

**UNIVERSITÉ SAAD DAHLEB DE BLIDA1**

**Faculté des sciences**

Département d'informatique



**MEMOIRE DE MASTER**

**En Informatique**

Option : Ingénierie Des Logiciels

**THÈME :**

**Raisonnement probabiliste des  
connaissances ontologiques  
incertaines.**

Réalisé par

Belaroui Yasmine

Deriouch Youssra Meriem

Encadré par

Dr. M. FAREH

Dr. I. RIALI

Mme.LAHIANI. Présidente

Mme.CHIKHI. Examinatrice

21 Juin 2023

# Remerciements

Nous remercions tout d'abord ALLAH le tout puissant qui, par sa grâce nous a permis d'arriver au bout de nos efforts en nous donnant la santé, la force, la volonté et en nous faisant entourer des merveilleuses personnes dont nous tenons à remercier.

Nous tenons à remercier chaleureusement **Mme. FAREH** pour son soutien, sa confiance, son aide inestimable tout au long de notre parcours. Nous remercions également notre co-promoteur **Mr. RIALI** pour son aide, sa disponibilité et ses orientations qui ont tellement contribué à la réalisation de notre projet. Nous sommes très heureux de travailler avec votre équipe et d'apprécier vos qualités, vos valeurs, votre sérieux et votre compétence.

Nous souhaitons adresser tous nos remerciements aux personnes qui ont participé de près ou de loin à la réalisation de ce mémoire, plus particulièrement à : **Dr. Moualfi Abdelghani** médecin spécialiste en cardiologie.

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

# Dédicaces

Je dédie ce travail à plusieurs personnes qui ont joué un rôle essentiel dans ma vie :

Tout d'abord, à ma chère mère, mon premier professeur, ma source de force et ma raison d'être. Peu importe ce que je fais ou dis, je ne pourrai jamais vous remercier suffisamment. Votre présence, votre amour et vos conseils ont été inestimables. Merci infiniment maman.

Ensuite, à mon cher père, la source de ma confiance, celui qui m'a toujours soutenu et protégé comme une perle rare. Aucun mot ne peut décrire la profondeur de mes sentiments à envers vous, votre amour inconditionnel est une bénédiction pour moi.

À ma grande sœur Mouna, je te suis reconnaissante pour ton soutien constant et tes encouragements. Tu as toujours été là pour moi, corrigeant mes erreurs et me guidant vers ce qui est juste. Je ne peux pas exprimer à quel point je suis chanceuse de t'avoir dans ma vie.

À chère Selma, je voulais te dire combien tu es spéciale pour moi. Ta présence est un cadeau précieux qui apporte bonheur et amour. Merci pour ta sollicitude et ton esprit bienveillant qui illuminent ma vie chaque jour.

À mon seul et petit frère Hamza, toutes les choses que tu as faites pour moi, tous les mots que tu as dits et les rires que nous avons partagés sont encore gravés profondément dans mon cœur. Je veux simplement te dire merci.

À mon grand-père Abdelkader, puisse Dieu t'accorder bénédiction et miséricorde, ainsi qu'à ma grand-mère Mébarka, mes oncles et mes tantes.

Je rends hommage à mon amie, l'infirmière Yasmine, pour son aide inestimable dans ma recherche. Sa passion, ses connaissances et sa générosité ont été une véritable source d'inspiration. Merci du fond du cœur pour ton soutien précieux.

À toutes les belles amitiés que j'ai pu construire pendant mon parcours universitaire : Meriem, Hafsa, Ikram, Ihcene, Yasmine, Manel, Rania, Lina et Rima.

Enfin, je tiens à exprimer ma gratitude envers mon binôme Meriem, avec qui j'ai partagé des moments inoubliables, bien que parfois difficiles. Grâce à son ambition et à son énergie positive, nous avons réussi à les surmonter avec succès.

**YASMINE**

## Dédicaces

À ma chère mère, un pilier de soutien et d'encouragement dans ma vie. Ta présence bienveillante et ton amour inconditionnel ont été une source d'inspiration inestimable. Je te dédie ces mots en reconnaissance de tout ce que tu as fait pour moi.

À mon cher père, qui a toujours été là pour moi avec sagesse, force et bienveillance. Tes sacrifices et ton dévouement ont façonné la personne que je suis aujourd'hui. Cette dédicace est un témoignage de mon amour et de ma gratitude éternels.

À mes deux frères Youcef et Mohamed, Vous êtes mes compagnons de vie, mes meilleurs amis. Notre lien fraternel est indéfectible et précieux . Je vous dédie cette dédicace pour toutes les aventures que nous avons partagées, pour les rires et les larmes, et pour le soutien mutuel qui nous a guidés à travers les hauts et les bas.

À ma famille, qui m'a entouré de son amour et de son soutien. Votre présence dans ma vie a apporté bonheur, réconfort et force.

À mes meilleures amies Ihcene, Ikrame, Yasmine et Amel qui m'ont toujours écouté, soutenu et supporté. Votre amitié sincère et votre soutien inconditionnel sont des trésors précieux.

Avec dédicace spéciale à Yasmine, qui as été ma partenaire inébranlable dans cette aventure. Ensemble, nous avons traversé des moments de travail acharné et de défis, mais nous avons également connu des réussites inoubliables. Je te dédie cette dédicace pour notre collaboration exceptionnelle et notre soutien mutuel.

À mon amie Anfal, future médecin, qui m'a apporté une aide précieuse dans ma recherche. Ta passion, tes connaissances et ta générosité ont été une source d'inspiration. Cette dédicace est un hommage à notre amitié et à ton soutien précieux.

À ma chère enseignante ,Ton départ cette année a laissé un vide immense dans nos cœurs. Tu as été une source d'inspiration et de bienveillance pendant mes années d'école primaire. Je te dédie cette dédicace en souvenir des enseignements que tu as partagés et de l'impact positif que tu as eu sur ma vie. Tu resteras à jamais dans nos pensées et nos souvenirs.

Que ces mots soient porteurs de reconnaissance, d'amour et de gratitude envers ceux qui ont joué un rôle significatif dans votre vie.

**YOUSSRA MERIEM**



## Résumé

Au cours des dernières années, l'utilisation des ontologies pour conceptualiser les connaissances de divers domaines d'application a connu une popularité croissante. Cependant, il est crucial de reconnaître les limites des ontologies lorsqu'il s'agit de traiter des problèmes impliquant des connaissances incertaines. Contrairement aux ontologies, les réseaux bayésiens sont spécifiquement conçus pour modéliser et raisonner sur l'incertitude.

L'objectif principal de notre travail de recherche est de concevoir un modèle de raisonnement probabiliste appliqué à une ontologie, en intégrant la connaissance incertaine. Pour atteindre cet objectif, nous avons élaboré une ontologie classique en définissant les concepts et les relations qui la composent. Ensuite, nous avons construit un réseau bayésien en définissant sa structure auprès d'un expert du domaine et en ajustant ses paramètres à partir d'un dataset sélectionné. Afin d'incorporer les connaissances incertaines du réseau bayésien standard à l'ontologie classique et de concevoir notre ontologie probabiliste, nous avons créé une méta-ontologie. De plus, nous avons développé un moteur de raisonnement capable de répondre à des tâches ou requêtes spécifiques.

Afin de valider notre travail, nous avons soumis nos contributions à des tests approfondis en collaboration avec l'expert du domaine, et leur efficacité a été confirmée pour la modélisation et le raisonnement sur des connaissances médicales, et plus particulièrement sur les maladies cardiovasculaires.

### **Mots clés :**

Raisonneur probabiliste, Connaissance incertaine, Réseau Bayésien, Ontologie, Ontologies Probabiliste.

## **Abstract**

In recent years, the use of ontologies to conceptualize knowledge in various application domains has grown in popularity. However, it is crucial to recognize the limitations of ontologies when dealing with problems involving uncertain knowledge. Unlike ontologies, Bayesian networks are specifically designed to model and reason about uncertainty.

The main objective of our research work is to design a probabilistic reasoning model applied to an ontology, integrating uncertain knowledge. To achieve this goal, we developed a classical ontology by defining the concepts and relations that compose it. We then built a Bayesian network by defining its structure from a domain expert and adjusting its parameters based on a selected dataset. To incorporate the uncertain knowledge of the standard Bayesian network into the classical ontology and design our probabilistic ontology, we created a meta-ontology. In addition, we have developed a reasoning engine capable of responding to specific tasks or queries.

To validate our work, we subjected our contributions to extensive testing in collaboration with the domain expert, and their effectiveness has been confirmed for modeling and reasoning on medical knowledge, and more specifically on cardiovascular diseases.

### **Key words :**

Probabilistic reasoner, Uncertain knowledge, Bayesian network, Ontology, Probabilistic ontologies.

## ملخص:

في السنوات الأخيرة، أصبح استخدام الأنطولوجيات سائغًا بشكل متزايد لتمثيل عدة مجالات تطبيقية. ومع ذلك، من الضروري التسليم بالقيود المفروضة على الأنطولوجيات عند التعامل مع المشاكل التي تنطوي على المعارف غير المؤكدة. على عكس الأنطولوجيات، تم تصميم شبكات بايز خصيصًا للنمذجة والاستدلال في حالة عدم اليقين.

الهدف الرئيسي لعملنا البحثي هو تصميم نموذج منطقي احتمالي مطبق على أنطولوجيا دامجة للمعرفة غير المؤكدة. لتحقيق هذا الهدف، طورنا أنطولوجيا كلاسيكية من خلال تحديد المفاهيم والعلاقات التي تؤلفها. تم قمنا ببناء شبكة بايز من خلال تعريف هيكلها من قبل خبير المجال وتعديل معالماتها بناءً على مجموعة بيانات مختارة. من أجل دمج المعرفة غير المؤكدة لشبكة بايز القياسية في الأنطولوجيا الكلاسيكية وتصميم الأنطولوجيا الاحتمالية الخاصة بنا، قمنا بإنشاء ميطا انطولوجيا. بالإضافة إلى ذلك، قمنا بتطوير محرك منطقي قادر على الاستجابة لمهام أو استفسارات محددة.

من أجل التحقق من صحة عملنا، أخضعنا مساهماتنا لاختبارات مكثفة بالتعاون مع خبير المجال، وتم تأكيد فعاليتها للنمذجة والاستدلال في المعرفة الطبيعية، وبشكل أكثر تحديدًا في أمراض القلب والأوعية الدموية.

## كلمات مفتاحية:

محرك الاستدلال، معرفة غير مؤكدة، شبكة بايزي، أنطولوجيا، أنطولوجيا احتمالية.

