblent aux facteurs de certitude, et cette ressemblance a suggéré qu'ils pourraient combiner la rigueur de la théorie des probabilités avec la flexibilité des systèmes basés sur des règles [20].

5. Matrice comparative

Un aperçu est donné de l'analyse de chaque approche sur les métriques, leurs natures, leurs caractéristiques et leurs performances. Le tableau -1- ci-dessous résume les résultats analytiques [20].

Metrics Approaches	Complexity	Accuracy	Ease of Implementation	Reasoning Mechanism
BayesOWL	SIMPLE/ MODERATE	MEDIUM	EASY	AVAILABLE
MEBN	COMPLEX	HIGH	MEDIUM/HARD	AVAILABLE
PR-OWL	COMPLEX	HIGH	MEDIUM/HARD	AVAILABLE
Dempster- Shafer	MODERATE	MEDIUM	MEDIUM	AVAILABLE

Tableau 1 : La matrice comparative des approches [20].

Après la définition de chaque approche et langages nous avons pu choisir l'approche et le langage correspondant à notre projet qui est MEBN et PR-OWL.

Les théories MEBN utilisent des graphes dirigés pour spécifier des distributions de probabilité conjointes pour une collection de variables aléatoires connexes. Comme tous les langages basés sur la logique, MEBN peut représenter Les valeurs des fonctions n-ary et des relations d'une manière naturelle, il permet de construire un ensemble de modèle et des processus d'inférence et de définir les distributions de probabilité avec un nombre presque infini d'hypothèses liées entre elles.

Pour ce qui concerne la représentation de l'incertitude, le PR-OWL le fournit à l'aide des constructions OWL qui sont basées sur la logique de MEBN [36].

6. Conclusion

Dans ce chapitre nous avons présenté deux types d'ontologie classique et probabiliste qui permettent de représenter les connaissances ainsi que leurs caractéristique, leurs construction, leurs composants et leurs langages puis, nous avons introduit la notion d'incertitude et l'incapacité de l'intégrer dans les ontologies classiques qui nous conduisent à la modéliser dans les ontologies probabilistes.

Ce chapitre nous amène à la représentation des connaissances incertaines qui nécessite des outils puissants et qui permette de faire les inférences.

Chapitre 2 :

L'inférence

dans les

ontologies

1. Introduction

La logique de description permet de modéliser un domaine d'application ainsi que des services d'inférences, dans les réseaux bayésiens, l'inférence concerne le calcul de la probabilité. Il s'agit de déterminer les probabilités conditionnelles d'événements reliés par des relations d'influences.

Les ontologies sont devenues un moyen indispensable pour représenter et exploiter les données et les connaissances d'un domaine, et plus particulièrement celles du domaine médical. Les données et les connaissances médicales ont été structurées au moyen de taxonomies. Il existe plus de 100 classifications médicales, ce qui rend difficile l'exploitation des données codées par l'une ou l'autre.

2. Logiques de Description

2.1. Définition

Les logiques de description (LD) forment une famille de langages de représentation de connaissances qui peuvent être utilisés pour représenter les connaissances d'un domaine d'application d'une façon structurée et formelle. Une caractéristique fondamentale de ces langages est qu'ils ont une sémantique formelle. Les logiques de description sont utilisées pour de nombreuses applications.

Les logiques de description utilisent les notions de concept, de rôle et d'individu. Les concepts correspondent à des classes d'individus, les rôles sont des relations entre ces individus [23].

2.2. L'inférence dans la Logiques de Description

Un système de LD permet de modéliser un domaine d'application ainsi que des services d'inférence. Le raisonnement dans la LD permet de découvrir des connaissances implicites à partir des connaissances explicites.

Une base de connaissance de LD se compose de deux composants TBOX et ABOX, par conséquent nous pouvons avoir deux classes de services :

2.2.1. Inférences élémentaires sur la TBOX

- **Vérification de satisfiabilité** : un concept C est satisfiable (ou consistant) par rapport à TBOX T , s'il existe un modèle I de T tels que C n'est pas vide (I est un modèle de C).

- **Test de subsumption** : c'est l'inférence de base sur les expressions d'un concept, typiquement écrit $C \sqsubseteq D$ (D subsume C). La détermination de subsumption est un problème de vérification si un concept D (le subsumer) est considéré plus général que C (le subsumé).

En utilisant la relation de subsumption sur tous les couples possible de concepts contenant dans la TBOX, on atteint d'autre tâche ou service du raisonnement qui est la classification.

Cette tâche détermine l'endroit d'un concept par rapport aux autres concepts dans la hiérarchie.

- **Vérification d'équivalence** : deux concepts C et D sont équivalents (notée $C \equiv D$) par rapport à une TBOX T si et seulement si $C \sqsubseteq D$ pour tout modèle I de T .

- **Vérification de disjonction** : deux concepts C et D sont disjoints (notée $C \neq D$) par rapport à une TBOX T si et seulement si $C \sqcap D = \emptyset$ pour tout modèle I de T [24].

2.2.2. Inférences élémentaires sur la ABOX

Le raisonnement sur l'ABOX se concentre sur le test de correction d'un modèle d'un domaine d'application [25]. Nous avons deux tâches essentielles :

- **La vérification d'instance** : consiste à vérifier si un individu a d'une ABOX est une instance d'une description de concept donné C ($a \in C$).

- **La vérification de consistance** : cette tâche permet de vérifier s'il existe une interprétation pour l'ABOX et la TBOX en même temps.

3. Les inférences dans les Ontologies probabilistes

3.1. Les réseaux bayésiens

Un réseau bayésien est un système représentant la connaissance et permettant de calculer des probabilités conditionnelles apportant des solutions à différentes sortes de problématiques.

- Un réseau bayésien $B = (G, \theta)$ est défini par :
- Un graphe orienté sans circuit (DAG) $G = (V, E)$, où V est l'ensemble des nœuds de G , et E l'ensemble des arcs de G ;
- Un espace probabilisé fini (Ω, Z, p)
- Ensemble des variables aléatoires associées aux nœuds du graphe et définies sur (Ω, Z, p) tel que :

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa(X_i))$$

Où $Pa(X_i)$ est l'ensemble des causes (parents) de X_i dans le graphe G .

Un réseau bayésien est donc un graphe causal auquel ils ont associé une représentation probabiliste sous-jacente. Cette représentation permet de rendre quantitatifs les raisonnements sur les causalités que l'on peut faire à l'intérieur du graphe [26].

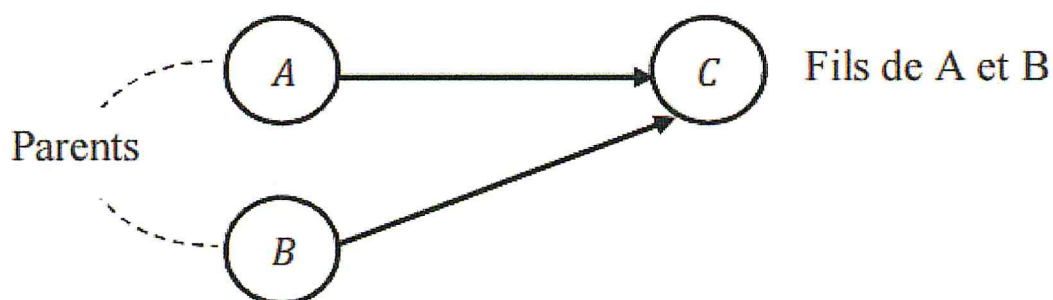


Figure 4 : Exemple d'un réseau bayésien simple [26]

3.2. Le rôle des réseaux bayésiens

Selon le type d'application, l'utilisation pratique d'un réseau bayésien peut être envisagée au même titre que celle d'autres modèles : réseau de neurones et arbre de décision. Naturellement, le choix de la méthode fait intervenir différents critères, comme la facilité, le coût et le délai de mise en œuvre d'une solution. En dehors de toute considération théorique, les aspects suivants des réseaux bayésiens les rendent préférables que d'autres modèles :

- ✓ Acquisition des connaissances. La possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle : retour d'expérience (données historiques ou empiriques), expertise (exprimée sous forme de règles logiques, d'équations, de statistiques ou de probabilités subjectives), observations. Dans le monde industriel, par exemple, chacune de ces sources d'information, quoique présente, est souvent insuffisante individuellement pour fournir une représentation précise et réaliste du système analysé.

- ✓ Représentation des connaissances. La représentation graphique d'un réseau bayésien est explicite, intuitive et compréhensible par un non spécialiste, ce qui facilite à la fois la validation du modèle, ses évolutions éventuelles et surtout son utilisation. Typiquement, un décideur est beaucoup plus enclin à s'appuyer sur un modèle dont il comprend le fonctionnement qu'à faire confiance à une boîte noire.

- ✓ Utilisation des connaissances. Un réseau bayésien est polyvalent : on peut se servir du même modèle pour évaluer, prévoir, diagnostiquer, ou optimiser des décisions, ce qui contribue à rentabiliser l'effort de construction du réseau bayésien.

- ✓ Qualité de l'offre en matière de logiciels. Il existe aujourd'hui de nombreux logiciels pour saisir et traiter des réseaux bayésiens. Ces outils présentent des fonctionnalités plus ou moins évoluées : apprentissage des probabilités, apprentissage de la structure du réseau bayésien, possibilité d'intégrer des variables continues, des variables d'utilité et de décision [27].

3.3. L'inférence dans les réseaux bayésiens

L'inférence dans un réseau bayésien concerne le calcul de la probabilité de n'importe quelle variable ou sous ensemble de variables à partir des autres variables observées. Il s'agit donc de déterminer les probabilités conditionnelles d'événement relié par des relations d'influence.

Les algorithmes d'inférence dans les réseaux bayésiens se répartissent en deux groupes : algorithmes *d'inférence exacte* et algorithmes *d'inférence approchée*. Les algorithmes d'inférence exacte exploitent les indépendances conditionnelles contenues dans le réseau pour calculer les probabilités a posteriori exactes [28], [29]. Concernant la deuxième catégorie d'algorithmes, les méthodes utilisées donnent des estimations approchées des probabilités à posteriori [30], [31].

3.3.1. Inférence exacte

La tâche de base de tout système d'inférence probabiliste consiste à calculer la distribution de probabilités a posteriori d'un ensemble de **variables de requête**, étant donné un **événement** observé autrement dit, une affectation de valeurs à un ensemble de **variables d'observation**.

X représente la variable de requête ; E représente l'ensemble de variable d'observation $E_1, \dots, E_m$, et e est un événement observé particulier ; Y les variables non observées $Y_1, \dots, Y_l$ (parfois nommées **variables cachées**). D'où l'ensemble complet de variables $X = \{X\} \cup E \cup Y$.

Une requête de type demande la distribution de probabilité à posteriori : $P(X|e)$ [32].

3.3.2. Inférence approchée

Etant donné que l'inférence exacte est impraticable dans de grands réseaux multiplement connectés, il est essentiel d'envisager des méthodes d'inférence approchée. Exemple l'algorithme d'échantillonnage aléatoire, également nommé algorithme de Monte-Carlo, qui fournit des réponses approchées dont l'exactitude dépend du nombre d'échantillons générés. Ces dernières années, les algorithmes de Monte-Carlo ont été largement utilisés en informatique pour estimer des quantités difficiles à calculer exactement [33].

3.4. Quelques domaines d'application

Il existe plusieurs domaines d'application des réseaux bayésiens dont :

- **Santé** : Les premières applications des réseaux bayésiens ont été développées dans le domaine du diagnostic médical.

Les réseaux bayésiens sont particulièrement adaptés à ce domaine parce qu'ils offrent la possibilité d'intégrer des sources de connaissances hétérogènes (expertise humain et données statiques), et surtout parce que leur capacité à traiter des requêtes complexes peuvent constituer une aide véritable et interactive pour le praticien.

Dans le domaine de la santé, une application intéressante des algorithmes issus des réseaux bayésiens a permis d'améliorer considérablement la recherche de la localisation de certains gènes, dans le cadre du projet Human Genom [26].

- **Informatique** : Les réseaux bayésiens sont probablement l'une des technologies les plus adaptées pour construire l'intelligence des agents. Ils assurent en effet ses différentes propriétés :

- **L'autonomie** : Représentée par la capacité des réseaux bayésiens de fournir des décisions en présence d'incertitude, ou en l'absence de certaines informations.

- **La motivation** : peut être représentée par certains types d'inférences, ou par un système de planification.

- **La réactivité** : c'est le principe même de l'inférence dans les réseaux bayésiens.

- **L'adaptation à l'environnement** : elle est rendue possible par les capacités d'apprentissage incrémental des réseaux bayésiens.

- **Gestion de connaissance** : Le domaine de la gestion des connaissances, qui connaît un intérêt croissant, est également un champ d'application potentiel pour les réseaux bayésiens, dans la mesure où ceux-ci offrent un formalisme riche et intuitif de représentation de la connaissance [26].

4. Les ontologies dans le domaine médical

4.1. L'ontologie en médecine

L'ontologie médicale est un modèle de la connaissance d'un domaine clinique. Il contient tous les concepts pertinents liés au diagnostic, au traitement, aux procédures cliniques et aux données sur les patients [37].

L'ontologie est associée avec des outils génériques, qui sont appropriés pour la modélisation des connaissances et de raisonnement, mais la plupart du temps, les définitions de la maladie dans la description logique dans une ontologie existante ne sont pas suffisantes pour classer les caractéristiques des patients sous une maladie particulière parce qu'ils ne formalisent pas des définitions opérationnelles des maladies (association des signes et symptômes = critères de diagnostic).

L'objectif principal de l'étude d'ontologie médicale est de proposer une représentation ontologique qui prend en compte les différents diagnostics sur lequel les conditions particulières de patients peuvent être classées dans une maladie spécifique. Cette méthode a besoin comme condition préalable une liste claire des différents diagnostics nécessaires et suffisantes tel que défini pour un bon nombre de maladies par les sociétés savantes [38].

4.2. La terminologie médicale

L'utilisation d'ontologies en médecine est principalement axée sur la représentation et le (ré) organisation des terminologies médicales. Les médecins ont développé leurs propres langues et lexiques spécialisés pour les aider à stocker et à communiquer efficacement les connaissances médicales générales et les informations relatives aux patients. Ces terminologies se caractérisent par une quantité significative de connaissances implicites. Les systèmes d'information médicale, d'autre part, doivent être capables de communiquer des notions médicales complexes et détaillés (éventuellement exprimés en différentes langues) sans ambiguïté. C'est évidemment une tâche difficile, nécessite une analyse approfondie de la structure et des concepts des terminologies médicales. Mais cela peut être réalisé en construisant des ontologies de domaine médical pour représenter des systèmes de terminologie médicale [39].

Les médecins emploient un vocabulaire spécifique qui compte entre 15000 et 20000 mots. L'étude de ses termes s'appelle la terminologie médicale. Un mot médical est une composition de plusieurs unités de sens différents, souvent dérivés de mots grecs et latins. Les unités composant le mot médical sont les radicaux, préfixes, suffixes Le radical et racine, ou racine est la partie centrale du mot. Son sens peut être précisé par un préfixe (= placé avant la racine) ou un suffixe (= placé après le radical). Les préfixes et les suffixes s'appellent les affixes. Le nombre de radicaux, de préfixes et de suffixes varie d'un mot à l'autre [40].

4.3. Les ontologies médicales existantes

Le développement des classifications médicales illustre la nécessité pour une communauté scientifique de disposer d'un vocabulaire de consensus. La multiplicité de ces classifications, particulièrement remarquable dans le domaine médical, illustre aussi le fait que ces besoins se sont manifestés dans différentes communautés, qui y ont répondu d'une manière opportuniste, en fonction de leurs moyens et de leurs besoins. Plusieurs classifications ont ainsi été trouvées, chacun répond à une catégorie de besoins, mais avec différentes structuration, bien que les concepts qu'elles définissent et présentent un taux de recouvrement important [41].