---

# Table des matières

---

<b>Table des figures</b>	<b>ix</b>
<b>Liste des tableaux</b>	<b>x</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 GESTION DE L'INCERTITUDE DANS LES ONTOLOGIES</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Les ontologies en ingénierie de connaissance . . . . .	4
1.2.1 Définition des ontologies . . . . .	5
1.2.2 Les composants d'une ontologie . . . . .	5
1.2.3 Les langages de représentation et de manipulation d'ontologies . . . . .	6
1.2.4 Raisonnement dans les ontologie classique . . . . .	7
1.2.5 Limitation des ontologies . . . . .	8
1.3 Connaissance incertaine . . . . .	8
1.3.1 Définition . . . . .	8
1.3.2 Gestion de l'incertitude dans les ontologies . . . . .	8
1.4 Réseaux bayésiens . . . . .	9
1.4.1 Définition . . . . .	9
1.4.2 Types des réseaux bayésiens . . . . .	9
1.4.3 Apprentissage dans le réseau bayésien . . . . .	10
1.4.4 Raisonnement incertain dans les réseaux bayésiens . . . . .	12
1.5 Les ontologies probabilistes . . . . .	12
1.5.1 La définition de l'ontologie probabiliste . . . . .	13
1.5.2 Raisonnement dans les ontologies probabilistes . . . . .	13
1.6 Travaux existants . . . . .	13
1.6.1 Comparaison des travaux . . . . .	15
1.6.2 Analyse des travaux : . . . . .	16
1.7 Conclusion . . . . .	17

<b>2</b>	<b>CONCEPTION DU SYSTÈME</b>	<b>18</b>
2.1	Introduction . . . . .	18
2.2	Motivation de l'étude de cas . . . . .	18
2.2.1	Définition de la maladie . . . . .	18
2.2.2	Les symptômes . . . . .	19
2.2.3	Les facteurs de risque . . . . .	19
2.3	Le schéma global du système . . . . .	19
2.4	Description du schéma global . . . . .	20
2.5	Le Dataset . . . . .	21
2.6	Prétraitement des données . . . . .	22
2.7	Construction de l'ontologie classique . . . . .	24
2.8	Construction de réseau bayésien . . . . .	26
2.8.1	Apprentissage de la structure . . . . .	27
2.8.2	Apprentissage des paramètres . . . . .	27
2.9	Méta-ontologie du réseau bayésien . . . . .	28
2.10	Construction de l'ontologie probabiliste . . . . .	29
2.11	Raisonnement probabiliste des connaissances ontologiques . . . . .	34
2.12	Conclusion . . . . .	36
<b>3</b>	<b>IMPLEMENTATION ET TEST DU SYSTEME</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Langage de Programmation . . . . .	37
3.3	Outils de développement . . . . .	37
3.4	API Utilisés . . . . .	38
3.5	Les interfaces graphiques . . . . .	38
3.6	Validation du système . . . . .	45
3.6.1	Validation du système par l'expert du domaine . . . . .	45
3.6.2	Validation du système par les mesures de performance . . . . .	48
3.7	Interprétation des résultats . . . . .	50
3.8	Conclusion : . . . . .	51
	<b>Conclusion et perspectives</b>	<b>52</b>
	<b>Bibliographie</b>	<b>54</b>

---

# Table des figures

---

2.1	Schéma global du système. . . . .	20
2.2	Taxonomie de Thing. . . . .	25
2.3	Taxonomie de Risk-Factor. . . . .	25
2.4	Taxonomie de Sign. . . . .	25
2.5	La structure de l'ontologie classique. . . . .	26
2.6	Les classes, les propriétés d'objets et les propriétés de données de l'ontologie classique. . . . .	26
2.7	La structure du réseau bayésien. . . . .	27
2.8	Le réseau bayésien après l'apprentissage des paramètres. . . . .	28
2.9	la structure de méta-ontologie. . . . .	28
2.10	Les classes, les propriétés d'objets et les propriétés de données de la méta-ontologie. . . . .	29
2.11	Les classes, les object property et les data property de l'ontologie finale. . . . .	30
2.12	Le code en OWL de la classe "Patient". . . . .	31
2.13	Le code en OWL de l'instance "Patient1". . . . .	31
2.14	Le code en OWL de la classe "Glucose". . . . .	31
2.15	Le code en OWL de l'instance "High_Glucose". . . . .	32
2.16	Le code en OWL de la classe "IsDiabetic". . . . .	32
2.17	Le code en OWL de l'instance "True_Diabetes". . . . .	33
2.18	Le code en OWL de la classe "IsSuspectedToCHD". . . . .	33
2.19	Le code en OWL de l'instance False_CHD. . . . .	34
3.1	Interface d'accueil. . . . .	39
3.2	Interface de raisonneur. . . . .	40
3.3	Interface de la fonctionnalité "subsomption probabiliste". . . . .	41
3.4	Interface de la fonctionnalité "Satisfiabilité probabiliste". . . . .	42
3.5	Interface de la fonctionnalité "Récupération d'instance". . . . .	43
3.6	Interface de la fonctionnalité "Incertitude existentielle". . . . .	44
3.7	Interface de la fonctionnalité "Incertitude de type d'individu". . . . .	45
3.8	Histogramme 1 : Precision, Recall, F-score, Accuracy. . . . .	50

---

# Liste des tableaux

---

1.1	Comparaison entre les Travaux . . . . .	16
2.1	Description du dataset. . . . .	22
2.2	Catégorisation des données. . . . .	23
2.3	Les attributs des concepts. . . . .	24
2.4	Les relations entre les concepts. . . . .	24
2.5	Les tâches de raisonnement probabiliste des connaissances ontologiques. . . . .	35
3.1	Les évidences du " <i>Patient1</i> ". . . . .	46
3.2	Les évidences du " <i>Patient2</i> ". . . . .	47
3.3	Les évidences de " <i>Patient3</i> ". . . . .	47
3.4	Les évidences de " <i>Patient4</i> ". . . . .	48
3.5	La matrice de confusion. . . . .	49
3.6	La matrice de confusion de notre système. . . . .	49





---

# Introduction générale

---

## Contexte

Dans un large éventail de domaines de recherche et d'application, l'accès à une connaissance précise et fiable est essentiel. Cependant, dans de nombreuses situations, les connaissances disponibles sont sujettes à l'incertitude. La représentation et le raisonnement sur ces connaissances incertaines sont donc devenus des enjeux majeurs pour les chercheurs en intelligence artificielle et en sciences de l'information.

Face à cette incertitude, une approche prometteuse consiste à adopter une perspective probabiliste. Les méthodes probabilistes offrent un cadre formel pour modéliser et raisonner sur des connaissances incertaines.

Dans ce contexte, les ontologies et les réseaux bayésiens se révèlent être des outils puissants pour traiter l'incertitude de manière formelle et structurée. En effet, l'intégration des ontologies et des réseaux bayésiens permet de combiner les avantages de ces deux approches. Elle offre la possibilité de représenter des connaissances incertaines de manière formelle en utilisant des ontologies, tout en bénéficiant des capacités de raisonnement probabiliste des réseaux bayésiens.

## Problématique

Les ontologies sont utilisées pour représenter les connaissances dans un domaine spécifique, de manière structurée et partageable, dans un format lisible et traitable par les machines. Cependant, l'un des principaux défauts des ontologies est leur incapacité à représenter et à raisonner sur l'incertitude.

Ces dernières années, une grande importance a été accordée à la représentation et au raisonnement de l'incertitude dans le domaine du Web sémantique. En effet, le groupe URW3-XG (Uncertainty Reasoning for the World Wide Web Incubator Group) recommande l'utilisation de la théorie des probabilités et des outils associés pour modéliser et représenter les connaissances incertaines dans ce domaine, en particulier l'utilisation des réseaux bayésiens.

La représentation des connaissances n'est pas un objectif final en soi, il est essentiel que les

connaissances représentées puissent être manipulées par un mécanisme de raisonnement approprié.

Dans ce contexte, il est donc nécessaire de combiner à la fois la modélisation de la connaissance incertaine pour l'intégrer dans une ontologie, et un mécanisme de raisonnement probabiliste pour l'appliquer aux connaissances intégrées. Ce mécanisme de raisonnement probabiliste doit être capable de réaliser des inférences sur les représentations de ces connaissances.

cela conduit à poser les questions suivantes :

1. Comment raisonner dans le cas d'incertitude et d'incomplétude des connaissances ontologiques ?
2. Comment inférer des connaissances implicites à partir des connaissances explicites stockées dans une ontologie en utilisant la théorie de probabilité ?

## Objectif

L'objectif de ce travail est de modéliser le raisonnement probabiliste appliqué à une ontologie qui intègre les connaissances incertaines. Nous cherchons donc à développer un raisonneur capable de répondre aux requêtes concernant cette incertitude. Le raisonneur utilisera les mécanismes de base de raisonnement présents dans les réseaux bayésiens, qui se fondent sur la théorie de probabilité et les techniques associées. Les méthodes bayésiennes permettent au système, d'intégrer les connaissances des experts, avec l'apprentissage machine, afin de fournir des modèles compréhensibles. De plus, elles offrent la possibilité de générer des prédictions probabilistes qui tiennent compte de l'incertitude lors de la prise de décisions.

## Organisation du mémoire

Dans le cadre de notre étude, le mémoire sera présenté en trois chapitres :

- **Chapitre 1 : GESTION DE L'INCERTITUDE DANS LES ONTOLOGIES** : Dans ce chapitre, nous explorerons les différentes notions liées aux ontologies classiques, ainsi que leur défi de représenter formellement les connaissances incertaines. Nous présenterons également les réseaux bayésiens et les ontologies probabilistes, en expliquant comment ils traitent l'incertitude. Ensuite, nous discuterons des différentes méthodes de raisonnement utilisées dans ces formalismes. Enfin, nous concluons en présentant des travaux pertinents dans ce domaine et en effectuant une analyse comparative des approches présentées.
- **Chapitre 2 : CONCEPTION DU SYSTÈME** : Ce chapitre expose la motivation de notre cas d'étude, ainsi que le choix et le prétraitement du dataset utilisé. Nous décrivons la construction des composants de notre système, en suivant les directives de l'expert du domaine. Ensuite, nous explorons l'intégration de ces composants en construisant notre

ontologie probabiliste. Enfin, nous abordons la conception de notre raisonneur probabiliste en détaillant ses différentes tâches de raisonnement.

- **Chapitre 3 : IMPLEMENTATION ET TEST DU SYSTÈME** Ce chapitre présente les résultats de la phase de conception, où nous détaillons la mise en oeuvre de notre système. Ensuite, nous procédons à l'évaluation et aux tests des performances de notre système afin de mesurer son efficacité.

Ce mémoire se terminera par une conclusion générale qui résume les principaux points abordés et offre quelques perspectives d'amélioration de notre travail.

## *Chapitre 1*

---

# **GESTION DE L'INCERTITUDE DANS LES ONTOLOGIES**

---

## **1.1 Introduction**

L'ontologie, dans les domaines de l'informatique et de la philosophie, désigne une représentation formelle des connaissances dans un domaine spécifique. Les ontologies classiques sont utilisées pour raisonner de manière déterministe, en se basant sur des faits établis et des relations bien définies. Cependant, de nombreuses situations réelles impliquent une certaine dose d'incertitude. Ainsi, pour aborder ce défi, différents chercheurs ont proposé d'utiliser les réseaux bayésiens en conjonction avec les ontologies. Ceux-ci offrent une approche puissante pour modéliser et raisonner sur l'incertitude basée sur la théorie des probabilités.

Ce chapitre a pour objectif de présenter, en premier lieu, les différentes notions liées aux ontologies classiques. Ensuite, nous aborderons le problème de l'incertitude dans le domaine de la représentation formelle des connaissances. Une autre partie sera consacrée à la présentation des réseaux bayésiens et des ontologies probabilistes. De plus, nous expliquerons les diverses méthodes de raisonnement utilisées dans les ontologies classiques, les réseaux bayésiens et les ontologies probabilistes. Nous terminerons par la présentation de certains travaux liés à ce domaine et par une analyse comparative.

## **1.2 Les ontologies en ingénierie de connaissance**

Ontologie est un terme qui a tout d'abord été défini en Philosophie comme une branche de la Métaphysique qui s'intéresse à l'existence, à l'être en tant qu'être et aux catégories fondamentales de l'existant. En effet, ce terme est construit à partir des racines grecques *ontos* qui veut dire ce qui existe, l'Être, l'existant, et *logos* qui veut dire l'étude, le discours, d'où sa traduction par l'étude de l'Être et par extension de l'existence [1].

Au début des années 80, les chercheurs en intelligence artificielle ont emprunté le terme « Onto-

logie » au champ de la philosophie. Les ontologies sont devenues la définition des connaissances d'un domaine. Elles fournissent la possibilité de séparer les connaissances du domaine des connaissances opérationnelles [35].

Au début des années 1990, le terme 'Ontologie' s'est propagé dans la communauté de l'ingénierie des connaissances, et depuis il a un très grand succès [4].

### 1.2.1 Définition des ontologies

Dans le cadre de l'intelligence artificielle, les ontologies sont nées du besoin de représenter les connaissances dans les systèmes informatiques, de ce fait, elles sont toujours définies par rapport au processus général de la représentation des connaissances [13].

Des différentes définitions sont proposées dans la littérature pour la notion d'ontologie, Neeches et ses collègues [28], 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 [15], 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. Cette définition a été modifiée légèrement par Borst comme « spécification formelle d'une conceptualisation partagée ». Ces deux définitions sont regroupées dans celle de Studer [35], comme « spécification formelle et explicite d'une conceptualisation partagée ». Ils l'expliquent comme suit :

- **Spécification explicite** : signifie que les concepts, les propriétés, les relations, les fonctions, les restrictions et les axiomes de l'ontologie sont définis de façon déclarative.
- **Formelle** : réfère au fait qu'une ontologie doit être traduite dans un langage interprétable par une machine (machine-readable).
- **Conceptualisation** : réfère à un modèle abstrait d'un phénomène du monde en identifiant les concepts appropriés à ce domaine.
- **Partagé** : réfère au fait qu'une ontologie capture la connaissance consensuelle c'est-à-dire non réservée à quelques individus, mais partagée par un groupe ou une communauté.

### 1.2.2 Les composants d'une ontologie

Pour formalisées les connaissances d'un domaine en utilisant principalement les cinq composants d'une ontologie qui sont proposer par Gomez et Benjamins en 1999 [14], comme suit :

#### 1. Les concepts

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 constitue son extension. Un concept est composé de trois parties [17] :

- Une notion : elle correspond à la sémantique du concept, elle est définie à travers ses propriétés et ses attributs. Elle est appelée intention du concept.
- Un ensemble d'objets : il correspond aux objets définis par le concept, il est appelé extension du concept. Les objets sont les instances du concept.
- Un (ou plusieurs) terme(s) : les termes permettent de désigner le concept. Ces termes sont aussi appelés labels de concept.

## 2. Les relations

Une relation est un lien entre des entités, exprimée souvent par un terme ou un symbole littéral. Elle peut être caractérisée par une signature qui précise le nombre d'instances de concepts que la relation lie, leurs types et l'ordre des concepts [10].

## 3. Les fonctions

Elles présentent des cas particuliers de relations dans lesquelles le nième (n) élément de la relation est unique pour les (n-1) éléments précédents [14]. Formellement, les fonctions sont définies telles que :  $F : c_1 * c_2 * \dots * c_{n-1} \rightarrow c_n$ .

## 4. Les axiomes (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 [34].

5. **Instances** Les instances constituant la définition extensionnelle de l'ontologie, ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine du problème [24].

### 1.2.3 Les langages de représentation et de manipulation d'ontologies

#### 1. XML

Le langage eXtensible Markup Language (XML) soutenu par le W3C depuis février 1998 pour le balisage des documents, conçu et optimisé pour la transmission des données via le Web, XML fournit un mécanisme simple mais robuste pour l'encodage des informations sur le Web sémantique [3].

#### 2. RDF

Le langage RDF est un modèle de méta-données pour référencer des objets (ressources) et les liens entre ces objets. Dans ce langage, les ressources sont identifiées par les URIs (Uniform Resource Identifier). Il s'agit d'un langage qui permet de décrire la sémantique

des données, de manière compréhensible par les machines. Le modèle de données RDF est constitué de trois types de composantes [26] :

- Sujet : il représente la ressource à décrire. Il est nécessairement identifié par une URI.
- Prédicat ou Propriété : il s'agit d'une propriété utilisée pour caractériser et décrire une ressource. Un prédicat est nécessairement identifié par une URI. Il exprime un lien ou une relation entre le sujet et l'objet.
- Objet : représente une donnée ou une autre ressource (identifiée par une URI).

### 3. OWL (Web Ontology Language)

Dans le but d'étendre l'expressivité du langage RDFS, le groupe de travail sur les ontologies du W3C, WebOnt, a recommandé le langage OWL comme un standard pour la représentation d'ontologies depuis février 2004 [25]. En plus de ses constructeurs permettant de décrire des classes et des propriétés plus complexes, OWL intègre des outils de comparaison de propriétés et de classes. Il permet ainsi d'exprimer des propriétés telles que l'équivalence, la disjonction entre classes, la cardinalité, la symétrie, la transitivité d'une relation, etc. C'est un langage inspiré des logiques de description [2], qui offre aussi de nouvelles primitives permettant de définir des classes à l'aide de mécanismes ensemblistes (intersection de classes, union de classes, complément d'une classe). Il permet également d'effectuer des déductions. Ainsi, OWL peut être utilisé pour représenter explicitement une ontologie. De plus, en permettant la mise en relation d'ontologies différentes, OWL permet la prise en compte de l'aspect distribué des ontologies. Grâce à son pouvoir expressif et son pouvoir de raisonnement, il demeure un puissant langage de description d'ontologies [10].

## 1.2.4 Raisonnement dans les ontologie classique

Dans l'ontologie classique, le raisonnement fait référence à l'utilisation de l'ontologie pour déduire de nouvelles informations à partir des connaissances existantes. Les tâches de raisonnement dans l'ontologie classique incluent notamment :

### 1. La subsomption des concepts

La subsomption des concepts dans une ontologie est une relation qui exprime l'inclusion d'une classe (concept) dans une autre classe de manière hiérarchique. Elle indique qu'une classe est plus spécifique qu'une autre classe, selon [19], cette tâche vise à tester si un concept  $C1$  est plus général qu'un concept  $C2$  (Cette relation est notée  $C2 \subseteq C1$ ).

### 2. La satisfiabilité des concepts

La satisfiabilité des concepts dans l'ontologie indique si un concept donné peut être réalisé ou satisfait par au moins un individu dans de l'ontologie. Autrement dit un concept  $C$  est satisfiable si et seulement s'il possède au moins un individu  $I$  [19].

### 3. **Equivalence de concepts**

Permet de vérifier si la base de connaissances implique que deux classes  $C1$  et  $C2$  sont équivalentes [19], (on note  $C1 \equiv C2$ ). Plus précisément, lorsqu'on définit deux concepts comme équivalents, c'est-à-dire que tous les individus qui appartiennent à l'un de ces concepts appartiennent aussi à l'autre concept, et vice versa.

### 4. **La vérification d'instance**

cette tâche permet de vérifier si un individu donné peut être considéré comme une instance d'une classe spécifique dans l'ontologie.

### 5. **Récupération d'instances**

La récupération d'instances dans le domaine de l'ontologie fait référence au processus de recherche et d'extraction des individus.

### 6. **La vérification de disjonction entre deux concepts**

Selon [21], deux concepts  $C1$  et  $C2$  sont disjoints, si et seulement si  $C1 \cap C2 = \emptyset$  pour tous modèles  $I$ . En clair, la disjonction consiste à déterminer si les deux concepts sont mutuellement exclusifs et ne peuvent pas avoir d'instances en commun. (notée  $C1 \vee C2$ ).

## 1.2.5 **Limitation des ontologies**

Bien que les ontologies classiques jouent un rôle majeur dans la représentation formelle des connaissances, leur principal défaut réside dans leur incapacité à représenter et à raisonner sur l'incertitude, une caractéristique inévitable dans le monde réel.

## 1.3 **Connaissance incertaine**

La fiabilité de l'information est un enjeu crucial dans divers domaines scientifiques. Il est donc nécessaire de traiter cette incertitude afin d'obtenir une représentation adéquate du monde réel.

### 1.3.1 **Définition**

Le terme "incertitude" tel que le définit le dictionnaire français Larousse désigne une information qui n'est pas établie avec certitude, qui peut ou non se produire et qui peut être de nature vague. Le but sera alors d'attribuer une valeur de probabilité ou une possibilité à une proposition pour que cette dernière soit vraie ou fausse [20],

### 1.3.2 **Gestion de l'incertitude dans les ontologies**

Une représentation adéquate d'un domaine des connaissances réelles doit être capable de capturer l'incertitude. Il existe de nombreux formalismes différents qui peuvent être utilisés pour



mesurer et gérer l'incertitude. La théorie des probabilités est l'un des plus connus formalismes pour traiter l'incertitude [30].

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 [8].

## 1.4 Réseaux bayésiens

Les réseaux Bayésiens sont des techniques de choix dans la représentation de connaissances et dans l'exploitation de celles-ci qui tient compte de l'incertitude dans le raisonnement. Ils sont utilisés dans un bon nombre d'applications. Nous définissons, dans cette section, quelques notions de base qui nous seront utiles dans notre travail

### 1.4.1 Définition

Les réseaux Bayésiens sont des modèles qui représentent des connaissances incertaines sur des phénomènes complexes. Un réseau Bayésien peut se décomposer en deux parties :

- une représentation graphique par un graphe orienté acyclique. Les variables aléatoires sont représentées par des nœuds contenant leurs états, et les relations de dépendance entre ces variables par des arcs. Les nœuds sont reliés par la relation causale (relation de cause à effet).
- une représentation probabiliste par un ensemble de tables de probabilités conditionnelles (CPT). À chaque nœud est associé une CPT de taille exponentielle au nombre de parents et à leur nombre d'états [36].

### 1.4.2 Types des réseaux bayésiens

Il existe plusieurs types des réseaux bayésiens, Parmi les types il ya

#### 1. Réseaux bayésiens dynamiques

Les réseaux bayésiens dynamiques sont une extension des réseaux bayésiens dans laquelle l'évolution temporelle des variables est représentée[27], cela signifie, que l'état actuelle d'une variable peut dépendre de l'état à un instant passé. Ce type de réseau bayésien est inspiré des chaînes Markov cachées[27].

#### 2. Réseaux bayésiens orientés objet

Les reseaux bayesiens orientes objet sont utiliser dans le cas où une sous-structure apparaît de manière répétée, autrement, ils permettent également de modéliser des systèmes

complexes pour lesquels le même mode de raisonnement apparaît dans différents sous-systèmes[27].

### 3. Réseaux bayésiens distribués

Réseaux bayésiens distribués vient pour solliciter le problème de la complexité temporelle et de la complexité spatiale, il consiste à organiser le réseau bayésien en plusieurs réseaux de plus petite taille[27].

### 4. Les réseaux bayésiens multi-entités (MEBN)

Le réseau bayésien multi-entité est un formalisme qui unifie la logique du premier ordre et la théorie des probabilités. Les MEBN sont formés de fragments (appelés MFragments) qui représentent alors la distribution jointe d'un sous-ensemble de variables[11].

### 5. Réseaux bayésiens flous

Les Réseaux Bayésiens Flous (RBFs) combinent les capacités des réseaux bayésiens et de la logique floue pour bénéficier des avantages des deux modèles à la fois. En effet un RBF est une extension de réseau bayésien permettant de traiter les probabilités des événements sous des observations floues [19].

### 6. Le réseau bayésien hybride

Un réseau bayésien est dit hybride s'il contient un ensemble de variables aléatoires discrètes et continues représentées par les noeuds d'un graphe [9].

## 1.4.3 Apprentissage dans le réseau bayésien

La construction de réseaux bayésiens consiste à trouver une structure ou graphe et estimer les paramètres, Cependant, face à une base de données très volumineuse où l'on souhaite extraire une structure adaptée pour l'inférence dans un graphe, la tâche peut devenir complexe. C'est là qu'intervient la méthode d'apprentissage, qui vise à résoudre ce problème.