4.3.1. Ontologie « Liver_Cancer »

Le cancer du foie est un des cancers digestifs les plus agressifs. Mais c'est aussi un de ceux dont les traitements ont le plus progressé ces dernières années [45].

Le cancer du foie est le sixième cancer le plus courant au monde. Dans ce chapitre, on discute le diagnostic du cancer du foie qui contient des connaissances générales sur le foie.

Cette ontologie créée en 2015 par Aditya Khamparia et ses collègues, contient des différents facteurs de risque et symptômes de la maladie du cancer du foie.

A. Facteurs de risque :

Le cancer du foie est causé par une combinaison de différents facteurs de risque, qui sont le tabagisme, le diabète, le vieillissement, l'obésité ... etc [45].

B. Symptômes :

Les principaux symptômes de cancer du foie sont la fatigue, la jaunisse, les douleurs articulaires, la perte de poids, le foie élargie ... etc [45].

5. Conclusion

Dans ce chapitre nous avons focalisé sur les outils et les langages de représentation des connaissances et des ontologies d'où nous avons cité la logique de description et les inférences qui concernent les ontologies classiques et les réseaux bayésiens, qui concernent les ontologies probabilistes ainsi que les méthodes d'inférence les plus connues.

Le réseau bayésien multi entités (MEBN) est l'un des outils dans la représentation des connaissances incertaines qui concerne la prédiction dans le domaine médical, ce qui nous amène à choisir la méthodologie UMP-ST qui utilise cet outil pour modéliser ces connaissances et pour la construction des ontologies probabilistes qui sera l'objectif du troisième chapitre.

Chapitre 3 :

Conception du

système

1. Introduction

Le diagnostic médical est une tâche qui permet aux médecins de faire des prédictions sur une maladie selon les signes qui apparaissent sur un patient. Ces prédictions gèrent les connaissances incertaines qui doivent être modélisé dans une ontologie probabiliste, la construction de cette dernière se fait à l'aide de la méthodologie UMP-ST (Uncertainty Modeling Process for Semantic Technology).

2. Motivation de l'étude de cas

Le foie est un organe vital qui possède de nombreuses fonctions. Il fabrique la bile, filtre le sang, stocke le glucose et produit des substances indispensables au maintien de l'équilibre dans l'organisme. Le cancer du foie le plus fréquent est le carcinome hépatocellulaire ou hépato carcinome. Il se développe à partir des cellules spécialisées du foie, les hépatocytes [44].

Le choix de l'ontologie « Liver_Cancer » est à cause de l'importance de cette maladie dans le monde car le cancer du foie est le type de cancer le plus courant au monde et il cause majeure de décès (788 000 décès/ans), dont près de 80 % hommes [44].

3. Architecture globale du système

Notre conception consiste à faire les prédictions sur la maladie du cancer du foie dans une ontologie probabiliste. Qui était à la base une ontologie classique modélisé par la méthodologie UMP-ST.

La figure -5- suivante montre l'architecture globale de notre système :

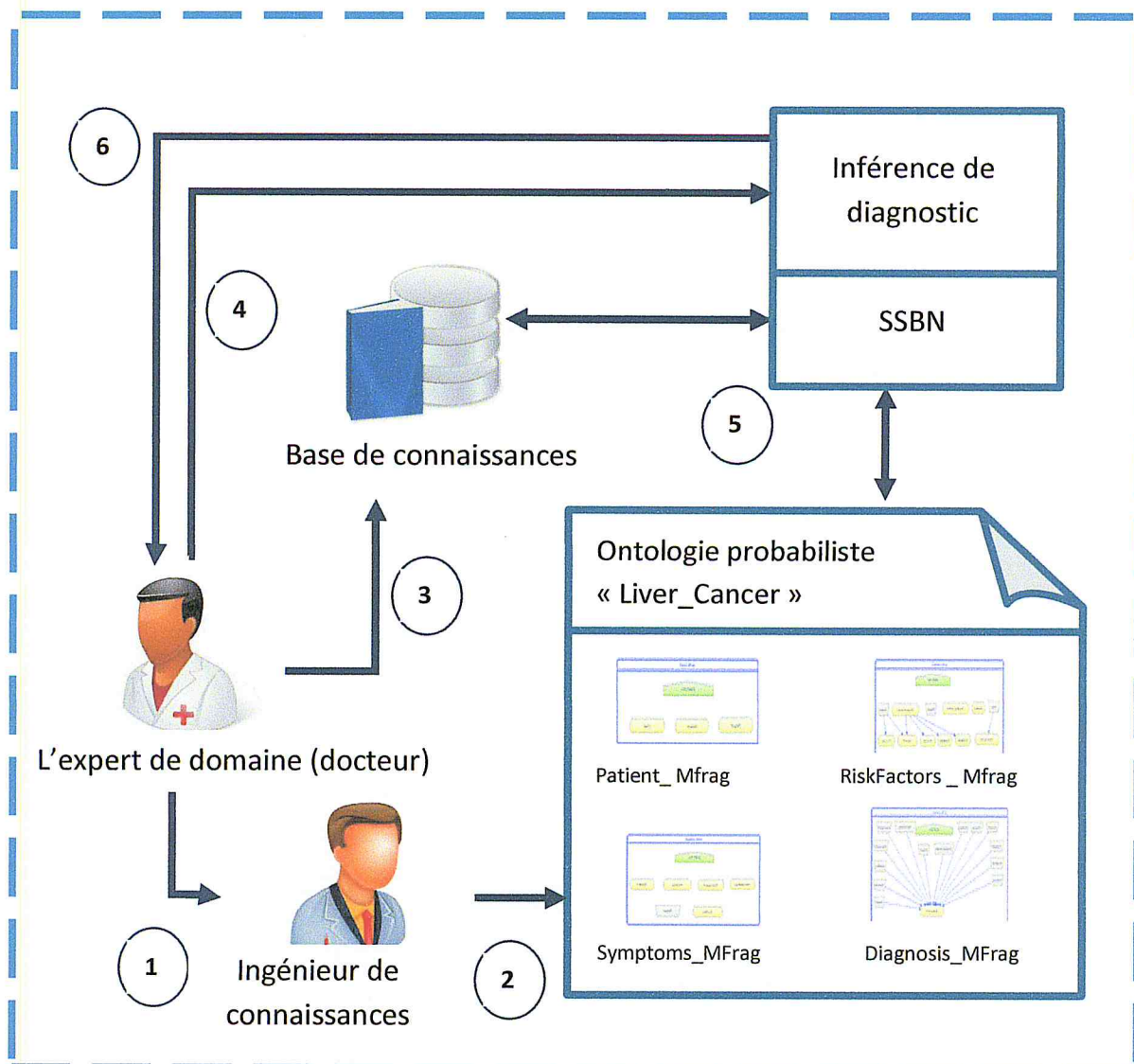


Figure 5 : L'architecture globale du système.

N	Description
1	L'expert de domaine enrichit les connaissances avec un ensemble de facteurs de risques et de symptômes relatives au cancer du foie.
2	L'Ingénieur construit un réseau bayésien multi entité, qui représente les composants de l'ontologie probabilistes en utilisant la méthodologie UMP-ST.
3	L'expert de domaine remplit la base de connaissance via une interface par des informations sur les patients, tel que ses symptômes.
4	L'expert de domaine lance une requête sur un patient dont il veut diagnostiquer.
5	Pour que l'inférence se fasse correctement, la présence des connaissances sur un patient donné « Base de connaissance » et de l'ontologie probabiliste est nécessaire, en générant le SSBN.
6	L'expert de domaine reçoit la réponse de l'inférence associée à sa requête sous une forme de probabilité.

Tableau 2 : La description d'architecture globale du système.

4. L'ontologie classique de l'étude de cas

Dans notre projet nous avons utilisé une ontologie « Liver_cancer ontology » déjà définie et nous l'avons modifié selon les discussions avec de l'expert.

4.1. L'ontologie Liver_cancer existante

La figure -6- ci-dessous représente l'ontologie classique de Liver_Cancer celle qui va être améliorée, en se basant sur des discussions avec l'expert Dr BENSALAM. S.M médecin spécialiste en oncologie médicale au niveau de la clinique fatema al Azhar pour valider les informations concernant les symptômes et les facteurs de risque de cette maladie.

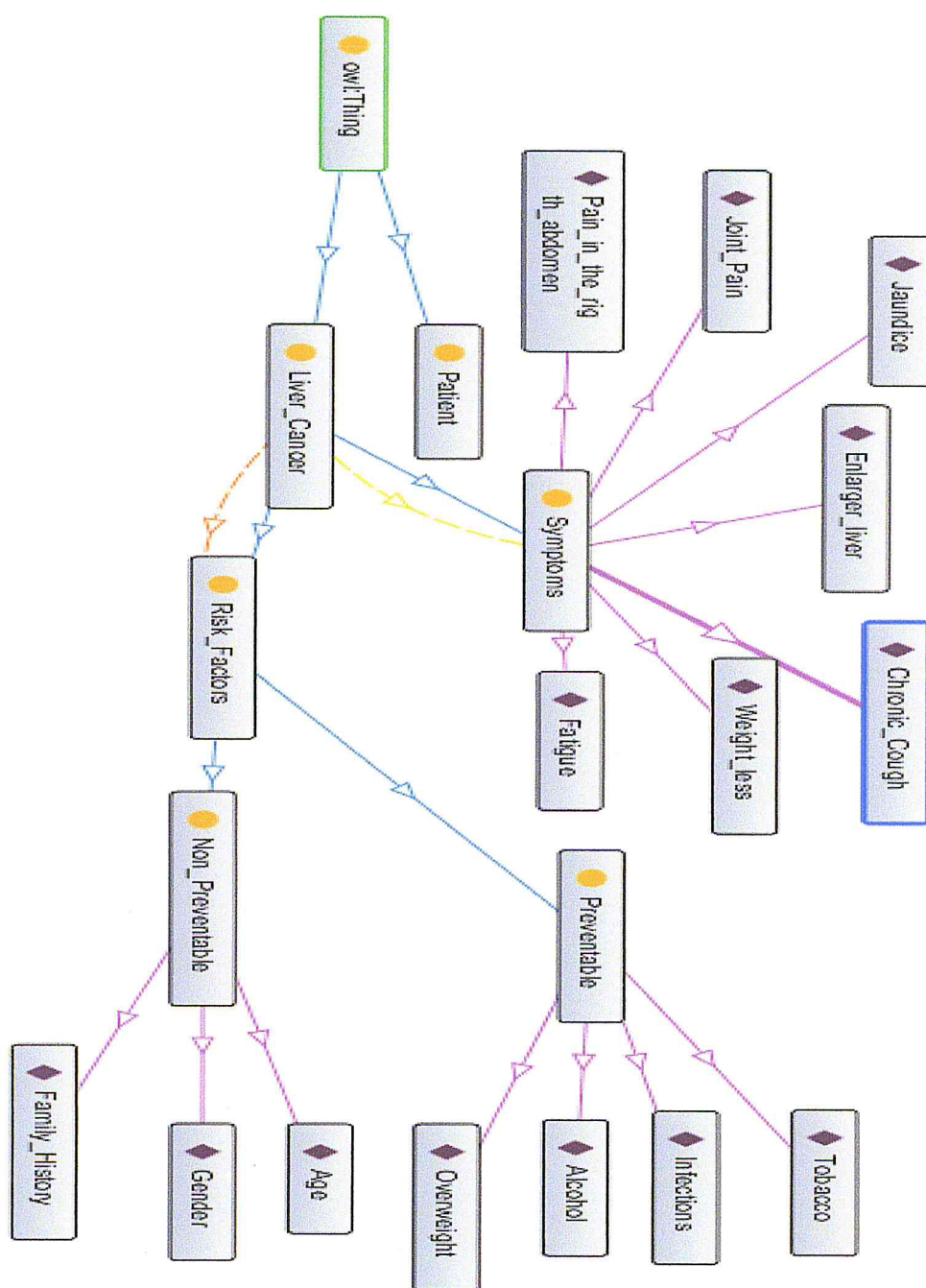


Figure 6 : L'ontologie Liver_Cancer existante [45].

4.2. Notre ontologie

Après avoir rendu visite à la clinique fatema al Azhar pour assurer la validité des notions existantes dans l'ontologie classique et d'ajouter quelques notions.

La figure-7- ci-dessous montre l'ontologie après modification :

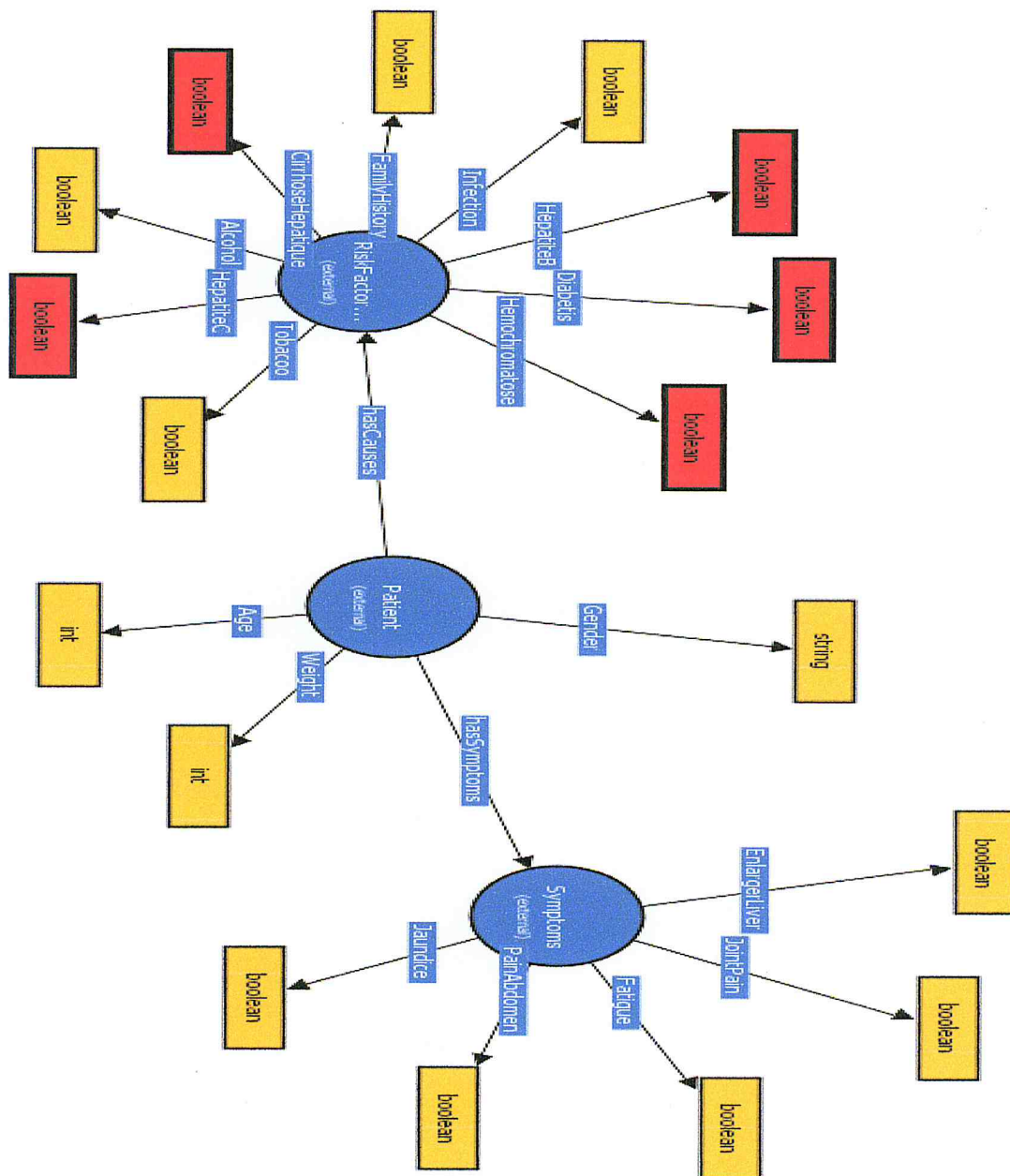


Figure 7 : L'ontologie Liver_Cancer après les modifications.

L'ontologie classique de Liver_Cancer mentionné précédemment contient la classe Patient avec les propriétés Age, Gender, Weight, la classes Symptoms avec les propriétés EnlargerLiver, Jaundice, fatigue, PainAbdomen et JointPain et la classe RiskFactors avec les propriétés Alcohol, Tobacco, Infection et FamilyHistory.

Les modifications apparaissent dans la classe de RiskFactors, en ajoutant les propriétés HepatiteC, HepatiteB, CirrhoseHepatique, Hemochromatose et Diabetis.

5. La création de l'ontologie probabiliste

L'objectif de la création de l'ontologie probabiliste est de faire le diagnostic qu'un patient est probablement touché par le cancer du foie, ce diagnostic provient des connaissances incertaines que l'ancienne ontologie ne pouvait pas traiter.

5.1. Définition de la méthodologie UMP-ST

L'UMP-ST est un processus itératif et incrémental, basé sur le processus unifié (UP), pour la conception d'une ontologie probabiliste. Alors que l'UP sert de point de départ, UMP-ST s'inspire aux processus d'ingénierie de l'ontologie.

Cette méthodologie est applicable à tout problème nécessitant le développement d'une ontologie probabiliste, dans laquelle PR-OWL a été choisi comme langage ontologique en raison de son traitement complet de l'incertitude [46].

Le diagramme dans la figure -8- montre une vision générale des différentes étapes du processus de la méthodologie

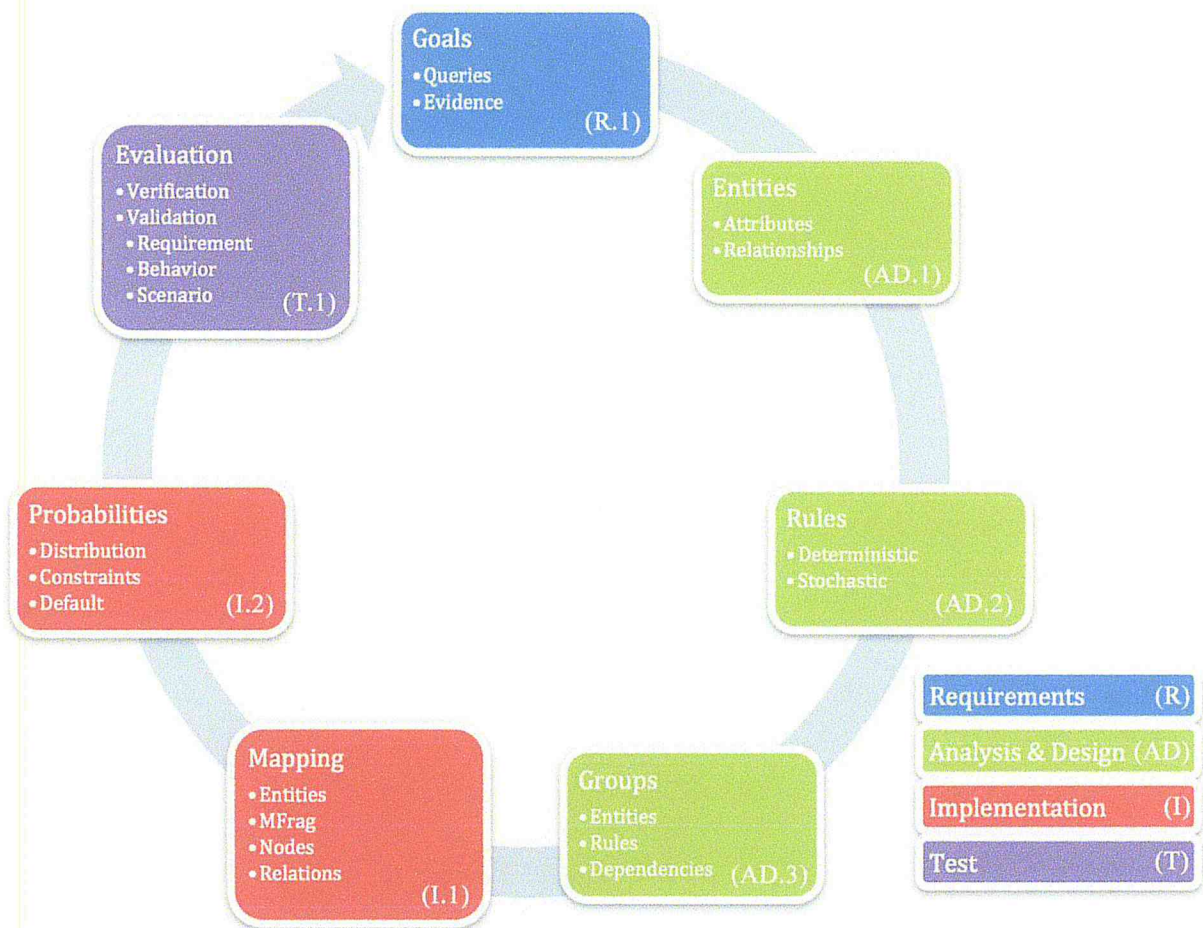


Figure 8 : Cycle de modélisation de l'ontologie probabiliste (POMC) [46].

5.2. Les exigences (case bleue)

Le POMC commence par la discipline des exigences « Requirements », pour cette discipline, il est important de définir les questions que le modèle devrait répondre, c'est-à-dire les requêtes à poser au système en cours de conception. Pour chaque question, un ensemble d'informations (évidence) pouvant aider à répondre [46].

5.2.1. Les besoins (Goals)

Les besoins (Goals), veut dire les différents types de questions ou requêtes que l'utilisateur veut que le système réponde via le raisonnement de l'ontologie probabiliste (OP) [46].

L'ensemble des besoins, des requêtes et des preuves (évidence) suivants concernant notre domaine de raisonnement pour qu'un patient soit touché par le cancer du foie, sont :

- **Besoin (Goal) :** Identifier si le patient est touché par le cancer du foie ?

- **Requête 1 (Query) :**

- Le patient a-t-il les facteurs de risque de la maladie ?

- **Evidence :**

- Le patient à l'Age >30.
- Le patient est en surpoids.
- Le patient a le cancer du foie dans la famille (HistoryFamily).
- Le patient a la Cirrhose Hépatique.
- Le patient a l'Hépatite B.
- Le patient a l'Hépatite C.
- Le patient est Diabétique.
- Le patient est un fumeur.
- Le patient est accro à l'alcool.
- Le patient a des infections virales.

- **Requête 2 (Query):**

- Le patient a-t-il les Symptômes de la maladie ?

-**Evidence :**

- Le patient est toujours fatigué.
- Le patient est maigre.
- Le patient a des douleurs articulaires
- Le patient a des douleurs abdomens.
- Le patient a la jaunisse.
- Le patient a Hypertrophie du foie.

5.3. Analyse et Conception (cases vertes)

Cette étape décrit les classes d'entités, leurs attributs, les relations entre eux et les règles qui sont applicables dans notre domaine. Cette étape est indépendante du langage utilisé pour implémenter le modèle [46].

5.3.1. Les entités

C'est de décrire les classes d'entités, leurs attributs, les relations entre eux, pour que nous puissions établir les règles qui sont applicables dans notre domaine.

- **Les entités:** ce sont les classes de l'ontologie
- Patient.
- Symptoms.
- RiskFactors.

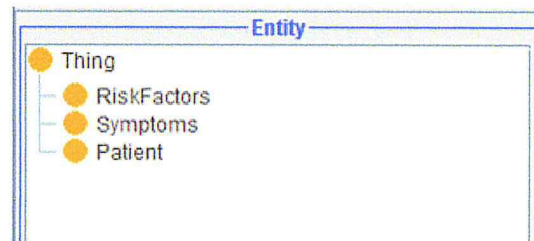


Figure 9 : Les entités.

- **Propriétés des entités (attributs) :**

Attributs	Domaine	Range
Gender	Patient	String
Weight	Patient	Int
Age	Patient	Int

Tableau 3 : Les attributs de l'entité patient.

Attributs	Domaine	Range
Alcohol	RiskFactors	Boolean
Tobacoo	RiskFactors	Boolean
FamilyHistory	RiskFactors	Boolean
Infection	RiskFactors	Boolean
HepatitisC	RiskFactors	Boolean

HepatitisB	RiskFactors	Boolean
CirrhosisHepatique	RiskFactors	Boolean
Hemochromatose	RiskFactors	Boolean
Diabetis	RiskFactors	Boolean

Tableau 4 : Les attributs de l'entité RiskFactors.

Attributes	Domaine	Range
Jaundice	Symptoms	Boolean
Fatigue	Symptoms	Boolean
EnlargerLiver	Symptoms	Boolean
PainAbdomen	Symptoms	Boolean
JointPain	Symptoms	Boolean

Tableau 5 : Les attributs de l'entité Symptoms.

- **Les relations entre entités (Relationships) :**

Relationship	Domaine	Range
hasCaused	Patient	RiskFactors
hasSymptoms	Patient	Symptoms

Tableau 6 : Les relations entre entités.

5.3.2. Les règles (Rules)

Généralement, les règles probabilistes sont décrites initialement en utilisant des énoncés de probabilité qualitative. La mise en œuvre d'une ontologie probabiliste nécessite de spécifier des probabilités numériques. Les valeurs de probabilité peuvent être obtenues auprès d'experts du domaine ou tirées de l'observation [46].

Les règles qui doivent être décrites lors de la mise en œuvre de notre modèle servent à calculer la probabilité (inférence), Ces règles conditionnelles ont été tirées d'une recherche approfondie sur le cancer du foie et des connaissances fournies par l'expert du domaine.

La liste suivante montre les règles conditionnelles :

A. Si le patient est un homme alors la probabilité qu'il soit touché par la maladie est de [true = 0.8, false = 0.2].

B. Si le patient est une femme alors la probabilité qu'elle soit touché par la maladie est de [true = 0.2, false = 0.8].

C. Si le patient a la Cirrhose Hépatique alors la probabilité qu'il soit touché par la maladie est de [true = 0.8, false = 0.2].

D. Si le patient a l'Hépatite B alors la probabilité qu'il soit touché par la maladie est de [true = 0.8, false = 0.2].

E. Si le patient a l'Hépatite C alors la probabilité qu'il soit touché par la maladie est de [true = 0.8, false = 0.2].

F. Si le patient est diabétique alors la probabilité qu'il soit touché par la maladie est de [true = 0.7, false = 0.3].

G. Si le patient a la Hémochromatose alors la probabilité qu'il soit touché par la maladie est de [true = 0.3, false = 0.7].

H. Si un patient est toujours fatigué, la probabilité qu'il soit touché par la maladie est très élevé [true = 0.9, false = 0.1].

I. Si le patient a le jaunissement alors la probabilité qu'il soit touché par la maladie est de [true = 0.87, false = 0.13]

J. Si le patient a un foie élargi alors la probabilité qu'il soit touché par la maladie est de [true = 0.85, false = 0.15].

K. Si le patient a des douleurs articulaire alors la probabilité qu'il soit touché par la maladie est [true = 0.83, false = 0.17].

L. Si le patient a des douleurs au ventre alors la probabilité qu'il soit touché par la maladie est de [true = 0.8, false = 0.2].

M. Si le patient a un surpoids alors la probabilité de quel soit touché par la maladie est de [true = 0.55, false = 0.45].

N. Si le patient est un homme et fumeur alors la probabilité de quel soit touché par la maladie est de [true = 0.82, false = 0.18].

O. Si le patient est une femme et fumeur alors la probabilité de quel soit touché par la maladie est de [true = 0.25, false = 0.75].

P. Si le patient est un homme et alcoolique alors la probabilité de quel soit touché par la maladie est de [true = 0.85, false = 0.15].

Q. Si le patient est une femme et alcoolique alors la probabilité de quel soit touché par la maladie est de [true = 0.35, false = 0.65].

R. Si le patient a la Cirrhose Hépatique et alcoolique alors la probabilité qu'il soit touché par la maladie est de [true = 0.85, false = 0.15].

S. Si le patient a la Cirrhose Hépatique et il a une infection alors la probabilité qu'il soit touché par la maladie est de [true = 0.9, false = 0.1].

T. Si le patient a la Cirrhose Hépatique et l'Hépatite B/C alors la probabilité qu'il soit touché par la maladie est de [true = 0.95, false = 0.05].

U. Si le patient a la Cirrhose Hépatique, une infection, l'Hépatite B/C et alcoolique alors la probabilité qu'il soit touché par la maladie est de [true = 0.99, false = 0.01].

V. Si le patient est âgé et il a le cancer du foie dans la famille alors la probabilité qu'il soit touché par la maladie est de [true = 0.68, false = 0.42].

5.3.3. Groupes (Groups)

Dans cette étape l'utilisation de diagrammes de classe UML est pour identifier les entités principales, les attributs et les relations pour le modèle [46].

La figure -10- au-dessous montre le diagramme de classe de notre modèle:

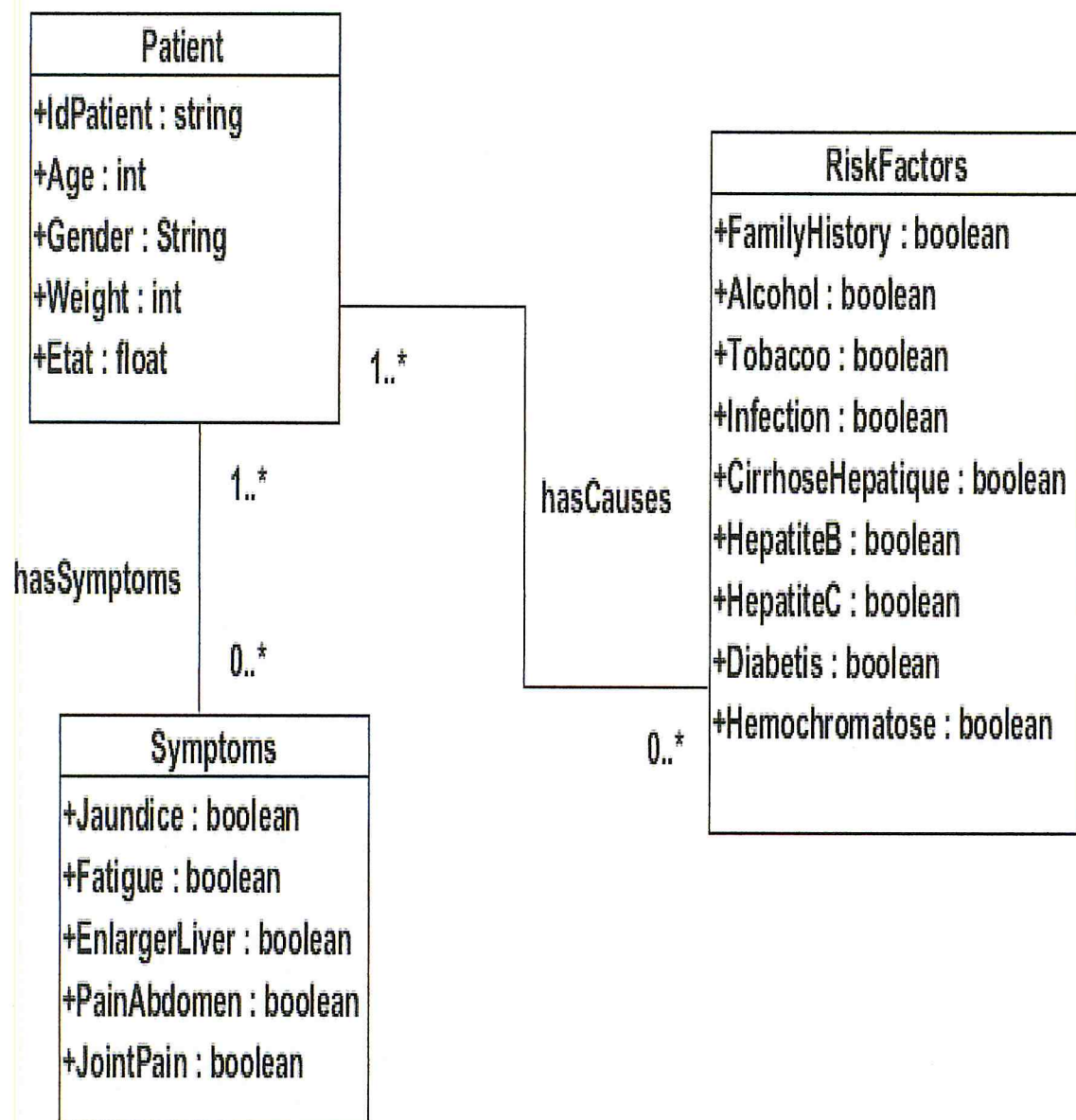


Figure 10 : Diagramme de classe

5.4. Implémentation (cases rouges)

Une fois que nous avons terminé la partie d'Analyse & conception, il est temps de commencer à implémenter notre modèle dans un langage spécifique. Cette section décrit comment modéliser l'ontologie probabiliste Liver_Cancer à partir de l'ancienne ontologie classique de la même maladie dans PR-OWL en utilisant UnBBayes.