L'apprentissage d'un RB se décompose en deux phases :

- Apprentissage de la structure : trouver la structure à partir de données observées.
- Apprentissage des paramètres : où les probabilités conditionnelles de chaque nœud de réseau devraient être évaluées.

#### 1. Apprentissage de la structure

L'apprentissage de la structure d'un RB permet de déterminer automatiquement une structure optimale d'un RB à partir de l'information contenue dans les données observées [29].

Beaucoup de travaux a été fait pour découvrir cette structure à partir des données complètes ou incomplètes. Vous trouverez ci-dessous quelques algorithmes fondés sur les données complètes et incomplètes :

##### a) Algorithmes basés sur un score (Données complètes)

Le principe de cette famille d'algorithme est de parcourir tous les graphes possibles

en leur donnant un score. Puis sélectionner le graphe qui présente le résultat le plus élevé, mais cette méthode est complexe, principalement en raison de la taille super-exponentielle de l'espace de recherche en fonction du nombre de variables.

**b) Algorithmes de structure (Données incomplètes)**

La méthode des données incomplète est basée sur le principe de l'algorithme EM. Cette méthode itérative part d'une structure initiale pour estimer la distribution de probabilité des variables cachées ou manquantes grâce à l'algorithme EM classique. L'espérance d'un score par rapport à ces variables cachées est ensuite calculée pour tous les réseaux bayésiens du voisinage afin de choisir la structure suivante [12].

**2. Apprentissage des paramètres**

L'apprentissage des paramètres consiste à estimer la distribution des probabilités au niveau des nœuds. Il est fondé sur des données qui ne sont pas nécessairement complètes. Différents algorithmes sont utilisés pour l'apprentissage des paramètres dans les deux cas.

**a) Apprentissage des données complètes**

Nous cherchons ici à estimer les distributions de probabilités à partir de données disponibles [22].

— **Apprentissage statistique**

Dans le cas où toutes les variables sont observées, la méthode la plus simple et la plus utilisée est l'estimation statistique qui consiste à estimer la probabilité d'un événement par la fréquence d'apparition de l'événement dans la base de données. Cette approche, appelée maximum de vraisemblance (MV) [22].

— **Apprentissage bayésien**

L'estimation bayésienne suit un principe quelque peu différent. Il consiste à trouver les paramètres  $\theta$  les plus probables sachant que les données ont été observées, en utilisant des a priori sur les paramètres [22].

**b) Apprentissage des données incomplètes**

Dans les applications pratiques, les bases de données sont très souvent incomplètes. Certaines variables ne sont observées que partiellement ou même jamais, que ce soit en raison d'une panne de capteurs, d'une variable mesurable seulement dans un contexte bien précis, d'une personne sondée ayant oublié de répondre à une question, etc [22].

Différentes approches ont été présentées pour cet apprentissage. Nous ne citerons ici que la méthode d'estimation utilisée, qui est l'algorithme itératif EM (Expectation Maximisation) [29].

— **Algorithme EM (Expectation-Maximisation)**

L'algorithme EM est fondé sur la répétition de deux phases jusqu'à ce que le

maximum de vraisemblance soit atteint, une fois les valeurs manquantes initialisées :

**Étape E « Expectation »** Utilisation des paramètres courants ( $\theta(t)_{ijk}$ ) pour estimer l'espérance d'apparition des différentes configurations [11].

**Étape M « Maximisation »** Estimation des nouveaux paramètres par maximum de vraisemblance (ou maximum a posteriori) en utilisant l'espérance des statistiques essentielles obtenue à l'étape précédente [11].

#### 1.4.4 Raisonnement incertain dans les réseaux bayésiens

Les réseaux bayésiens sont particulièrement utiles en raison de la possibilité d'inférer des probabilités à travers eux [31]. En effet, l'inférence bayésien est une sorte de calcul de probabilités conditionnelles. il s'agit de mettre à jour des probabilités des variables non observées après l'observation des valeurs d'un certain nombre d'autres variables afin de calculer les probabilités a posteriori [19]. Donc selon [27], les réseaux bayésiens permettent de mettre à jour les distributions de probabilités au niveau des variables : on commence par attribuer des valeurs aux variables d'observation, puis on utilise un algorithme d'inférence pour calculer les probabilités a posteriori des autres variables.

##### 1. Inférence exacte

L'inférence exacte, également appelée inférence déterministe ou inférence certaine, les réseaux bayésiens est le calcul précis des probabilités conditionnelles en fonction des informations contenues dans le réseau, à l'aide de règles probabilistes et d'algorithmes spécifiques. L'inférence exacte dans les RBs est un problème NP-complet. En pratique, la complexité des algorithmes d'inférence dépend de plusieurs facteurs dont notamment le nombre de nœuds du réseau bayésien, la structure du réseau bayésien, la taille des domaines des variables et le nombre de parents d'un nœud. Ce constat peut amener à utiliser des algorithmes d'inférence approchées [27].

##### 2. Inférence approchée

Lorsqu'un réseau bayésien est complexe ou si les tables de probabilités conditionnelles ne sont pas exactes, l'intérêt d'effectuer une inférence exacte avec ces valeurs approximatives n'est plus probant [11]. Pour cette raison, la méthode approximative a été proposée. Selon [27], ces méthodes donnent une estimation des probabilités a posteriori et sont non déterministes.

### 1.5 Les ontologies probabilistes

Selon Emna hlel et al. [18], l'un des principaux défauts de l'ontologie est leur incapacité de représenter et raisonner sur l'incertitude. Dans les années précédentes, divers chercheurs ont essayé de combiner les RBs avec les ontologies afin de représenter et raisonner avec l'incertitude

et par conséquence des extensions du formalisme Standard OWL ont été proposées comme BayesOWL, OntoBayes et PR-OWL.

### 1.5.1 La définition de l'ontologie probabiliste

Costa [5], définit une ontologie probabiliste comme étant une représentation formelle et explicite de la connaissance concernant un domaine d'application. Autre que les propriétés d'OWL, une ontologie probabiliste comprend :

- Régularités statistiques qui caractérisent le domaine.
- Les connaissances non concluantes, ambiguës, incomplètes, non fiables et dissonantes connaissances liées aux entités du domaine.
- L'incertitude relative à toutes les formes de connaissances, où le terme entité fait référence à tout concept (réel ou fictif, concret ou abstrait) qui peut être décrit et raisonné dans le domaine d'application.

### 1.5.2 Raisonnement dans les ontologies probabilistes

En plus des tâches de raisonnement présentées dans les ontologies classiques, le raisonnement dans l'ontologie probabiliste permet d'effectuer des tâches incertaines, d'après [32] trois grandes tâches de raisonnement doivent être mises en œuvre lors du traitement des connaissances probabilistes en ontologie.

1. L'incertitude existentielle.
2. L'incertitude de type de individu.
3. L'incertitude liée aux propriétés.
  - a) l'incertitude liée aux propriétés d'objets.
  - b) l'incertitude liée aux propriétés de données.

## 1.6 Travaux existants

### 1. Costa et al., 2008 [6] :

Cet article propose une extension d'OWL qui ajoute de nouvelles caractéristiques, PR-OWL fournit des constructeurs particuliers pour modéliser la connaissance incertaine, ainsi il permet de raisonner grâce à un moteur d'inférence probabiliste. PR-OWL est basé sur la logique de la Multi-Entity Bayesian Network (MEBN) qui est un système formel pour la représentation des connaissances intégrant la logique de premier ordre avec la théorie bayésienne des probabilités, à cause de ça il est flexible et expressif. Il fonctionne comme une ontologie de haut-niveau, composée d'un ensemble de classes, de sous-classes et de propriétés qui permettent de modéliser la partie incertitude de l'ontologie

en utilisant MEBN. L'ontologie probabiliste est modélisée à l'aide de la classe MTheory, composée par un ensemble de MFrag. Ces MFrag doivent collectivement former une MTheory cohérente. Les MFrag sont construits à partir de variables aléatoires, qui ont une distribution probabiliste et un ensemble exhaustif d'états possibles.

## 2. **Zhang et al., 2009 [37] :**

les auteurs ont créé une extension probabiliste appelée BayesOWL qui augmente et améliore OWL pour représenter et raisonner avec l'incertitude en se basant sur les réseaux bayésiens. BayesOWL fournit un ensemble de règles et de procédures pour la traduction automatique d'une ontologie OWL en une structure RB ainsi que des algorithmes qui incorporent des informations probabilistes sur les classes et les relations interclasses dans les CPT de RB. La traduction d'une ontologie OWL qui décrit la taxonomie des concepts en RB par le système prototype se fait en deux étapes :

- La première étape consiste à déconstruire la structure du RB à partir du fichier d'entrée de l'ontologie OWL et à initialiser les tables de probabilités conditionnelles (CPT) avec des valeurs par défaut. Ceci est fait par l'analyseur syntaxique de taxonomie (T-Parser en abrégé) et le constructeur de structure RB.
- La deuxième étape consiste à d'incorporer les informations probabilistes fournies par l'utilisateur concepts et des relations inter-concepts dans les CPT de la RB. Ceci est fait par l'analyseur de probabilités (P-Parser en abrégé) et le constructeur de CPT.

Le RB traduit est écrit dans le format de Netica (un logiciel de développement de réseaux bayésiens de Norsys Software Corporation ). En plus d'inclure une interface utilisateur graphique (GUI) pour une utilisation directe, BayesOWL a également fourni une interface de programmation d'applications (API) pour les développeurs des logiciels sous la forme de paquets Java.

## 3. **Foni A. Setiawan et al., 2019 [33] :**

Dans ce travail les auteurs ont proposé une nouvelle approche d'intégration des ontologie et des RB en transformant les informations logiques contenues dans une ontologie en un RB et les informations probabilistes contenues dans un RB en une ontologie. Selon les auteurs, ce travail est le premier à intégrer les RB dans OWL et vice versa. Cette Framework se compose de trois parties principales : l'application, le raisonneur et la base de connaissances. L'application peut être un logiciel indépendant ou un agent logiciel. Le Raisonneur comprend deux composants :le Raisonneur logique et le Raisonneur probabiliste. La base de connaissances contient deux formes : des connaissances ontologiques au format OWL/RDF et des connaissances RB au format XML of Decision Systems Laboratory (XDSL). Dans la base de connaissances, il existe un composant appelé Morpher qui effectue la transformation de données à partir d'ontologies standardisées (OWL/RDF) vers XDSL (et vice versa). Les principales contributions de cette recherche sont les suivantes :

- Enrichissement de l'ontologie basé sur les bases de connaissances RB.

— Ajustement structurel de RB par apprentissage structurel basé sur des bases de connaissances ontologiques.

— Support pour une clause probabiliste intégrée dans le format d'une SPARQL requête.

#### 4. **Hai, N et al., 2021 [16] :**

Cet article propose une méthode pour fournir une alerte de risque intégrée dans les corridors de transport. En combinant la méthode de décomposition WBS-RBS, la construction d'une base de connaissances ontologiques et la méthode du réseau bayésien, un système intégré d'identification et d'alerte précoce des risques liés aux corridors de pipelines est créé pour identifier les risques potentiels dans les corridors et fournir des alertes en temps réel. La méthode proposée a été évaluée sur un cas d'étude réel et les résultats montrent l'efficacité de la méthode pour fournir des alertes de risque intégrées en temps réel.

### 1.6.1 Comparaison des travaux

D'après les travaux présentés nous avons effectué une comparaison basée sur les critères suivants :

— Le langage utilisé : indique le langage de représentation d'ontologies.

— Le type de réseau bayésien : indique le type de réseau bayésien utilisé.

— Le principe de l'approche : décrit décrire la méthode utiliser pour chaque travail.

— Mécanisme de raisonnement : décrit le mécanisme de raisonnement utilisé dans chaque travail.

Le tableau 1.1 présente cette comparaison.