La première chose à faire est de commencer à implémenter les entités, leurs attributs et leurs relations avec PR-OWL, qui utilise essentiellement des termes MEBN. Cette partie est différente des précédentes, car cela dépend du langage utilisé [46].

5.4.1. Mapping

Le mapping nous permet d'utiliser PR-OWL pour raisonner de façon probabiliste sur des connaissances incertaines d'une ontologie existante basée sur les connaissances déjà disponibles [46].

A. MTheory

Un MTheory est un ensemble de MFrag qui satisfait collectivement des contraintes de cohérence assurant l'existence d'une distribution unique de probabilité conjointe sur les instances des variables aléatoires représentées dans chacune des MFrag dans l'ensemble [46].

La création de la Mtheory est la première étape pour construire une ontologie probabiliste, il se fait juste après l'importation de l'ontologie qu'on veut modéliser « LiverCancer », ce MTheory va contenir les MFrag, dont notre ontologie probabiliste aura besoin.

B. MFrag

Un MFrag représente une distribution de probabilité conditionnelle pour les cas de ses variables aléatoires résidentielles données à leurs parents dans le graphique des fragments et les nœuds contextuels [46].

Après la création de notre MTheory, nous commençons par la création des MFrag qui contient les nœuds (nœuds résidents, nœuds de contexte, nœuds d'entrées) qui sont les variables aléatoires.

La figure-11- ci-dessous montre les MFrag de domaine :



Figure 11 : Les MFrag du domaine.

- **variables ordinaires :**

Pour la déclaration des variables ordinaires on aura besoin des classes OWL que nous avons déjà extrait. Une variable ordinaire est déclarée avec la relation 'est-un' (*IsA*), c'est une instantiation d'une entité (classe OWL). Par exemple : *IsA (p, Patient)* tels que *p* est notre variable ordinaire et *Patient* est l'entité (la classe OWL).

- **ResidentNode (nœuds résidents) :**

Ce sont les variables aléatoires réelles qui forment le cœur du sujet d'un MFrag. La logique MEBN exige que la distribution probabiliste locale de chaque nœud résident doit être unique et explicitement défini dans son MFrag original. Les valeurs possibles d'un nœud résident peuvent être une entité existante. [34].

- **ContextNode (nœud de contexte) :**

Ce sont des variables aléatoires booléennes représentant des conditions qui doivent être satisfaites pour faire une distribution dans un MFrag valide. Les nœuds de contexte peuvent représenter plusieurs types de modèles d'incertitude sophistiqués, tels que l'incertitude sur les relations entre les entités. Si on peut inférer de la base de connaissances (c'est à dire l'ontologie) que la valeur d'un nœud de contexte est vraie, la distribution de probabilité de l'MFrag sera appliquée dans le modèle d'inférence. Si cette valeur est fausse, une distribution par défaut sera utilisée. Si la valeur est inconnue, le nœud de contexte devient pratiquement un parent de tous les nœuds résidents dans le même MFrag [34].

- **InputNode (nœud d'entre) :**

Ces nœuds sont essentiellement des «pointeurs» faisant référence aux nœuds résidents d'un autre MFrag. Les nœuds d'entrée fournissent également un mécanisme permettant la réutilisation des nœuds résidents entre les MFrag. Les nœuds d'entrée influencent la distribution de probabilité des nœuds résidents qui sont leurs enfants dans un MFrag donné, mais leurs propres distributions sont définis ailleurs (c'est à dire dans leurs propres MFrag d'origine). Dans une MTheory complète, chaque nœud d'entrée doit pointer vers un nœud résident dans un MFrag [34].

C. Les Types des relations du réseau bayésien :

Le type de relation entre les nœuds qui vont être inclus dans la structure du réseau est de type causal, on distingue deux types de nœud dans une relation de causalité :

- ✓ Relation de causalité entre les nœuds d'entrées avec des nœuds résidents.
- ✓ Relation de causalité entre les nœuds résidents avec d'autres nœuds résidents

[34].

Après les définitions des différents nœuds de MFrag nous allons montrer chaque MFrag en détails :

- MFrag de Patient_MFrag contient les différents attributs de patient selon les besoin dans notre domaine :

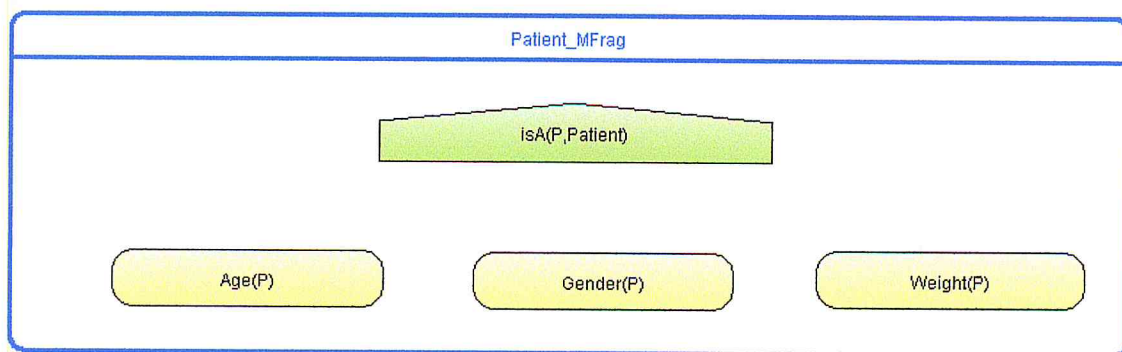


Figure 12 : Le MFrag de Patient.

- MFragment de RiskFactors_MFragment contient les différents facteurs de risque concernant le cancer du foie:

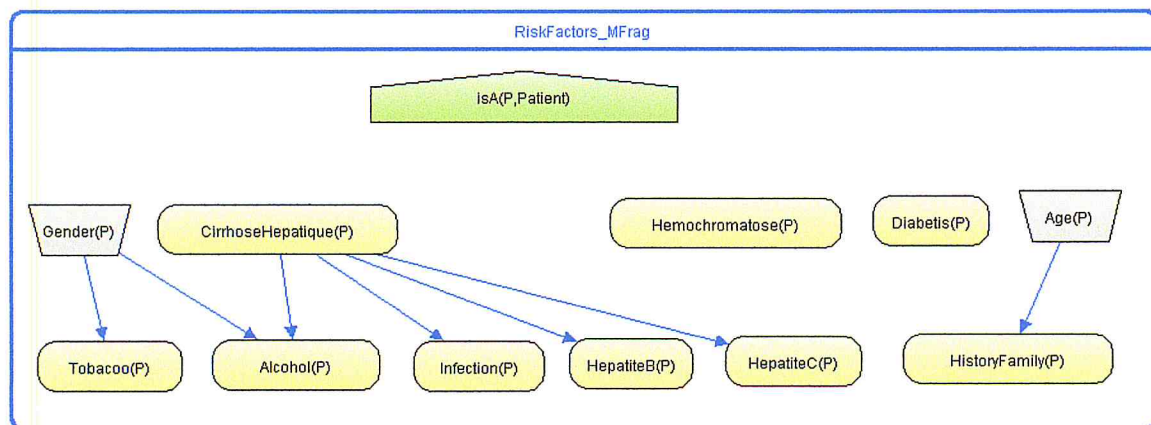


Figure 13 : Le MFragment de RiskFactors.

- MFragment de Symptoms_MFragment contient les différents symptômes concernant le cancer du foie :

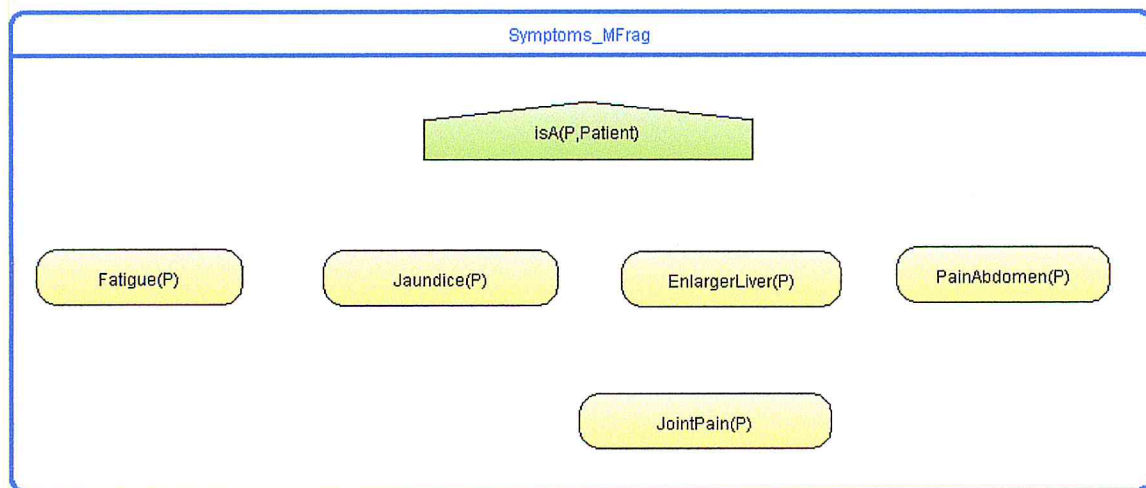


Figure 14 : Le MFragment des Symptoms.

- MFragment de Diagnosis_MFragment contient tous les symptômes et les facteurs de risque concernant le cancer du foie :

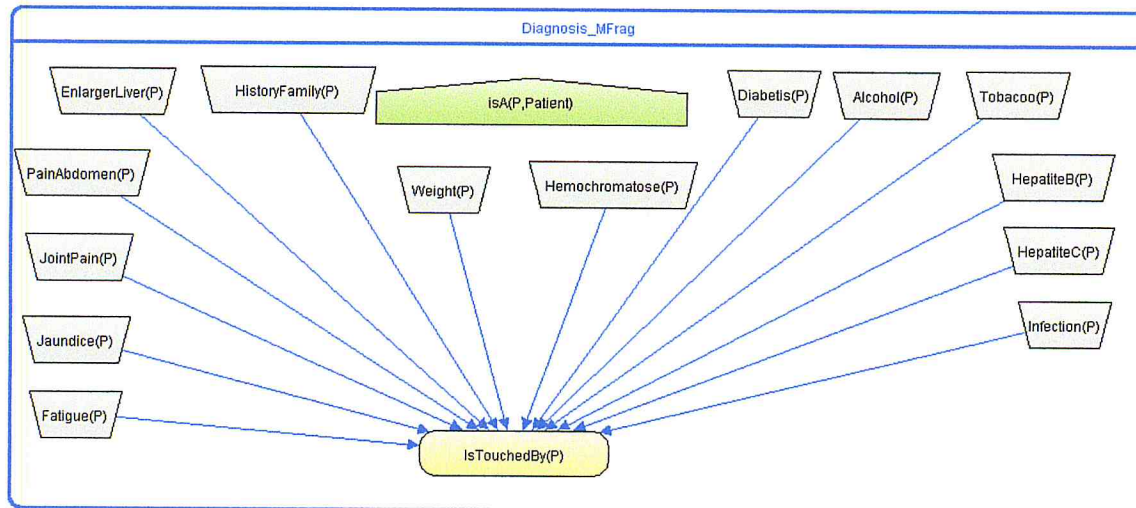


Figure 15 : Le MFragment de Diagnosis.

5.4.2. Probabilités

La dernière étape dans la construction d'une ontologie probabiliste dans UnBBayes est de définir les Distributions de Probabilité Locales (LPD) pour toutes les variables aléatoires (VR) résidents.

Chaque MFragment définit les LPD pour ses nœuds résidentiels, présentés sous forme d'ovales jaunes. Ces distributions sont conditionnées par la satisfaction des contraintes exprimées par les VR contextuels, présentés sous forme de pentagones verts. Les distributions locales peuvent dépendre des valeurs des VR récurrents, présentés sous forme de trapèzes gris, dont les distributions sont définies dans les MFragments dans lesquelles ils résident [46].

La figure -16- suivante montre le LPD pour le nœud résident IsTouchedBy, qui est la principale question à laquelle nous devons répondre afin d'atteindre le principal objectif de notre modèle qui est la prédiction. Cette distribution suit la grammaire UnBBayes-MEBN pour définir les LPD.

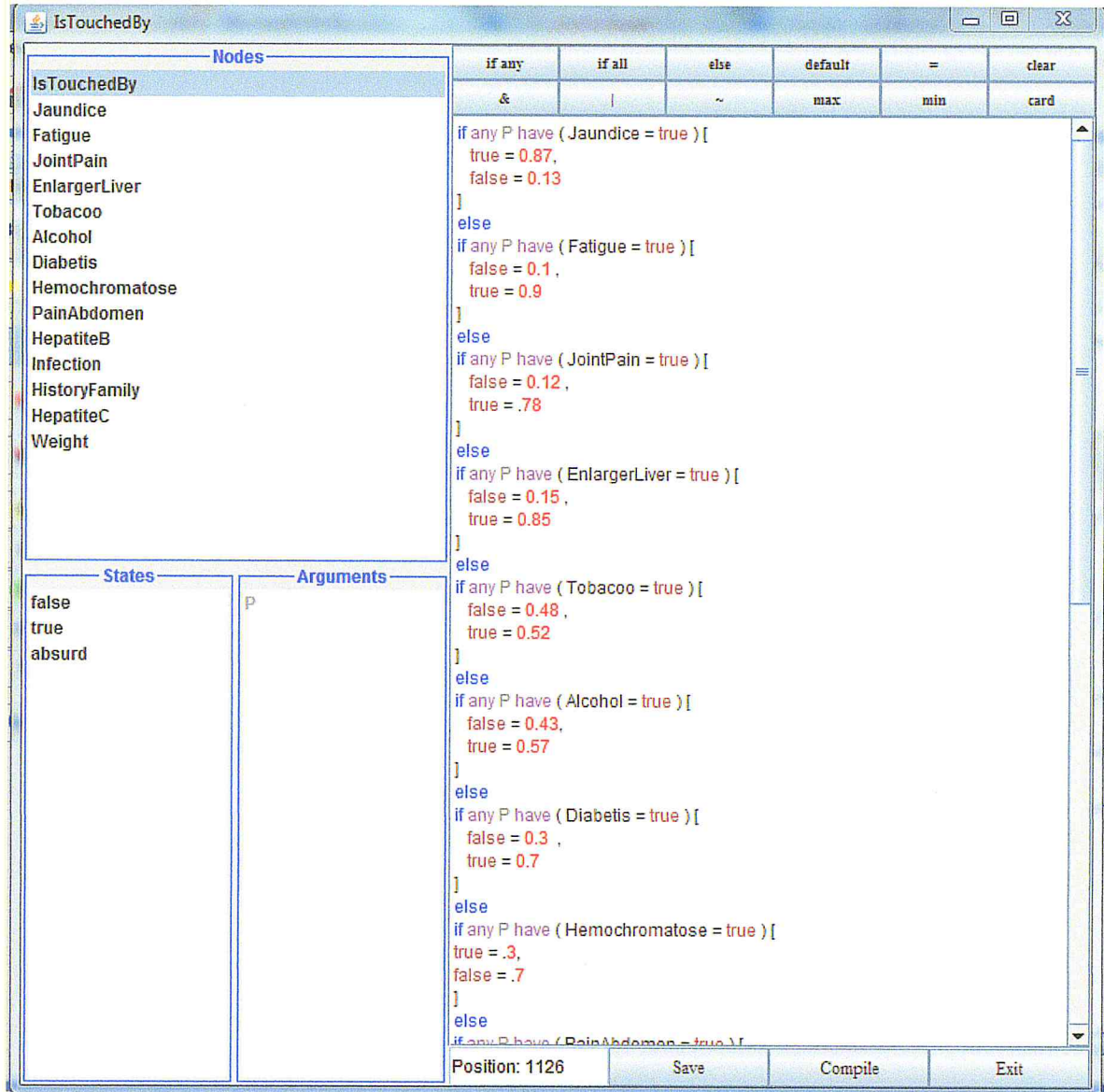


Figure 16 : LPD pour le nœud IsTouchedBy (patient).

Sur une utilisation typique d'une ontologie probabiliste PR-OWL, pendant le temps d'exécution (par exemple, en réponse à une requête), un processus de raisonnement logique créerait une instantiation des MFrag nécessaires pour répondre à la requête. Le résultat de ce processus est un réseau bayésien spécifique à la situation (SSBN), qui est un réseau bayésien minimal suffisant pour obtenir la distribution postérieure pour un ensemble d'instances de variables aléatoires cibles, compte tenu d'un ensemble d'occurrences variables aléatoires. Dans une ontologie probabiliste PR-OWL, les types d'entités correspondent aux classes OWL, les attributs correspondent aux propriétés OWL et les

relations correspondent aux relations OWL. Ainsi, PR-OWL permet au concepteur d'ontologie de spécifier des distributions de probabilité pour exprimer l'incertitude sur les propriétés et les relations dans une ontologie OWL [46].

La Figure-17- suivante présente les résultats d'une seule combinaison d'états pour les VR patient sans préciser ses évidences:

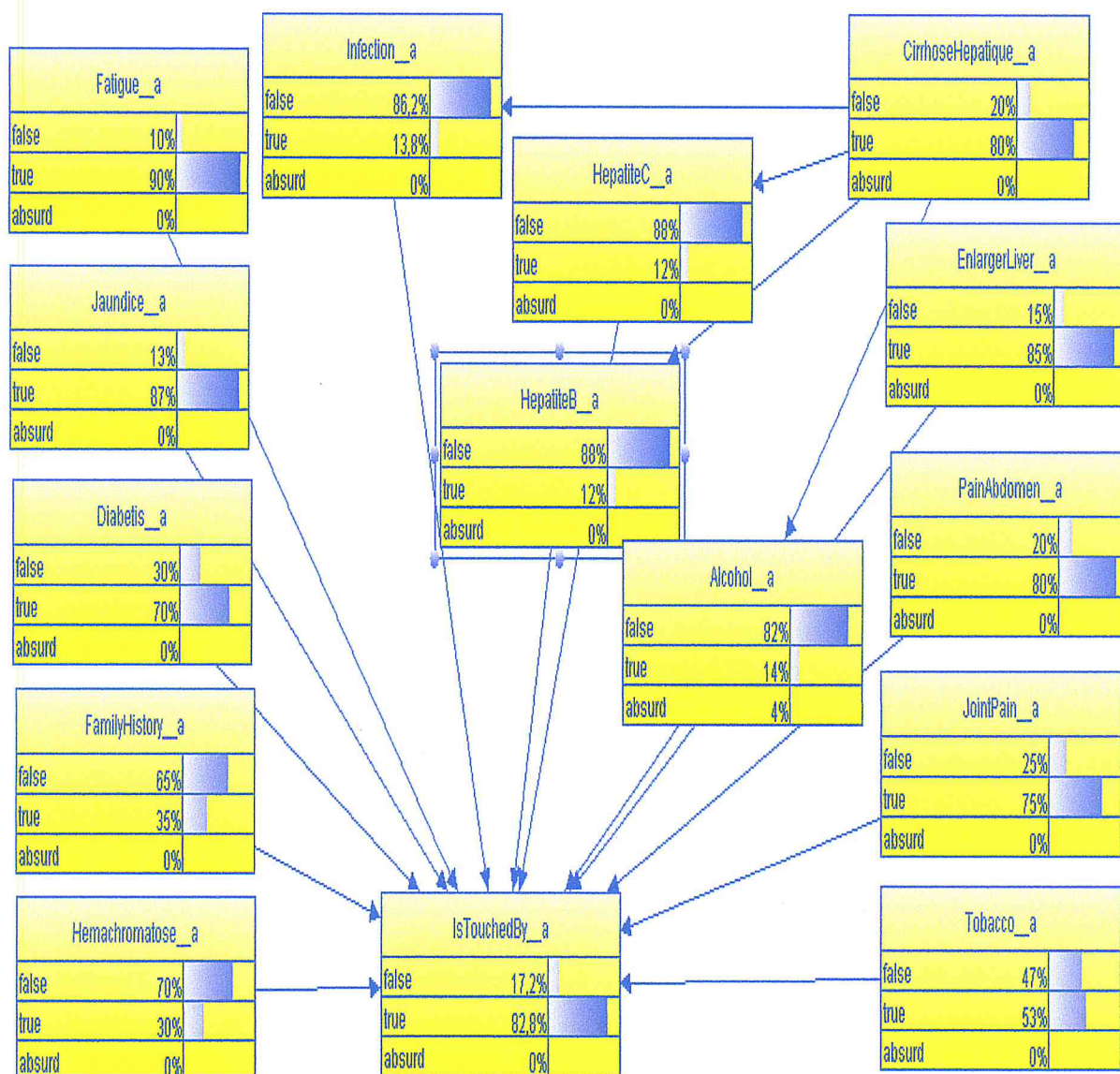


Figure 17 : Résultat SSBN pour le nœud IsTouchedBy (patient).

5.5. Test

Dans la plupart des méthodes de modélisation, le test joue un rôle essentiel. Comme le soulignent Laskey et Mahoney, Le test ne devrait pas être uniquement pour la vitrine et démontrer que le modèle fonctionne. L'objectif de la discipline de test est de trouver des défauts et des domaines d'amélioration dans le modèle.

Avant de commencer à décrire les activités dans la discipline Test, il est important de comprendre les différents types d'évaluation qui doivent être effectués. La littérature distingue deux types d'évaluation la vérification et la validation. D'une part, la vérification concerne la fourniture de toutes les fonctionnalités promises au client. Cela implique généralement l'examen des exigences, de la documentation, de la conception et du code. La vérification se fait souvent par des inspections et par des listes de contrôle. D'autre part, la validation concerne le bon comportement du système. La validation est le test réel du système et il est effectué après vérification [46].

5.5.1. Evaluation

Dans cette partie, nous nous concentrerons sur l'évaluation fondée sur des cas, qui définira des différents scénarios pour tester notre modèle. Nous voulons tester le comportement d'une partie du modèle, plus précisément, en vérifiant comment la variable centrée se comporte avec un ensemble de données distinct. Dans le cas de PR-OWL, nous pouvons analyser le comportement des variables aléatoires d'intérêt fournies par MFrag. Ce test MFrag est important pour saisir la cohérence locale du modèle [46].

a. Vérification

La vérification concerne l'établissement de «le système a été construit correctement», c'est-à-dire que les éléments du système sont conformes à leurs spécifications de performance définies [46].

Dans notre cas d'étude nous vérifierons essentiellement la façon dont la requête `IsTouchedBy (Patient)` se différencier entre les symptômes et les risque facteurs.

Nous définissons deux différents scénarios pour illustrer la mécanique de la définition et du test d'un scénario. Le premier est celui qui montre les évidences concernant les symptômes, Le second les risques facteurs.

b. Validation

La validation vise à établir que le «bon système a été construit», c'est-à-dire qu'il réalise son utilisation prévue dans son environnement opérationnel [46].

Lors de la définition d'un scénario, il est important de définir l'hypothèse a testée et quel est le résultat attendu, en plus de fournir les preuves qui seront utilisées.

Le premier scénario :

1. Hypothèse a testée

IstouchedBy(Patient).

2. Résultat attendu

Faible probabilité quand le patient n'a que les risques facteurs.

3. Evidence

- (a) FamilyHistory (Patient) = true.
- (b) Diabetis (Patient) = true.
- (c) Infection (Patient) = true.
- (d) HepatiteC (Patient) = true.
- (e) HepatiteB (Patient) = true.
- (f) Tobacoo (Patient) = true.
- (g) Alcohol (Patient) = true.
- (h) Hemochromatose (Patient) = true.

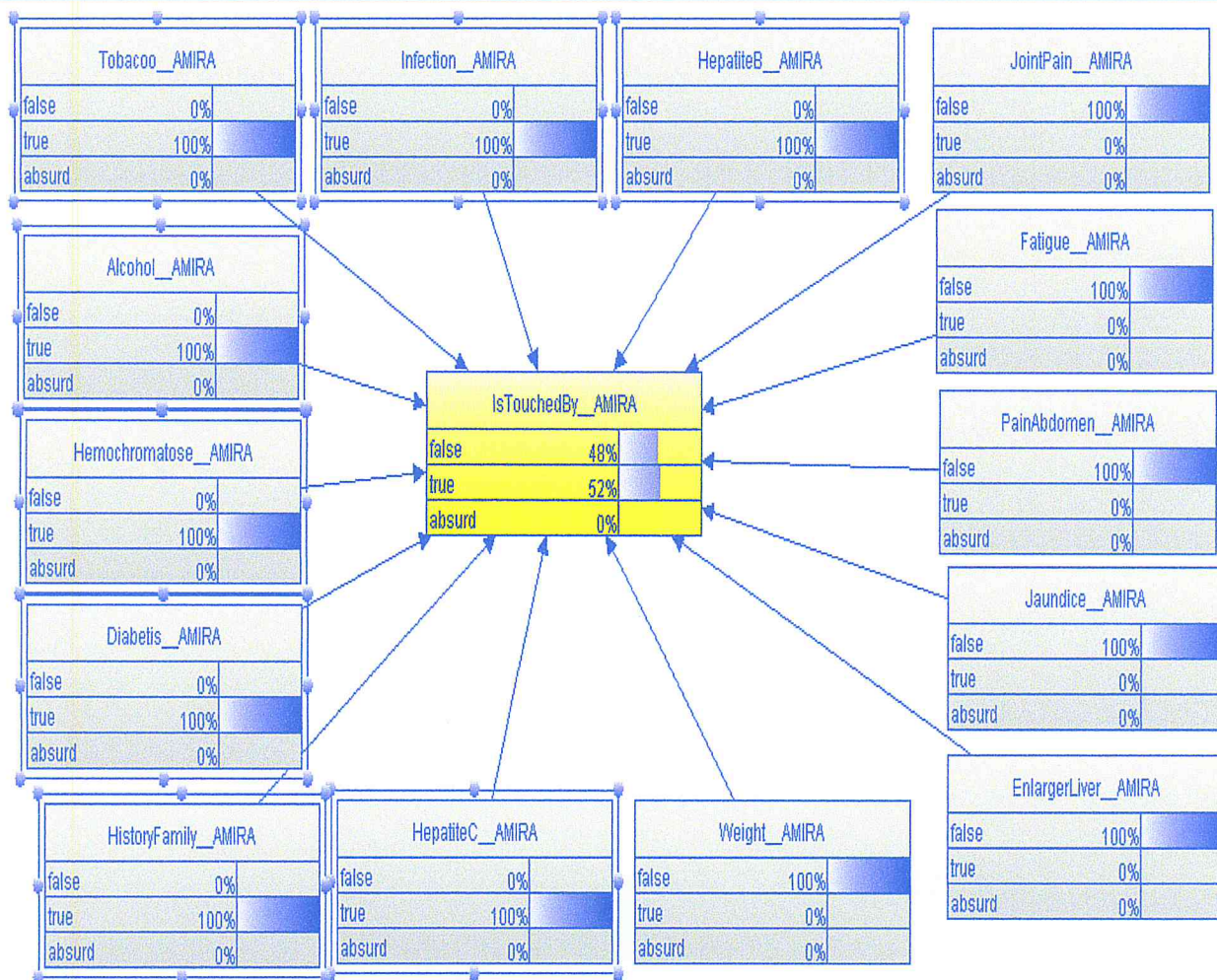


Figure 18 : SSBN généré pour le premier scénario.

Le second scénario :

4. Hypothèse a testée

IstouchedBy(Patient).

5. Résultat attendu

Haute probabilité quand le patient n'a que les symptômes.

6. Evidence

- (i) Jaundice (Patient) = true.
- (j) Fatigue (Patient) = true.
- (k) EnlargeLiver (Patient) = true.
- (l) PainAbdomen(Patient) = true.

(m) JointPain (Patient) = true.

(n) Wieght (Patient) = true.