Travaux	Langage utilisé	Type de RB	Principe de l'approche	Mécanisme de raisonnement
Costa et al., 2008	OWL	Standard / MEBN	Développement d'une extension du langage du web sémantique qui permet la construction d'un RB à partir d'une ontologie et l'intégration sémantique du RB dans l'ontologie	La génération d'un RB spécifique à la situation (SSBN), qui consiste à construire un RB minimale suffisante pour calculer les valeurs de probabilité des nœuds cibles sur la base des valeurs de preuve existantes
Zhang et al., 2009	OWL	Standard	Développement d'une extension du langage du web sémantique qui permet la construction d'un RB à partir d'une ontologie et l'intégration sémantique du RB dans l'ontologie	L'utilisation de l'algorithme de la procédure d'ajustement proportionnel itératif décomposé (D-IPFP) pour calculer les valeurs de probabilité pour chaque nœud du système.
Foni A.Setiawan et al., 2019	OWL / RDF	Standard / Dynamique / Orienté objet	Développement des algorithmes qui permettent d'intégrer les RB en ontologie et vice-versa dans une seule base de connaissances	Il est basé sur l'envoi des requêtes SPARQL. Il se déroule en deux étapes. La première étape consiste à effectuer un raisonnement probabiliste dans la base de connaissances RB. La deuxième étape, est un raisonnement dans la base de connaissances ontologique, qui se base sur le raisonnement de la première étape puis le résultat final de la requête est renvoyé.
Hai, N et al., 2021	OWL	standard	la combinaison d'une ontologie construite à partir d'une liste de risque établie en utilisant la méthode de décomposition WBS-RBS avec un réseau bayésien modélisé sur la base de cette ontologie et les connaissances des experts pour l'évaluation des facteurs de risque identifiés.	Il est basé sur la détection des risques dans les corridors de transport en calculant les probabilités à partir de données historiques et de l'expertise des spécialistes du domaine et les facteurs de risque ont été évalués avec l'inférence bayésienne.

TABLE 1.1 – Comparaison entre les Travaux

## 1.6.2 Analyse des travaux :

La comparaison entre ces études nous a permis de remarquer que chaque travail se base sur un langage, type de RB et une approche différente pour atteindre un objectif précis.



- Les travaux de Zhang et al., Costa et al. cherchent à développer le langage OWL en proposant des extensions probabilistes, les deux approches utilisent le langage OWL et les RBs Standards, mais l’approche de PR-OWL a également utilisé les RBs de type MEBNs.
- Foni A. Setiawan et al. ont proposé le Framework ByNowLife qui intègre les RB dans OWL et vice versa. Ce travail est basé sur les langages OWL / RDF pour modéliser les connaissances dans l’ontologie et utilise les RBs de type Standard ou Dynamique ou bien Orienté objet pour représenter les connaissances des RBs et pour effectuer la transformation des connaissances dans les deux sens le Morpher est utilisé.
- Dans leur étude, Hai, N. et al. ont proposé une approche qui combine une base de connaissances ontologique avec des réseaux bayésiens dans le but d’optimiser la détection et la gestion des risques dans les corridors de transport. Cette approche intègre les connaissances expertes ainsi que les données historiques pour une meilleure compréhension et modélisation des risques.

Bien que les chercheurs tentent de résoudre les problèmes liés à l’incertitude et au raisonnement, Cependant, leurs travaux reconnaissent des limitations :

- PR-OWL offre plusieurs avantages grâce aux MEBNs, mais cela ne préserve pas des problèmes tel que le délai d’exécution est lent, Le MEBN traite uniquement des données discrètes.
- Le Framework de BayesOWL est implémenté seulement dans une taxonomie OWL, sans inclure les propriétés et les types de données.
- Dans l’approche du Foni A. Setiawan et al. l’inconvénient principale est que les caractéristiques sémantiques des relations ontologiques ne sont pas prises en considération lors de la construction du BN.
- Dans l’article de Hai, N et al, les auteurs reconnaissent plusieurs limitations de leur modèle. Tout d’abord les types d’incidents à risque étudié dans ce document sont peu nombreux, et les facteurs de risque identifiés sont limités. En plus, la construction de l’ontologie et du réseau bayésien dépend de l’expertise des professionnels impliqués, ce qui peut être difficile à trouver.

## 1.7 Conclusion

Ce chapitre a abordé divers aspects des ontologies classiques, de l’incertitude dans la représentation des connaissances, des réseaux bayésiens et des ontologies probabilistes. Il a mis en évidence l’importance de ces notions pour une modélisation précise et robuste des connaissances, tout en expliquant les différentes méthodes de raisonnement des ontologies classiques, des réseaux bayésiens et des ontologies probabilistes. On a fini par la présentation de quelques travaux connexes relatifs à l’utilisation des ontologies en conjonction avec les réseaux bayésiens. Dans le prochain chapitre, nous nous concentrerons sur la conception de notre propre système.

## Chapitre 2

---

# CONCEPTION DU SYSTÈME

---

## 2.1 Introduction

Après la démonstration de tout ce qui touche aux composants fondamentaux de notre système, nous aborderons en détail la conception de notre raisonneur probabiliste, un système capable d'effectuer des inférences sur une ontologie probabiliste.

Au cours de ce chapitre, nous allons d'abord expliquer les raisons qui nous motivent à choisir la maladie dont les connaissances sont représentées dans notre ontologie probabiliste. Ensuite, nous aborderons le processus de construction de cette ontologie. De plus, nous mettrons en œuvre les tâches de raisonnement probabiliste dans notre raisonneur.

## 2.2 Motivation de l'étude de cas

Les maladies cardiovasculaires (MCV) sont la principale cause de décès dans le monde, représentant environ 31% de tous les décès. Les personnes qui souffrent de maladies cardiovasculaires ou qui présentent un risque élevé de développer ces maladies en raison de la présence de facteurs de risque tels que l'hypertension, le diabète, l'hyperlipidémie ou une maladie déjà présente, nécessitent une détection précoce et une prise en charge appropriée. Cette prise en charge doit inclure un soutien psychologique ainsi que des médicaments adaptés aux besoins individuels.<sup>1</sup>

### 2.2.1 Définition de la maladie

Le cœur est une puissante pompe musculaire dans le corps qui assure l'approvisionnement en oxygène et en nutriments grâce au sang qu'il contient. Pour accomplir cette fonction de pompage essentielle, le cœur est alimenté en sang oxygéné par les artères coronaires. Selon l'Organisation mondiale de la Santé (OMS), les maladies coronariennes font partie des maladies

---

1. [https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))

cardio-vasculaires et sont caractérisées par une altération du flux sanguin vers le muscle cardiaque en raison d'une réduction ou d'une obstruction des artères coronaires.<sup>2</sup>

### 2.2.2 Les symptômes

Selon la British Heart Foundation (BHF), la maladie coronarienne (CHD) se développe progressivement au fil du temps et les symptômes peuvent varier d'une personne à l'autre. L'angine est le terme utilisé pour décrire les symptômes les plus courants de la maladie coronarienne.<sup>3</sup>

Ces symptômes comprennent :

- Douleur thoracique
- Essoufflement
- Douleur voyageant à travers le corps
- Sentiment de faiblesse
- Nausée.

### 2.2.3 Les facteurs de risque

Selon la British Heart Foundation (BHF), de nombreux facteurs de risque peuvent augmenter le risque de maladie coronarienne.<sup>4</sup> Les principaux facteurs de risque comprennent :

- L'hypertension artérielle
- Taux de cholestérol élevé
- Diabète
- Fumeur
- Obésité

## 2.3 Le schéma global du système

Dans le cadre de notre recherche, notre objectif est de présenter un modèle de raisonnement probabiliste capable de traiter l'incertitude. Pour cela, nous devons à la fois représenter cette incertitude dans une base de connaissances ontologique et définir un mécanisme de raisonnement qui effectue des inférences probabilistes sur ces connaissances.

Pour démontrer l'efficacité de notre système, nous avons réalisé une étude de cas portant sur le

---

2. [https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))

3. <https://www.bhf.org.uk/information-support/conditions/coronary-heart-disease>

4. <https://www.bhf.org.uk/information-support/conditions/coronary-heart-disease>

diagnostic des maladies coronariennes. La figure 2.1 présente de manière visuelle le processus de conception de notre système.

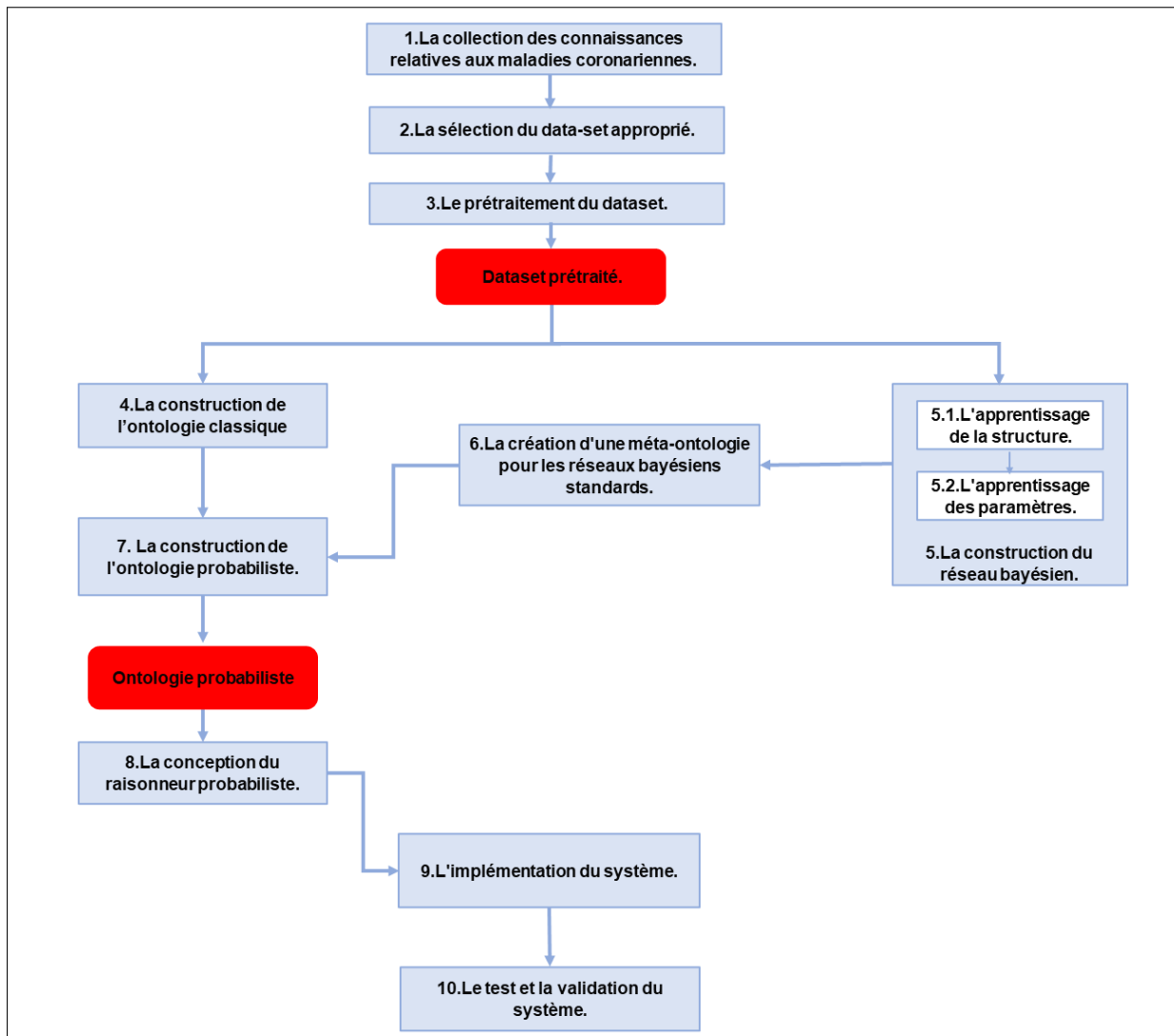


FIGURE 2.1 – Schéma global du système.