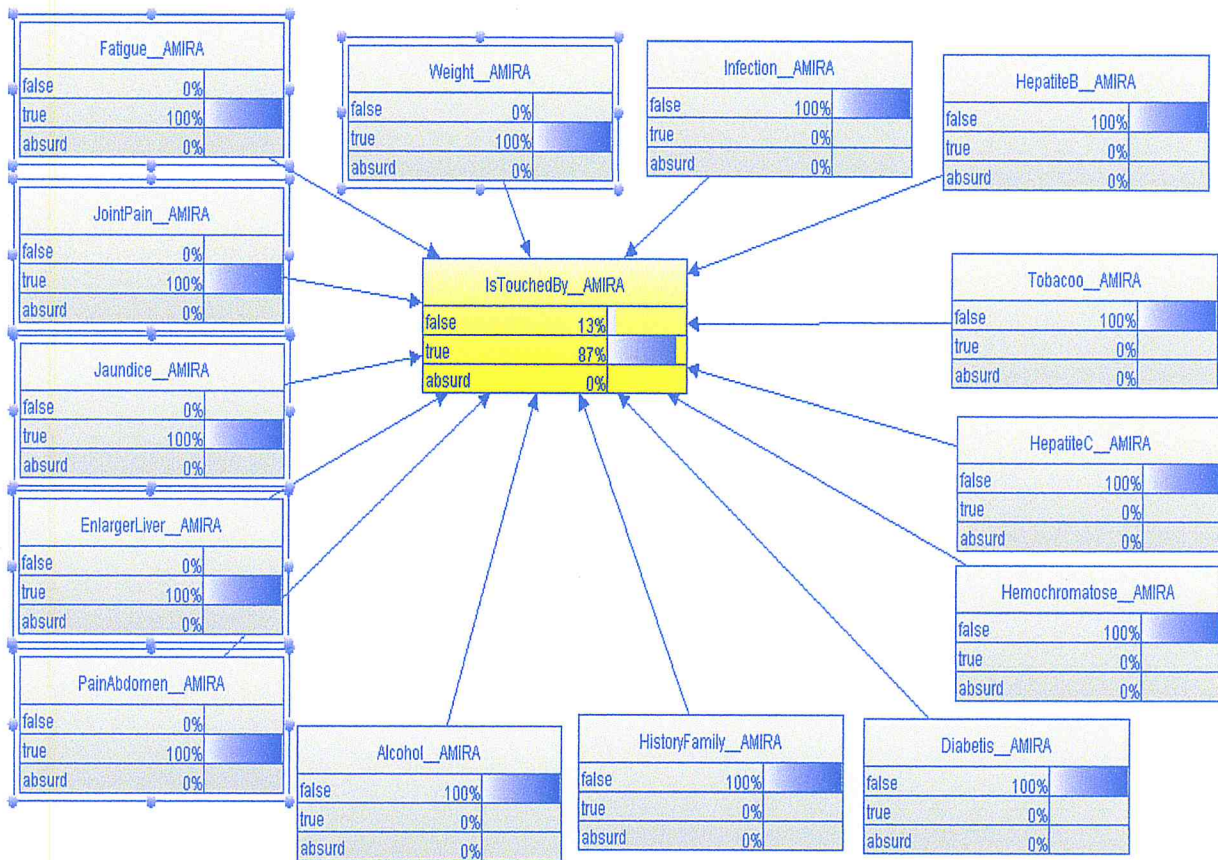


Figure 19 : SSBN généré pour le second scénario.

6. Résultat final de l'ontologie probabiliste

Après avoir suivie les étapes de la méthodologie (POMC), nous avons obtenu une ontologie probabiliste qui porte le Mtheory avec ses différents MFrag.

6.1. La différence entre l'ontologie classique et l'ontologie probabiliste obtenue

La hiérarchie des classes de l'ontologie classique (OWL) et de l'ontologie probabiliste (PR-OWL)

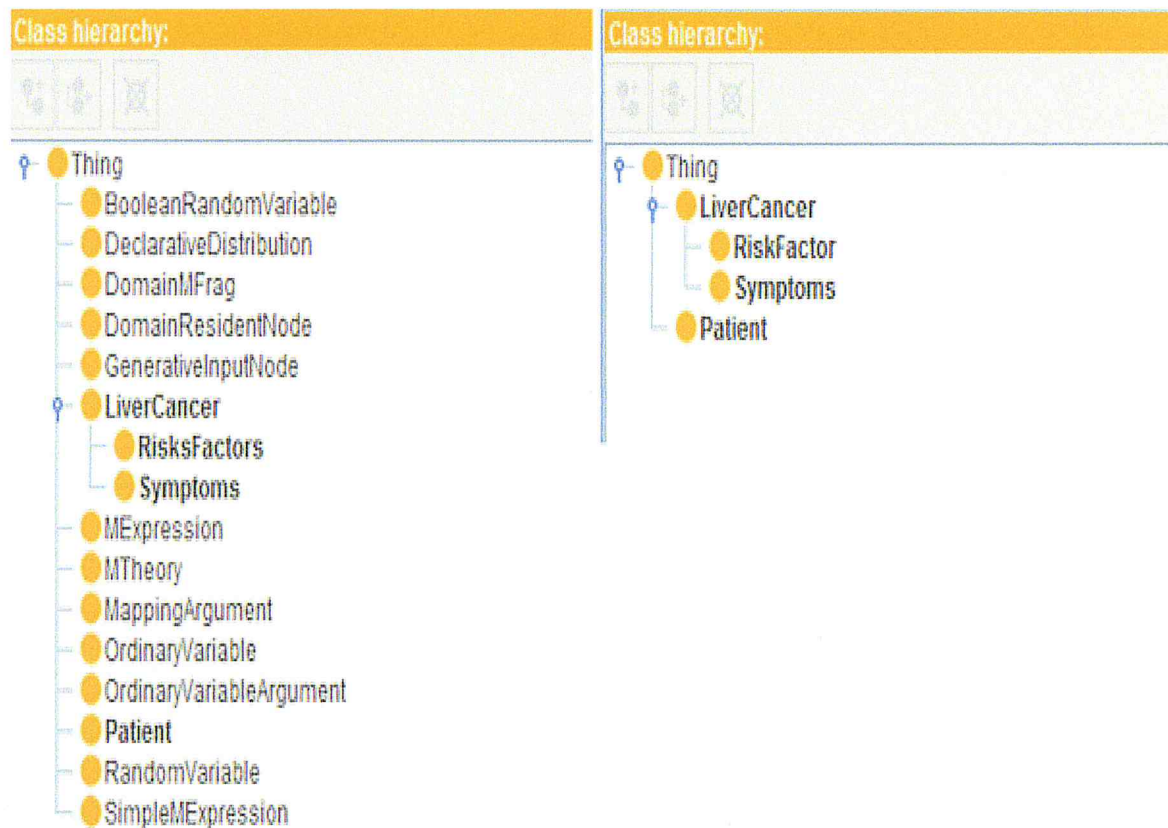


Figure 20 : La hiérarchie des classes de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).

- La hiérarchie des propriétés d'objet.

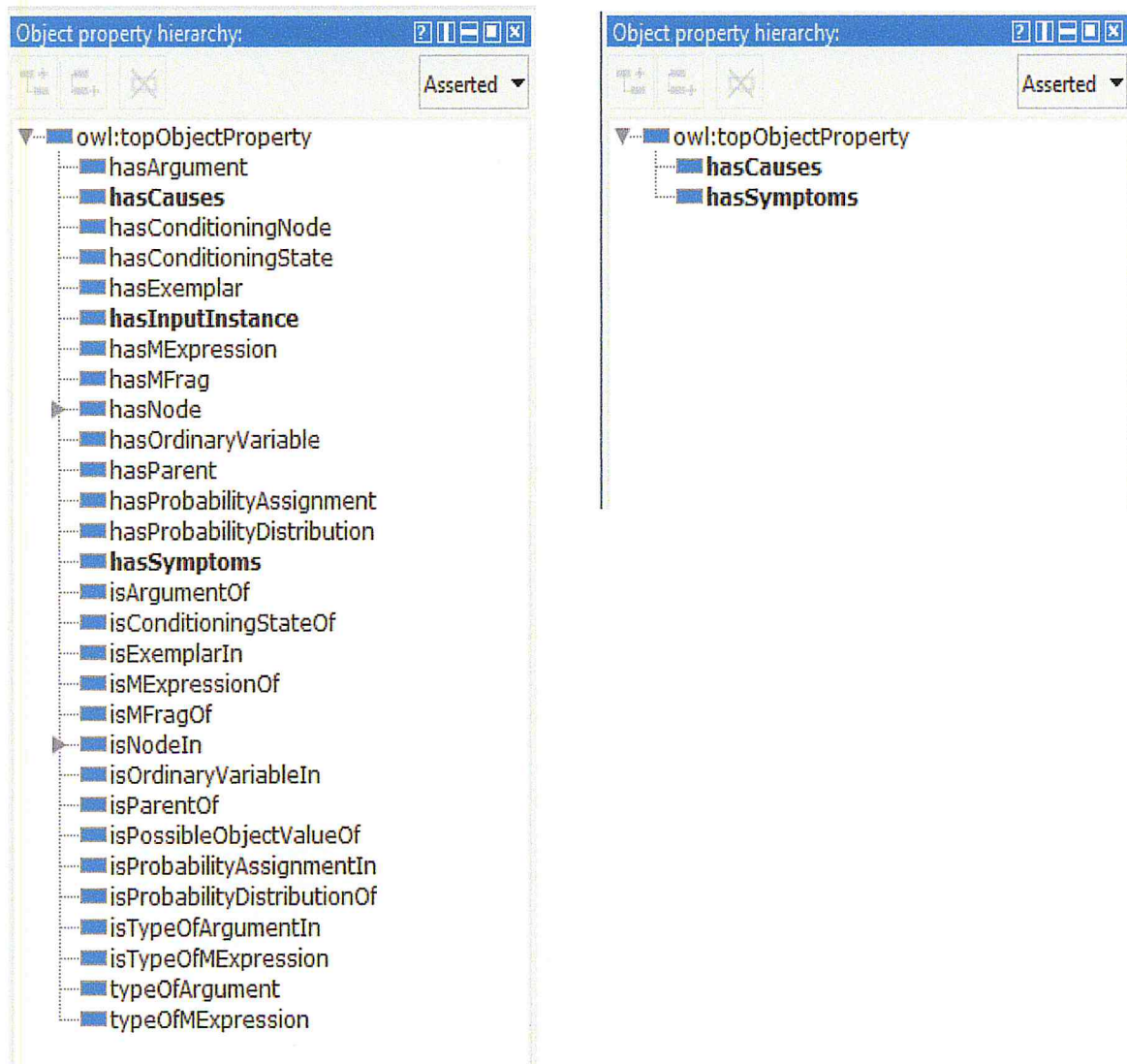
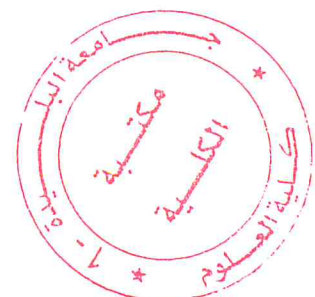


Figure 21 : La hiérarchie des propriétés d'objet de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).



- La hiérarchie des propriétés des données.

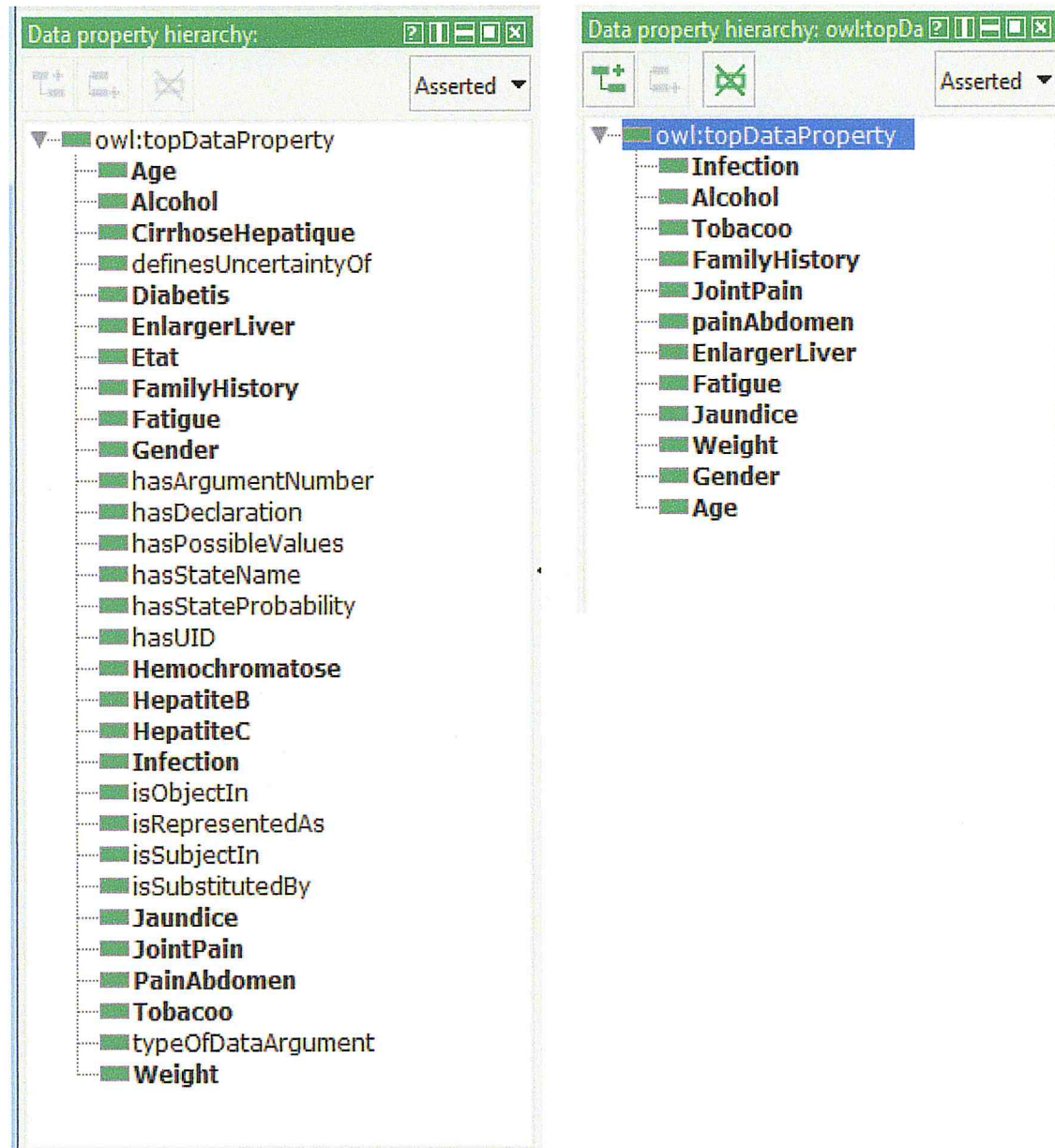


Figure 22 : La hiérarchie des propriétés des données de l'ontologie classique (à droite) et de l'ontologie probabiliste (à gauche).

7. Conclusion

Dans ce chapitre nous avons présenté notre architecture du système ainsi que toutes les étapes de la construction de notre ontologie probabiliste « Liver_Cancer », en suivant le processus de la méthodologie UMP-ST pour arriver finalement à faire les prédictions de la maladie, après avoir rempli les tables de distribution de chaque nœud appartenant à l'ontologie.

Après avoir fait les probabilités sous l'UnBBayes, nous passerons à la réalisation de notre système de prédiction qui doit aider les médecins à diagnostiquer le cancer du foie que nous allons par la suite l'implémenter dans le dernier chapitre.

Chapitre 4 :

La mise en œuvre du système

1. Introduction

Après avoir terminé toutes les étapes de la méthodologie nous commençons à implémenter notre ontologie probabiliste et notre système qui permet de faire l'inférence de notre cas d'étude dans un langage spécifique dans PR-OWL en utilisant UnBBayes.

Dans ce chapitre nous allons décrire de façon visuelle l'implémentation de notre système, en effectuant des captures d'écran des différentes interfaces du système.

2. Outils de développement de notre système

- **Eclipse et JDK**

Lors d'un développement d'application ou de la création d'un site web, il est plus qu'indispensable d'avoir un outil comme Eclipse IDE pour simplifier la tâche. Eclipse IDE « intégré » (en anglais, pour Integrated Development Environment, le produit d'un outil peut servir de matière première pour un autre outil) n'est autre qu'un outil de développement professionnel. Il offre de nombreuses fonctionnalités pratiques. Ainsi qu'un environnement «Java Development Kit » JDK est requis pour les développements en Java de notre système.

- **Unbbayes API**

UnBBayes est un Framework de réseau probabiliste écrit en Java. Il a à la fois une interface graphique et une API avec inférence, échantillonnage, apprentissage et évaluation. Il supporte BN, ID, MSBN, OOBN, HBN, MEBN / PR-OWL, structure, paramètre et apprentissage progressif [47].

- **Protégé**

Protégé est un éditeur d'ontologies distribué en open source par l'université en informatique médicale de Stanford. Protégé n'est un outil spécialement dédié à OWL, mais un éditeur hautement extensible, capable de manipuler des formats très divers. Le support d'OWL, comme de nombreux autres formats, est possible dans protégé grâce à un plugin dédié. Protégé est un outil employé par les développeurs et des experts de domaine pour développer des systèmes basés sur les connaissances (Ontologies). Des applications développées avec Protégé sont employées dans la résolution des problèmes et la prise de décision dans un domaine particulier. Protégé est aussi une plate-forme extensible, grâce au

système de plug-ins, qui permet de gérer des contenus multimédias, interroger, évaluer et fusionner des ontologies, etc. L'outil Protégé possède une interface utilisateur graphique (GUI) lui permettant de manipuler aisément tous les éléments d'une ontologie : classe, méta-classe, propriété, instance,...etc. Protégé peut être utilisé dans n'importe quel domaine où les concepts peuvent être modélisés en une hiérarchie des classes [48].

✓ Protège-OWL API

Est une bibliothèque Java open source pour la Web Ontology Language (OWL) et RDF (S). L'API fournit des classes et des méthodes pour charger et enregistrer des fichiers OWL, pour interroger et manipuler des modèles de données OWL, et pour effectuer un raisonnement basé sur les moteurs Description Logic. En outre, l'API est optimisée pour la mise en œuvre d'interfaces utilisateur graphiques [49].

✓ OWL API

L'API OWL est une interface et une implémentation Java pour le W3C Web Ontology Language (OWL), utilisé pour représenter les ontologies du Web sémantique. L'API est axée sur OWL DL et l'OWL à venir 2 [50].

OWL-API s'agit d'un ensemble d'interfaces riches en fonctionnalité, permettant une manipulation flexible d'ontologies. De nombreux projets utilisent OWL-API comme outil de développement [51].

3. La mise en œuvre du système

Après l'effectuation des étapes de la création de l'ontologie probabiliste nous avons eu un fichier Avec une extension UBF qui porte la MTheory avec ses différents MFrag du domaine ainsi que les différents Nœuds résident (Risque facteurs, Symptômes) de chaque MFrag. Et un fichier PLM (Base de Connaissances) qui comportent un exemple d'un patient probablement touché par le cancer du foie par remplir toutes Les évidences dans le fichier UBF sauf le nœud résident « IsTouchedBy » qui est un nœud qui sert à faire la décision.

4. La base de connaissance

La figure -23- ci-dessus montre un modèle d'une Base de connaissances (Fichier PLM) :

```
(DEFMODULE "/PL-KERNEL-KB/PL-USER/GENERATIVE_MODULE_1/FINDINGS_MODULE_1"
:CASE-SENSITIVE? FALSE)

(IN-MODULE "/PL-KERNEL-KB/PL-USER/GENERATIVE_MODULE_1/FINDINGS_MODULE_1")

(IN-DIALECT :KIF)

(ASSERT (PATIENT_LABEL SAMIRA_1))
(ASSERT (PATIENT_LABEL MOUNIR_1))
(ASSERT (PATIENT_LABEL AHMED_1))
(ASSERT (PATIENT_LABEL AHMED_2))
(ASSERT (THING_LABEL BELOW30))
(ASSERT (THING_LABEL NO))
(ASSERT (THING_LABEL OVER))
(ASSERT (THING_LABEL WOMEN))
(ASSERT (THING_LABEL ABOVE30))
(ASSERT (THING_LABEL LESS))
(ASSERT (THING_LABEL MAL))
(ASSERT (THING_LABEL YES))
(ASSERT (THING_LABEL MEN))
(ASSERT (not (CIRRHOSEHEPATIQUE SAMIRA_1)))
(ASSERT (CIRRHOSEHEPATIQUE MOUNIR_1))
(ASSERT (not (CIRRHOSEHEPATIQUE AHMED_1)))
(ASSERT (not (DIABETIS SAMIRA_1)))
(ASSERT (DIABETIS MOUNIR_1))
(ASSERT (not (DIABETIS AHMED_1)))
(ASSERT (not (HEMOCHROMATOSE SAMIRA_1)))
(ASSERT (not (HEMOCHROMATOSE MOUNIR_1)))
(ASSERT (not (HEMOCHROMATOSE AHMED_1)))
(ASSERT (not (ALCOHOL SAMIRA_1)))
(ASSERT (ALCOHOL MOUNIR_1))
(ASSERT (not (ALCOHOL AHMED_1)))
(ASSERT (not (TOBACOO SAMIRA_1)))
(ASSERT (TOBACOO MOUNIR_1))
(ASSERT (not (TOBACOO AHMED_1)))
```

Figure 23 : Base de connaissances.

5. Les interfaces

5.1. L'interface d'accueil

La première interface à afficher est la suivante :

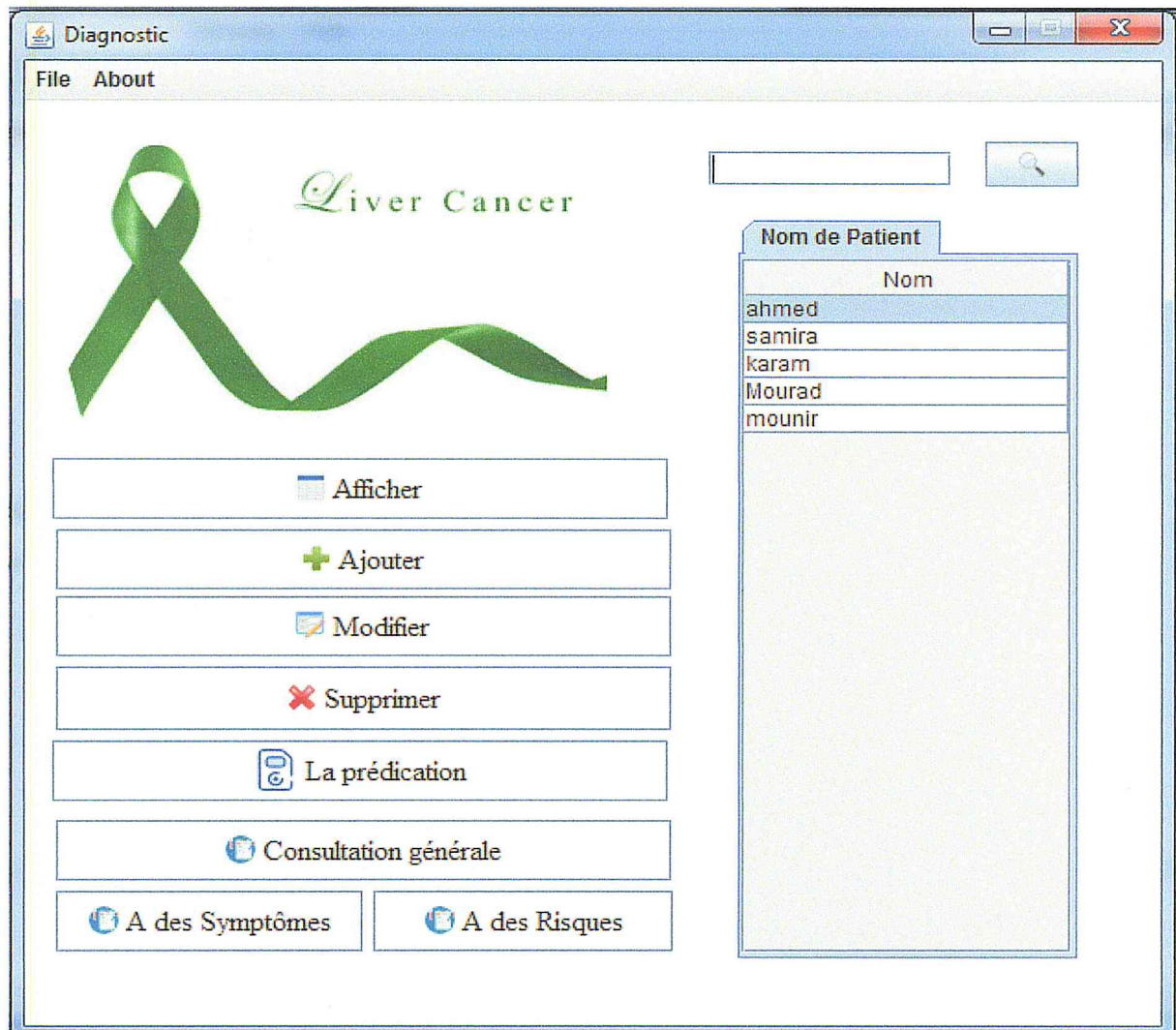


Figure 24 : Interface d'accueil

Cette interface permet l'accès aux différentes fonctionnalités qu'on peut faire sur la base de connaissance après l'importation des deux fichiers (UBF, PLM).

- **Afficher**

Cette même interface permet d'afficher tous les noms des patients existant dans notre base de connaissances.

- **Supprimer**

Cette même interface permet de supprimer un patient existant dans la base de connaissances.

- **Modifier**

Cette interface permet de modifier les informations d'un patient existant dans la base de connaissances.

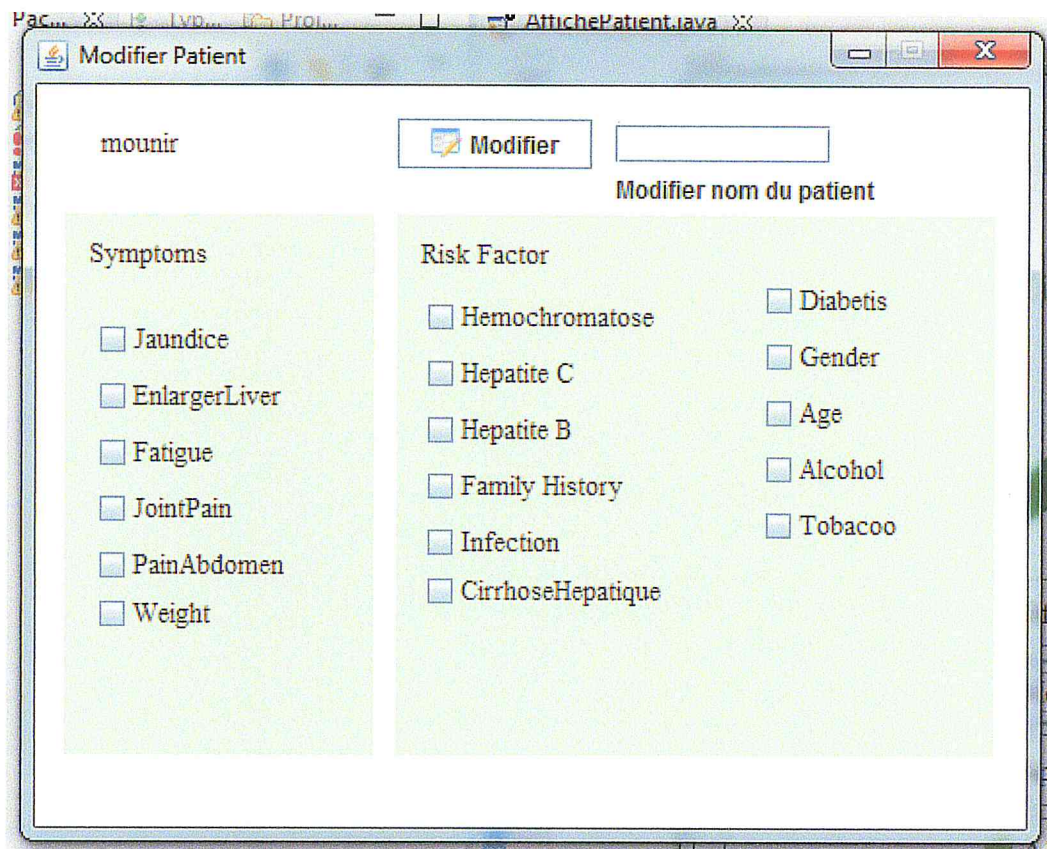


Figure 25 : L'interface « Modifier patient ».

- **Ajouter**

Cette interface permet d'ajouter un patient qui n'existe pas déjà dans la base de connaissances, ainsi elle permet de spécifier ses facteurs de risque et ses symptômes.

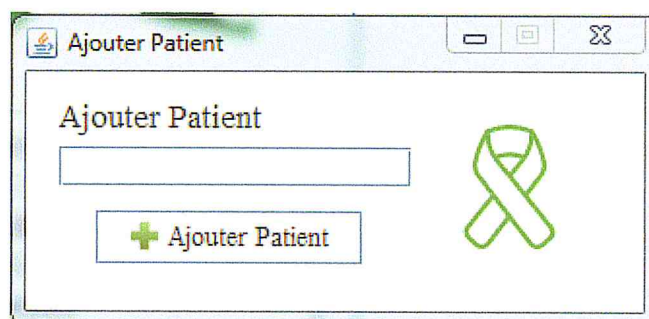


Figure 26 : L'interface « Ajouter patient ».

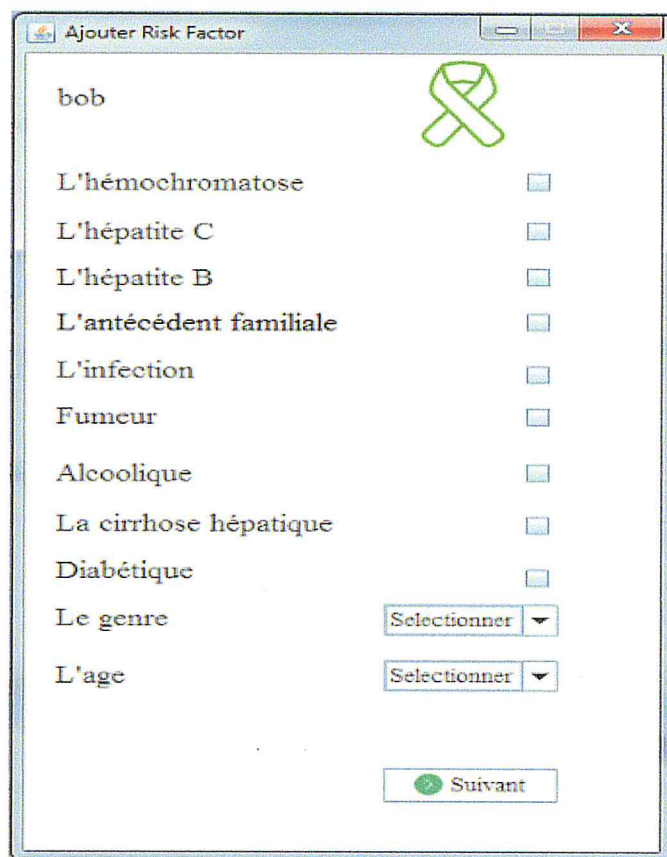


Figure 27 : L'interface « Ajouter Facteurs de risque ».

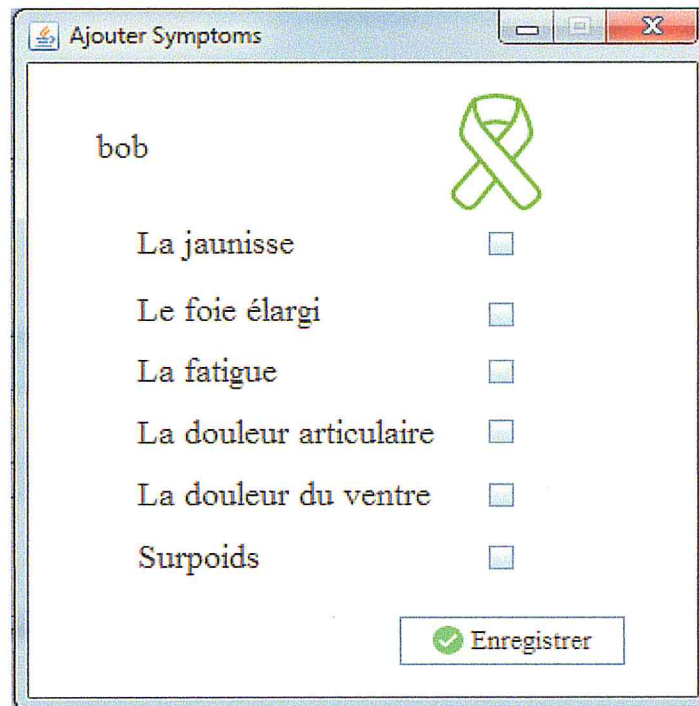


Figure 28 : L'interface « Ajouter Symptômes ».