## 2.4 Description du schéma global

Dans le cadre de notre recherche, nous avons suivi les étapes suivantes :

1. Le processus de conception de notre système commence par la consultation d'un expert du domaine, un cardiologue, afin d'acquérir les connaissances nécessaires concernant les maladies coronariennes.
2. En nous basant sur ces connaissances, nous avons choisi le dataset approprié à notre cas d'étude.
3. Par la suite, nous avons effectué les prétraitements nécessaires sur le dataset choisi.

4. En se basant sur les connaissances acquises, nous avons établi les concepts et les relations de notre ontologie classique.
5. Pour pouvoir construire le réseau bayésien nous avons passé par deux étapes :
  - 5.1. En suivant les directives de l'expert, nous avons établi les nœuds et les liens de la structure du réseau bayésien.
  - 5.2. En utilisant l'algorithme EM, nous avons effectué l'apprentissage des paramètres du réseau bayésien défini.
6. Afin d'incorporer les connaissances incertaines du réseau bayésien à notre ontologie, nous avons construit un méta-modèle capable de représenter les réseaux bayésiens standards.
7. Ensuite, nous avons intégré l'ontologie de méta-modèle à l'ontologie classique afin de construire notre ontologie probabiliste.
8. Pour mettre en œuvre le raisonneur, nous avons développé des algorithmes de raisonnement probabiliste basés sur les inférences des réseaux bayésiens.
9. Afin d'évaluer les performances de notre raisonneur, nous avons construit un système d'aide au diagnostic des maladies coronariennes qui se base sur l'ontologie probabiliste en tant que base de connaissances.
10. En collaboration avec l'expert du domaine, nous avons procédé au test et à la validation du diagnostic de notre système, ainsi qu'utilisé des indicateurs de performance pour l'évaluer.

## 2.5 Le Dataset

Le dataset que nous avons choisi est issu d'une étude en cours sur les maladies cardiovasculaires menée auprès des résidents de la ville de Framingham, dans le Massachusetts. Il est disponible sur le site web de Kaggle<sup>5</sup>, sous le nom de « Framingham Heart Study dataset ».

Nous avons sélectionné ce jeu de données pour plusieurs raisons. Tout d'abord, il dispose d'un nombre d'instances assez important, avec 4240 instances, ce qui le distingue des autres ensembles de données disponibles gratuitement.

De plus, ce dataset présente une combinaison intéressante de 15 attributs qui caractérisent les signes et les facteurs de risque des patients, ainsi qu'un attribut cible dans qui indique la prédiction du risque de maladie coronarienne (MC) sur une période de 10 ans.

La présence de valeurs manquantes dans la dataset permet d'appliquer l'algorithme EM pour traiter l'incertitude due à l'incomplétude des données.

Nous présentons dans le tableau 2.1 la description des attributs du dataset :

---

5. <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>

<b>Attribut</b>	<b>Description</b>
Sex	Le sexe du patient.
Age	L'âge du patient en années.
Education	Le niveau d'éducation du patient.
Smoking	Si le patient est fumeur ou non.
CigsPerDay	Le nombre de cigarettes fumées par jour.
BPMeds	La prise ou non de médicaments contre la pression artérielle.
PrevalentStroke	S'il y a déjà eu un antécédent d'accident vasculaire cérébral chez le patient ou non.
PrevalentHyp	Si le patient était hypertendu ou non.
Diabetes	Si le patient est diabétique ou non.
TotChol	Le niveau de cholestérol dans le sang du patient.
SysBP	Le niveau de pression artérielle systolique.
DiaBP	Le niveau de pression artérielle diastolique.
BMI	L'indice de masse corporelle (IMC).
HeartRate	Le rythme cardiaque.
Glucose	Le taux de glucose.
TenYearCHD	Le risque d'avoir une maladie coronarienne sur une période de 10 ans.

TABLE 2.1 – Description du dataset.

## 2.6 Prétraitement des données

Le prétraitement du dataset est essentiel pour garantir la qualité et la pertinence des données utilisées dans notre système, afin d'améliorer la précision des résultats obtenus.

Le processus de prétraitement de l'ensemble des données se déroule en trois étapes :

### 1. Nettoyage des données

Dans cette étape, en collaboration avec l'expert du domaine, dans le but d'améliorer la précision du diagnostic, nous avons effectué deux actions. D'une part, nous avons éliminé les attributs qui n'ont pas d'impact sur le diagnostic de la maladie coronarienne. D'autre part, nous avons ajouté l'attribut "obesity" au dataset.

### 2. Catégorisation

Après avoir nettoyé l'ensemble des données et bénéficié des connaissances de l'expert du domaine, nous avons pu déterminer les intervalles pour chaque attribut numérique. Le tableau 2.2 présente les catégories déterminées à partir des intervalles élaborés.

Attribut	Intervalle	Valeurs
Age	[0-48]	Young
	>48	Old
cigsPerDay	[0-4]	Low
	[5-14]	Normal
	>15	High
totChol	<200	Normal
	[200-239]	Borderline-high
	>=240	High
SysBP	<100	Low
	[100-120[	Optimal
	[120-130[	Normal
	[130-139]	High
	[140-159]	Mild-hypertension
	[160-179]	Moderate-hypertension
	>=180	Severe-hypertension
diaBP	<60	Low
	[60-80[	Optimal
	[80-85[	Normal
	[85-89]	High
	[90-99]	Mild-hypertension
	[100-109]	Moderate-hypertension
	>=110	Severe-hypertension
BMI	<18,5	Under-wieght
	[18,5-24 ,99]	Normal
	[25-29 ,99]	Over-weight
	[30-34,99]	Moderate-obesity
	[35-39,99]	Severe-obesity
	>=40	Moribide-obesity
heartRate	<60	Low
	[60-110]	Medium
	>110	High
Glucose	<70	Low
	[70-100]	Normal
	[100-125]	High
	>125	Extremely-high

TABLE 2.2 – Catégorisation des données.

### 3. Équilibrage du dataset

Il est essentiel d'avoir un ensemble de données équilibré pour garantir une performance optimale de l'apprentissage automatique. Dans notre cas, l'ensemble de données était dés-équilibré, avec une proportion de 15,18% pour la classe "True" et de 84,81% pour la classe "False". Pour remédier à cela, nous avons utilisé la technique de sous-échantillonnage en appliquant la méthode du "seuil de dureté des instances" (Instance Hardness Threshold). C'est une méthode qui identifie les instances "dures" de la classe minoritaire (c'est-à-dire

les plus difficiles à classer) et les exclut de l'ensemble d'entraînement. Savoir quelles instances sont souvent mal classées ou difficiles à classer correctement peut améliorer le processus d'apprentissage de l'algorithme en réduisant le bruit dans les données d'entraînement [7]. Cette méthode utilise le théorème de Bayes pour calculer la probabilité de classer correctement les données en fonction des caractéristiques d'entrée et des étiquettes [23]. Après avoir appliqué cette méthode nous avons obtenu un dataset équilibré avec un pourcentage de 52,60% pour la classe du diagnostic positif et 47,40% pour la classe du diagnostic négatif.

## 2.7 Construction de l'ontologie classique

Après avoir effectué l'étape précédente, nous avons créé notre ontologie classique en nous basant sur les attributs de notre dataset prétraité. Nous avons spécifié les concepts, leurs attributs et leurs relations afin de représenter formellement les connaissances implicites de l'expert.

Le tableau 2.3 représente les attributs (data properties) des concepts de l'ontologie :

Attributs	Domaine	Range
Age_value	Patient	Int
Sex_value	Patient	String
Glucose_value	Glucose	Float
TotalCholesterol_value	TotalCholesterol	Float
BMI_value	BMI	Float
DiastolicBP_value	DiastolicBP	Float
HeartRate_value	HeartRate	Int
SystolicBP_value	SystolicBP	Float
IsDiabetic_state	IsDiabetic	Boolean
IsHypertensive_state	IsHypertensive	Boolean
Obesity_state	Obesity	Boolean
IsTouchedbyStroke_state	IsTouchedbyStroke	Boolean
CigsPerDay_value	CigsPerDay	Int
IsSmoking_state	IsSmoking	Boolean
IsSuspectedToCHD_state	IsSuspectedToCHD	Boolean

TABLE 2.3 – Les attributs des concepts.

Le tableau 2.4 représente les relations (object properties) entre les concepts :

Relations	Domaine	Range
Has_Risk_Factor	Patient	Risk_Factor
Has_Sign	Patient	Sign
Has_Diagnostic	Patient	IsSuspectedToCHD

TABLE 2.4 – Les relations entre les concepts.



Les taxonomies représentent les hiérarchiques des classes. Nous présenterons ci-dessous la taxonomie des classes de l'ontologie.

### 1. Taxonomie de l'ontologie

La figure 2.2 représente la taxonomie de l'ontologie.

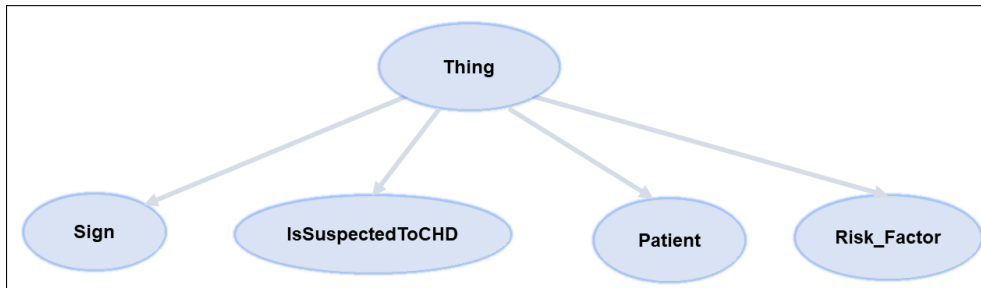


FIGURE 2.2 – Taxonomie de Thing.

### 2. Taxonomie de Risk-Factor

La figure 2.3 représente la taxonomie de la classe Risk-Factor.

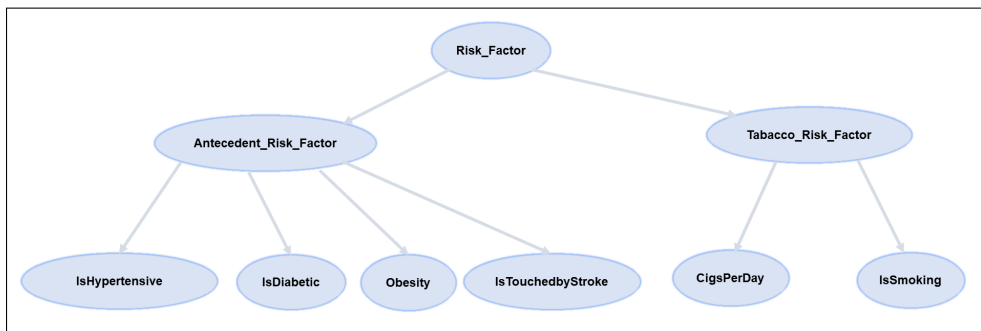


FIGURE 2.3 – Taxonomie de Risk-Factor.

### 3. Taxonomie de Sign

La figure 2.4 représente la taxonomie de la classe Sign.