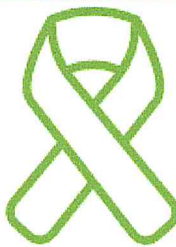
5.2. L'interface de consultation

- Consultation générale

Cette interface permet de consulter tous les facteurs de risque et symptômes d'un patient sélectionné

Consultation Générale

samira



Symptoms

- ☐ Jaundice
- ☒ EnlargerLiver
- ☒ Fatigue
- ☒ JointPain
- ☐ PainAbdomen
- ☒ Weight

Risk Factor

☐ Hemochromatose	☐ Diabetis
☒ Hepatite C	☐ Gender
☐ Hepatite B	☐ Age
☐ Family History	☐ Alcohol
☐ Infection	☒ Tobacoo
☐ CirrhoseHepatique	

Figure 29 : L'interface « Consultation Générale ».

- Consulter facteurs de risque

Cette interface permet de consulter les facteurs de risque.

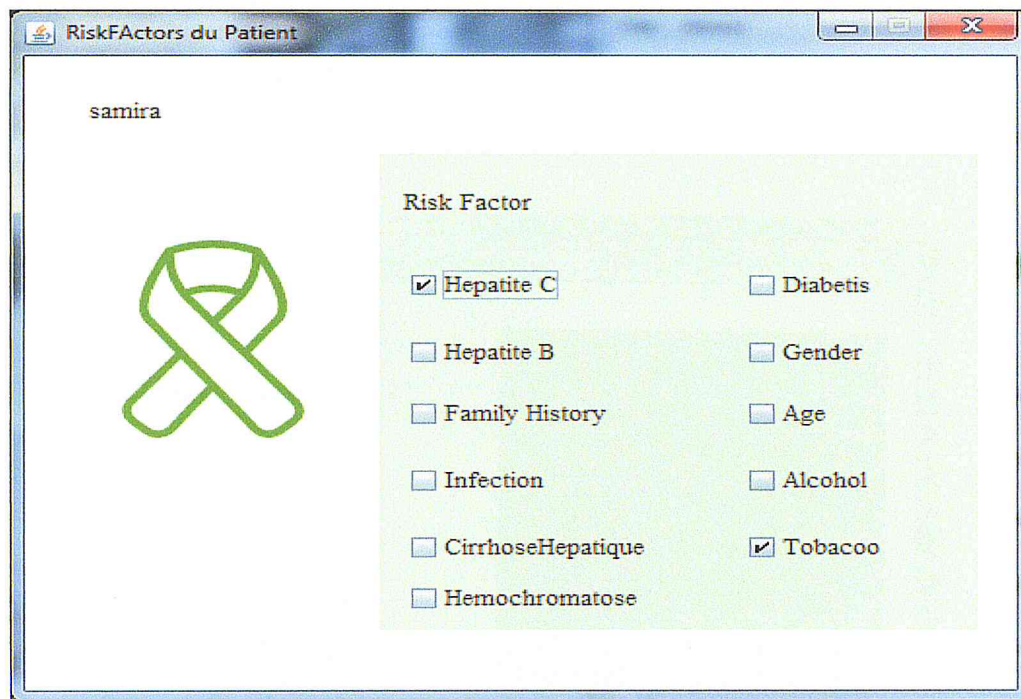


Figure 30 : L'interface «RiskFactors du patient».

- Consulter symptômes

Cette interface permet de consulter les symptômes d'un patient sélectionné.

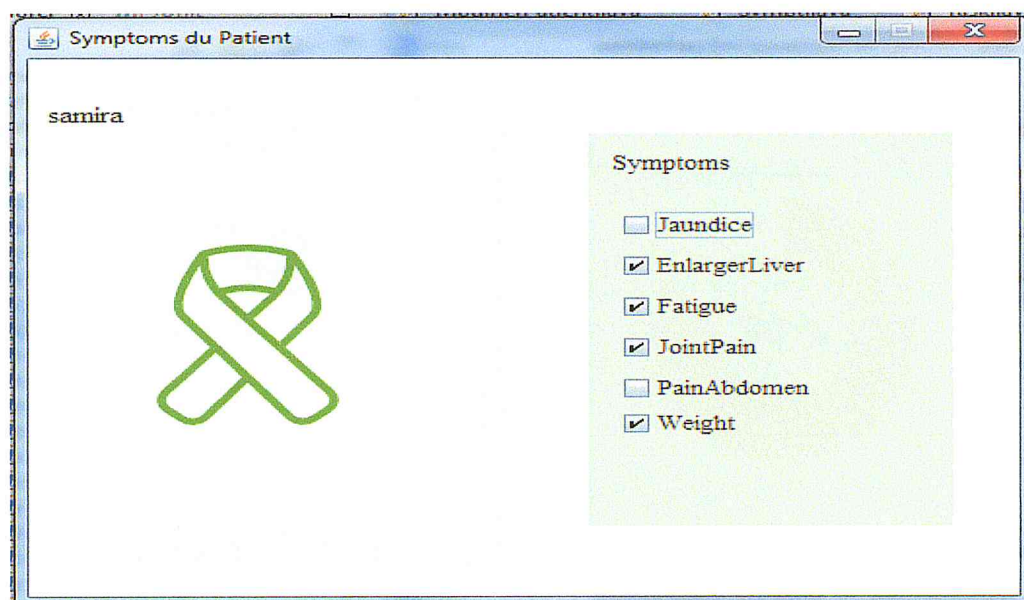


Figure 31 : L'interface «Symptoms du patient».

6. Conclusion

Dans ce chapitre nous avons implémenté notre système de prédiction après avoir importé l'ontologie probabiliste et la base de connaissance, avec des captures de notre application.

Conclusion générale

Conclusion générale

Les ontologies sont une approche puissante qui permettent le partage des connaissances dans un domaine quelconque et de soutenir l'interopérabilité entre les personnes et les systèmes informatiques.

Le domaine médical est basé sur les diagnostics. Mais les ontologies classiques (traditionnelles) n'intègre pas de connaissances incertaines, alors que l'ontologie probabiliste joue un rôle important dans la représentation de l'incertitude.

Ce mémoire présente l'inférence sur une ontologie probabiliste qui est faite à l'aide de la méthodologie UMP-ST, qui présente notre cas d'étude en une ontologie probabiliste en suivant toutes ses étapes de cette dernière qui sont :

- **Les exigences (Requirements) :** cette discipline définit les objectifs qui doivent être atteints en raisonnant avec la sémantique fournie par notre modèle. Habituellement, lors de la conception d'un OP, le but est de pouvoir automatiser un processus de raisonnement impliquant une incertitude. Par objectifs, c'est en prenant en considération les différentes questions posé par l'utilisateur, pour que le système puisse y répondre via le raisonnement de l'OP.

Le principal objectif du domaine médical et spécialement la maladie du cancer du foie est de savoir si le patient est touché par cette maladie tout en posant la question, est ce que le patient a les symptômes et les facteurs de risques de cette maladie et tout en donnant les preuves (évidence) qui concernent ces questions.

- **Analyse & Conception :** cette discipline décrit les classes d'entités, leurs attributs et les relations entre eux qui présentent le modèle notre cas d'étude en utilisant le langage de modélisation unifiée (UML) pour faciliter la communication avec l'expert de notre domaine toute en enregistrant séparément les règles déterministes dans l'ontologie probabiliste PR-OWL.
- **Implémentation :** Une fois l'étape Analyse et conception terminée, l'étape suivante consiste à implémenter le modèle dans un langage spécifique. Notre ontologie probabiliste « Liver-Cancer » a été construite en utilisant le langage de l'ontologie probabiliste PR-OWL. Ce dernier utilise MEBN pour exprimer l'incertitude sur les propriétés ou les relations de notre étude de cas qui sont définies dans les classes OWL.

- **Test** : il est responsable d'évaluer le comportement prévu du modèle développé au cours de la discipline de mise en œuvre à partir des règles définies lors de l'analyse et de la conception et aussi l'aboutissement des objectifs suscités pendant les exigences. Comme indiqué précédemment, il est judicieux de tester certaines des règles et hypothèses avant même la mise en œuvre. C'est une étape cruciale pour atténuer les risques.

Dans notre cas nous souhaitons vérifier si le modèle répond aux objectifs que nous avons développées au cours de l'étape des exigences, par tester l'hypothèse d'expert de domaine en vérifiant et validant qu'un même patient à une probabilité élevé lorsqu'il a seulement les symptômes et une faible probabilité lorsqu'il a que les facteurs de risque.

L'ontologie probabiliste PR-OWL est compatible avec OWL, cette ontologie basée sur les réseaux bayésiens avec entités multiples est capable de prédire tout en tenant compte des incertitudes autour des données.

Notre cas d'étude sert à faire la prédication qu'un patient atteint d'un cancer du foie tout en sachant ses facteurs de risque et ses symptômes qui sont représentés sous la forme de nœuds résidents, qui sont définis par distributions de probabilité locales (LPD).

Afin de réaliser notre conception nous avons implémenté notre système en important notre nouvelle ontologie probabiliste, ainsi nous avons montré les différentes interfaces de notre système.

Cependant nos perspectives c'est de rendre le système capable à suivre les nouveautés de la médecine, par exemple ajouter des nouveaux symptômes ou facteurs de risque.

- [25] Pham, T. A. L. (2008). *De l'optimisation à la décomposition de l'ontologie dans la logique de description* (Doctoral dissertation, Nice).
- [26] P. Naïm, P-H. Willemin, P.Leray, O. Pourret, and A. Becker. Réseaux bayésiens. 423 pages, Eyrolles, Paris, 2007.
- [27] Patrick Naïm, Pierre-Henri Willemin, Philippe Leray, Olivier Pourret, Anna Becker. Réseaux Bayésiens, Eyrolles 2008.
- [28] Pearl, J. (1988). Morgan Kaufmann series in representation and reasoning. Probabilistic reasoning in intelligent systems: Networks of plausible inference.
- [29] Lauritzen, S. L., & Spiegelhalter, D. J. (1990). Local computations with probabilities on graphical structures and their application to expert systems. *Readings in uncertain reasoning*, 415-448.
- [30] Santos, F. (2015). L'algorithme EM: une courte présentation.
- [31] Shachter, R. D., & Peot, M. A. (2013). Simulation approaches to general probabilistic inference on belief networks. *arXiv preprint arXiv:1304.1526*.
- [32] Fung, R., & Del Favero, B. (1994, July). Backward simulation in Bayesian networks. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence* (pp. 227-234). Morgan Kaufmann Publishers Inc.
- [33] Intelligence, A., & Approach, A. M. (2010). Stuart Russell and Peter Norvig.
- [34] Laskey, K. B., Costa, P. C., & Janssen, T. (2010). Probabilistic ontologies for Multi-INT fusion. *Ontologies and Semantic Technologies for Intelligence*, 213, 147.
- [35] Ravi, K., & Singh, S. (2013). Bayesian network for uncertainty representation in semantic web: a survey. *International Journal of Computer Applications Technology and Research*, 2(5), 530-538.
- [36] Jovic, A., Prcela, M., & Gamberger, D. (2007, June). Ontologies in medical knowledge representation. In *Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on* (pp. 535-540).
- [37] Bertaud-Gounot, V., Duvauferrier, R., & Burgun, A. (2012). Ontology and medical diagnosis. *Informatics for Health and Social Care*, 37(2), 51-61.
- [38] Smail, L. H., Supriyanto, E., Satria, H., & Priya, R. S. (2012). Building and reusing medical ontology for tropical diseases management.
- [39] <https://fr.slideshare.net/salahabdessemed1/terminologie-mdical>

Reference

- [1] Neches, R., Fikes, R. E., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling technology for knowledge sharing. *AI magazine*, 12(3), 36.
- [2] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- [3] Staab, S., & Maedche, A. (2000, August). Ontology engineering beyond the modeling of concepts and relations. In *Proceedings of the ECAI-2000 Workshop on Ontologies and Problem-Solving Methods*. Berlin.
- [4] Psyché, V. (2007). *Rôle des ontologies en ingénierie des EIAH: cas d'un système d'assistance au design pédagogique* (Doctoral dissertation, Université du Québec à Montréal).
- [5] Mizoguchi, R., & Ikeda, M. (1998). Towards ontology engineering. *Journal-Japanese Society for Artificial Intelligence*, 13, 9-10.
- [6] Bachimont B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In *Ingénierie des connaissances. Évolution Récentes et nouveaux défis*. Paris: Eyrolles.
- [7] Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, 11(2).
- [8] Paulo C. G. Costa & Kathryn B. Laskey, PR-OWL: A Bayesian extension to the OWL Ontology Language, 01/04/2016. <http://www.pr-owl.org/mebn/index.php>.
- [9] Farquhar, A., Fikes, R., & Rice, J. (1997). The ontolingua server: A tool for collaborative ontology construction. *International journal of human-computer studies*, 46(6), 707-727.
- [10] Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive science*, 9(2), 171-216.
- [11] Brickley, D., & Guha, R. V. (2004). RDF vocabulary description language 1.0: RDF schema.
- [12] D. Fensel, S. Decker, M. Erdmann, and R. Studer. Ontobroker: The very high idea. In *Proceedings of the 11th International Flairs Conference (FLAIRS-98)*, Sanibal Island, Florida, May 1998.

- [40] Temal, L., Lando, P., Dojat, M., Fürst, F., Gibaud, B., Kassel, G., & Lapujade, A. (2007). OntoNeuroLOG: une ontologie modulaire et multi-niveaux pour gérer l'hétérogénéité sémantique des métadonnées. *Actes de la journée thématique «Ontologies et Gestion de l'hétérogénéité sémantique» du GDR I*, 3.
- [41] Hardiker, N. R., Hoy, D., & Casey, A. (2000). Standards for nursing terminology. *Journal of the American Medical Informatics Association*, 7(6), 523-528.
- [42] Guefack, D. (2013). *Modélisation des signes dans les ontologies biomédicales pour l'aide au diagnostic* (Doctoral dissertation, Rennes 1).
- [43] Haug, P. J., Ferraro, J. P., Holmen, J., Wu, X., Mynam, K., Ebert, M., ... & Jones, J. (2013). An ontology-driven, diagnostic modeling system. *Journal of the American Medical Informatics Association*, 20(e1), e102-e110.
- [44] <http://www.chu-rouen.fr/page/tumeurs-du-foie>.
- [45] Pakhale, H., & Xaxa, D. K. (2016). A Survey on Diagnosis of Liver Disease Classification. *International Journal of Engineering and Techniques*, 2(3), 2395-1303.
- [46] Carvalho, R. N., Laskey, K. B., & Da Costa, P. C. (2016). Uncertainty modeling process for semantic technology. *PeerJ Computer Science*, 2, e77.
- [47] <https://sourceforge.net/projects/unbbayes>.
- [48] <http://dspace.univ-tlemcen.dz/bitstream/112/1062/9/Annexe-Protege>
- [49] https://protegewiki.stanford.edu/wiki/ProtegeOWL_API_Programmers_Guide
- [50] http://semanticweb.org/wiki/OWL_API.html
- [51] <http://www.pressinnov.com/2016/04/owl-api-une-api-java-pour-manipuler-des-ontologies>.