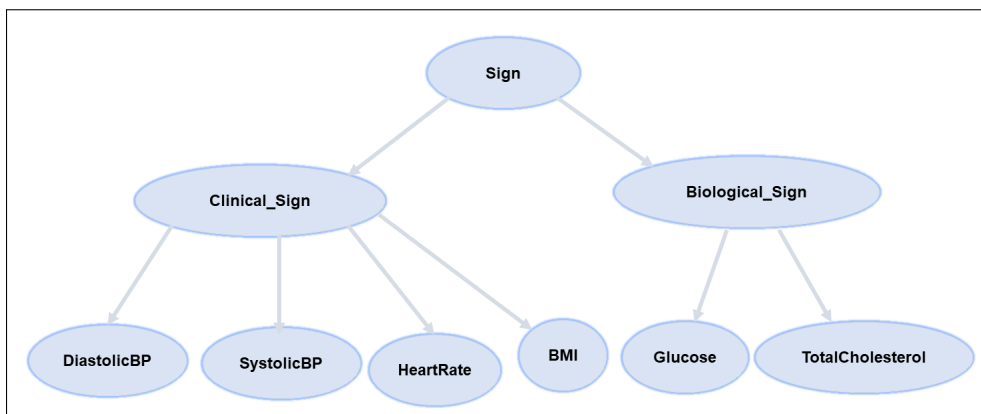


FIGURE 2.4 – Taxonomie de Sign.

Après avoir spécifié les concepts, les propriétés et les relations de notre ontologie classique, nous avons obtenu sa structure globale représenté dans la figure 2.5.

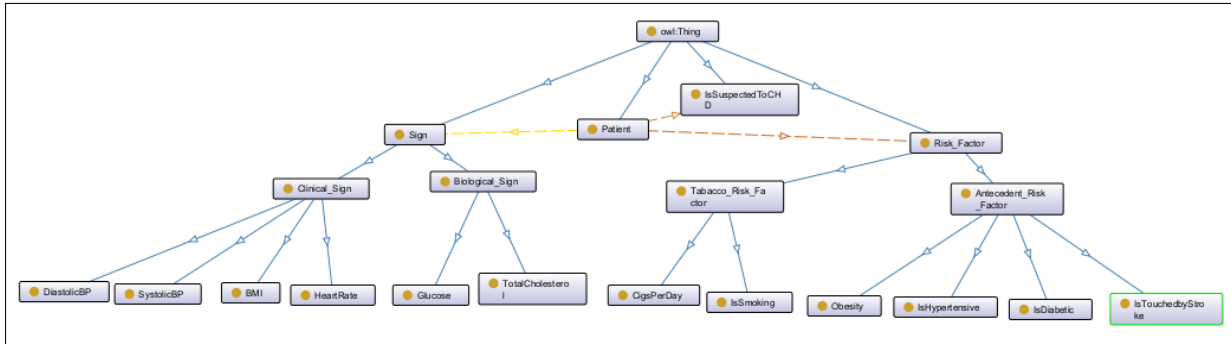


FIGURE 2.5 – La structure de l'ontologie classique.

La figure 2.6 illustre la hiérarchie des classes, ainsi que les propriétés d'objets et les propriétés de données de l'ontologie classique.

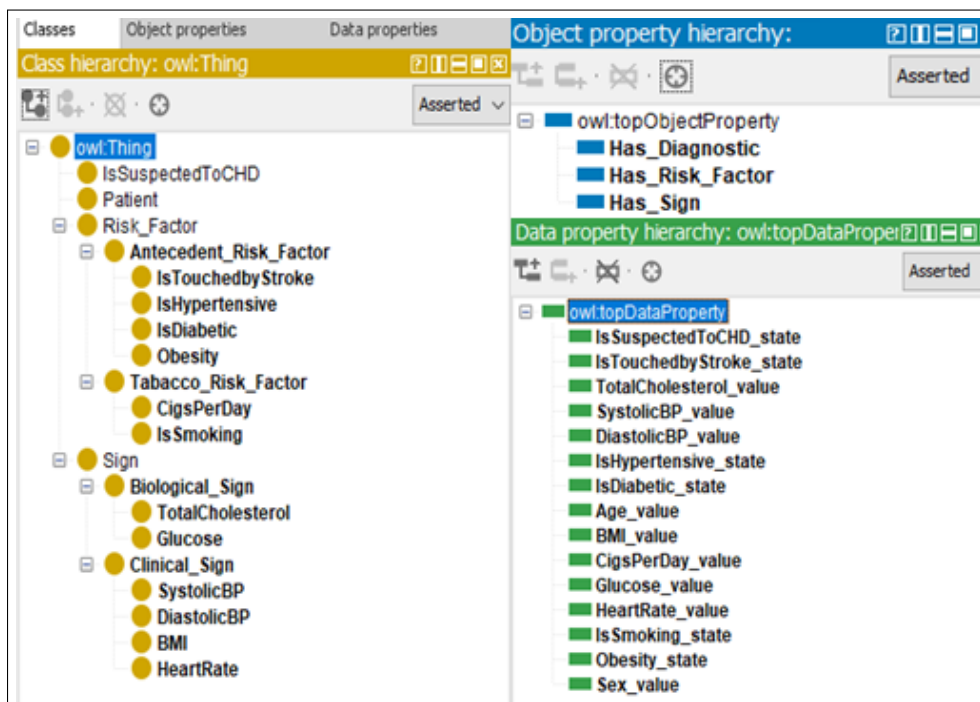


FIGURE 2.6 – Les classes, les propriétés d'objets et les propriétés de données de l'ontologie classique.

## 2.8 Construction de réseau bayésien

Afin de mettre en place un mécanisme de raisonnement probabiliste, nous avons procédé à la construction de notre réseau bayésien, qui nous permet de réaliser des inférences probabilistes. Le processus de conception s'est déroulé en deux phases distinctes :

## 2.8.1 Apprentissage de la structure

La première étape consiste à apprendre la structure du réseau bayésien. Au cours de cette étape, en nous appuyant sur l'expertise du cardiologue, nous avons déterminé les nœuds du réseau bayésien ainsi que les liens de dépendance entre ces nœuds. La figure 2.7 illustre la structure de notre réseau bayésien.

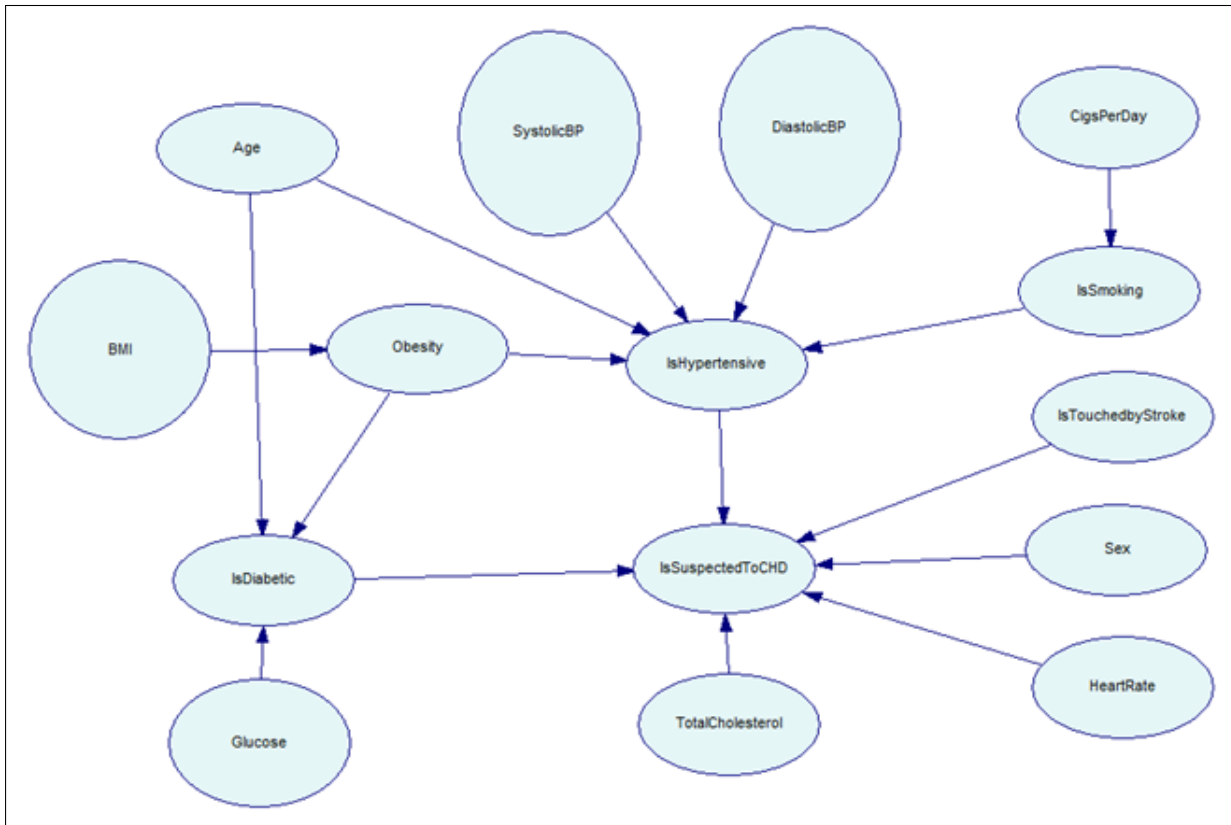


FIGURE 2.7 – La structure du réseau bayésien.

## 2.8.2 Apprentissage des paramètres

Dans le cadre de la deuxième étape, nous nous sommes concentrés sur l'apprentissage des paramètres. Une fois que la structure de notre réseau bayésien était établie, nous avons procédé à l'estimation des tables de probabilités conditionnelles pour chaque nœud. Cette étape cruciale a été réalisée en utilisant l'algorithme EM. La figure 2.8 présente notre réseau bayésien après avoir effectué cet apprentissage des paramètres.

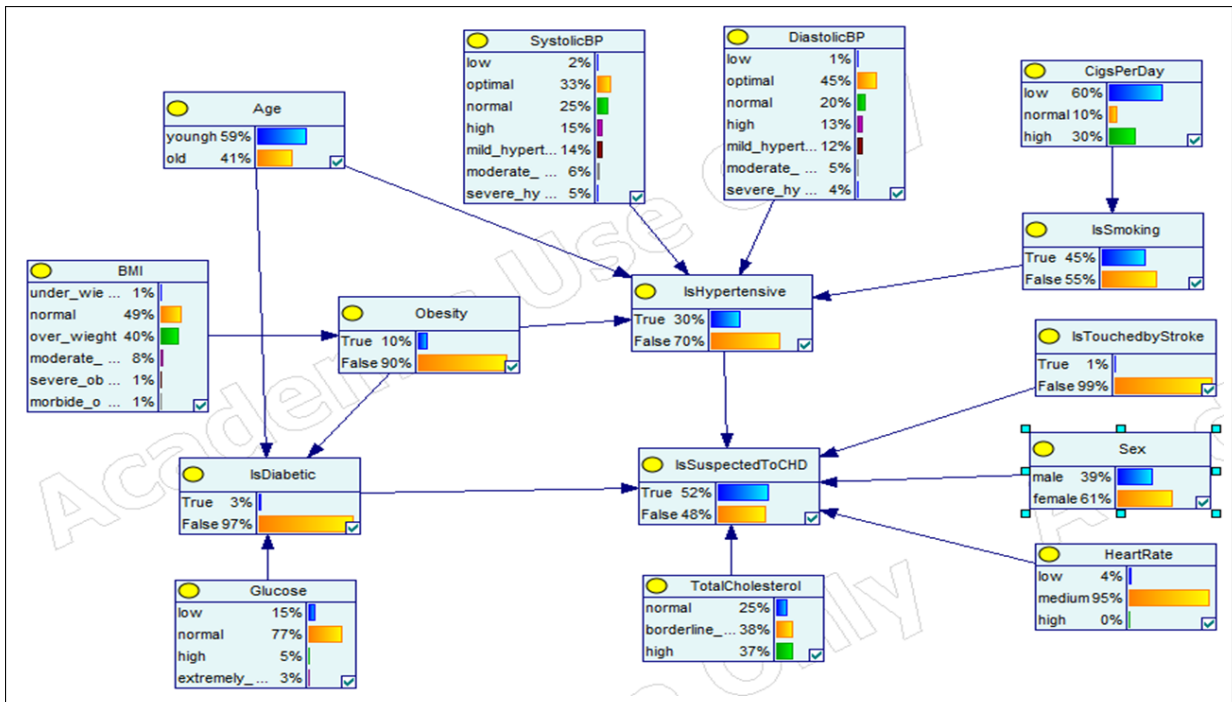


FIGURE 2.8 – Le réseau bayésien après l'apprentissage des paramètres.

## 2.9 Méta-ontologie du réseau bayésien

La construction de la méta-ontologie pour le réseau bayésien vise à représenter de manière formelle les connaissances probabilistes. Cette méta-ontologie permet de définir les combinaisons nécessaires pour l'instanciation d'un réseau bayésien standard. Grâce à cette méta-ontologie, il devient possible d'exploiter l'incertitude.

La structure globale de la méta-ontologie est illustrée dans la figure 2.9.

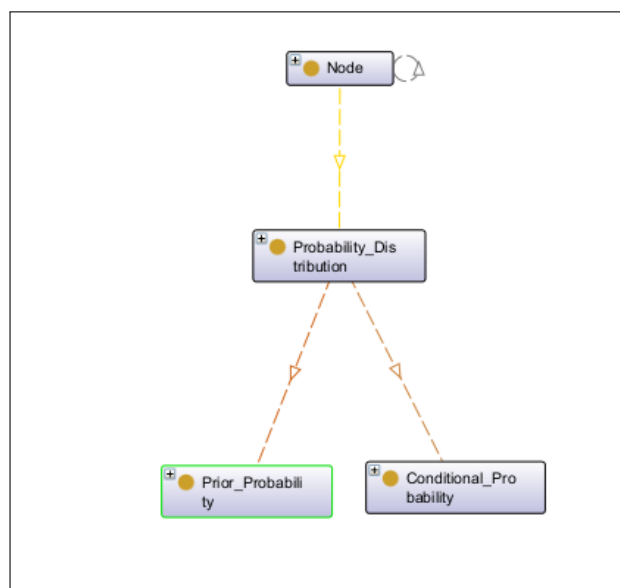


FIGURE 2.9 – la structure de méta-ontologie.

La figure 2.10 représente la hiérarchie des classes, ainsi que les propriétés d'objets et les propriétés de données de méta-ontologie.

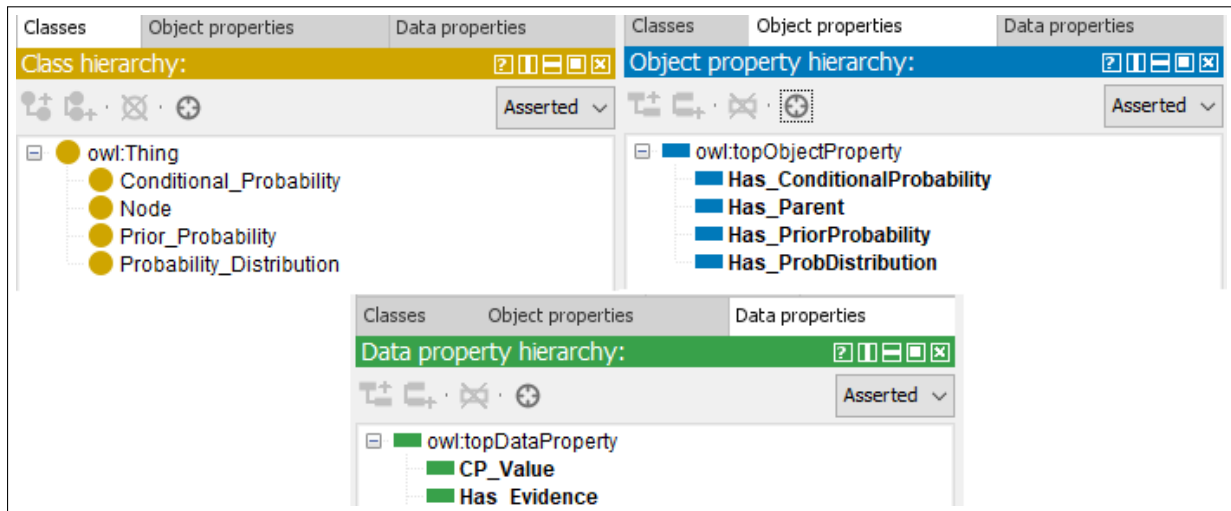


FIGURE 2.10 – Les classes, les propriétés d'objets et les propriétés de données de la méta-ontologie.

## 2.10 Construction de l'ontologie probabiliste

Cette étape vise à incorporer les connaissances probabilistes dans notre ontologie classique. Afin de réaliser cela, nous avons développé un algorithme qui intègre l'ontologie classique avec notre méta-ontologie du réseau bayésien. L'algorithme de l'intégration se déroule comme suite :

1. Lire (*Ontologie\_classique*).
2. Lire(*Réseau\_bayésien*).
3. Créer un concept nommé "Probabilistic".
4. Attribuer à ce concept les propriétés du concept "Node" et leurs co-domaines représentés dans la méta-ontologie.
5. Identifier les concepts probabilistes de l'*Ontologie\_classique* et les placer comme sous-concepts du concept "Probabilistic".
6. Ajouter les évidences pour chaque instance des concepts probabilistes.
7. Utiliser le *Réseau\_bayésien* pour attribuer des valeurs de probabilité à ces instances.
8. Sauvegarder le résultat sous le nom d'*Ontologie\_probabiliste*.
9. Retourner l'*Ontologie\_probabiliste*.

L'exécution de l'algorithme précédent a abouti à notre ontologie probabiliste. Nous avons obtenu les classes, les propriétés de données et les propriétés d'objets présentés dans la figure 2.11.

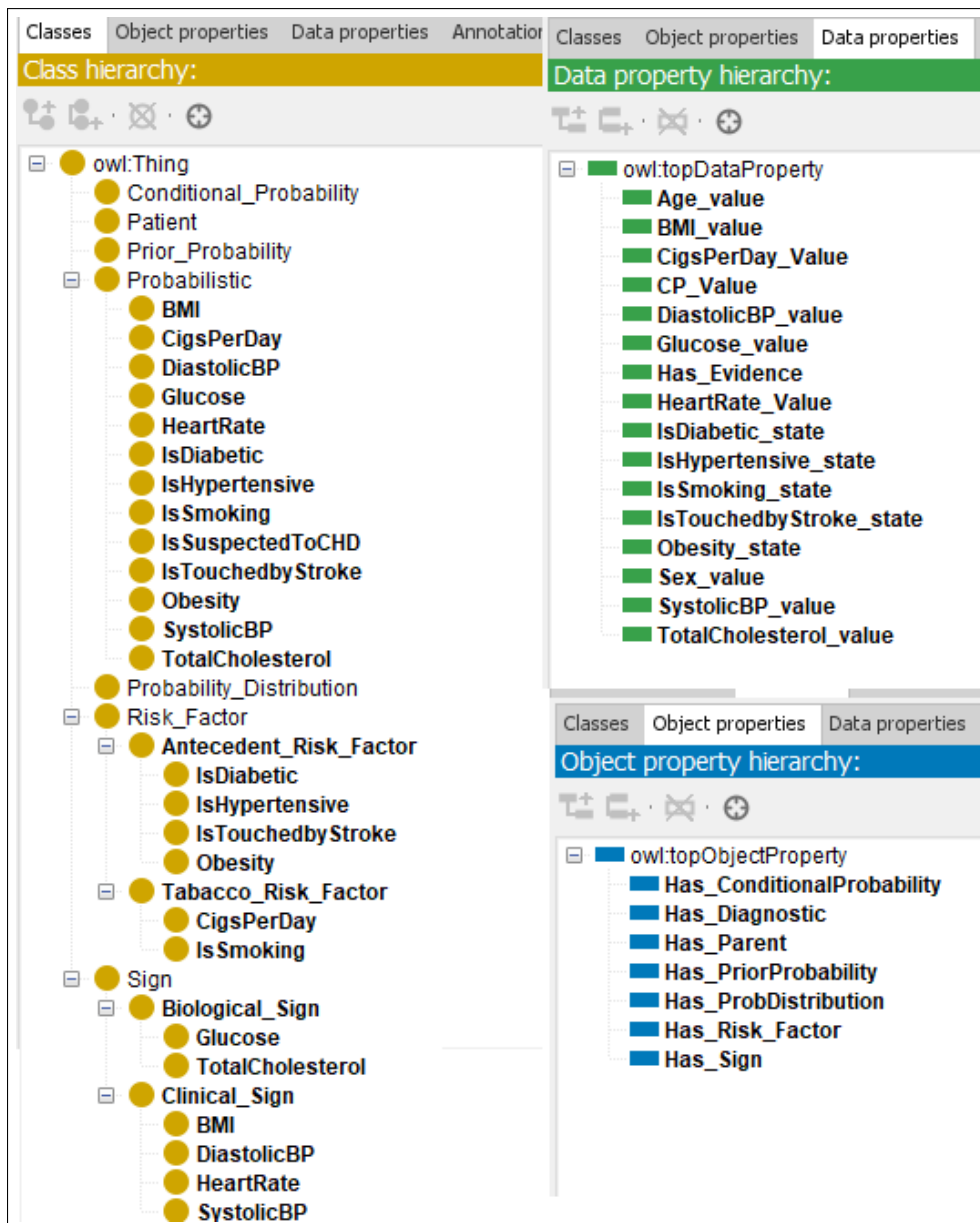


FIGURE 2.11 – Les classes, les object property et les data property de l'ontologie finale.

Afin d'illustrer la capture des connaissances médicales dans notre ontologie probabiliste finale, nous avons procédé à une instantiation d'exemple représentant les connaissances concernant un patient nommé "Patient1". Ce patient est diabétique, de sexe masculin, âgé de 68 ans, présentant un niveau élevé de glucose et ayant reçu un diagnostic négatif de maladie coronarienne sur une période de dix ans.

Le code OWL correspondant à notre exemple est présenté dans les figures suivantes : La figure 2.12 représente le code OWL de la classe "Patient".

```
<owl:Class rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Patient"/>
```

FIGURE 2.12 – Le code en OWL de la classe "Patient".

La figure 2.13 représente le code OWL de l'instance "Patient1" de la classe "Patient"

```
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Patient1">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Patient"/>
  <untitled-ontology-86:Has_Diagnostic rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#False_CHD"/>
  <untitled-ontology-86:Has_Risk_Factor rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#True_Diabetes"/>
  <untitled-ontology-86:Has_Sign rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#High_Glucose"/>
  <untitled-ontology-86:Age_value rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">68</untitled-ontology-86:Age_value>
  <untitled-ontology-86:Sex_value Male</untitled-ontology-86:Sex_value>
</owl:NamedIndividual>
```

FIGURE 2.13 – Le code en OWL de l'instance "Patient1".

La figure 2.14 représente le code OWL de classe probabiliste "Glucose"

```
<owl:Class rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Glucose">
  <rdf:s:subClassOf rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Biological_Sign"/>
  <rdf:s:subClassOf rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probabilistic"/>
</owl:Class>
```

FIGURE 2.14 – Le code en OWL de la classe "Glucose".

La figure 2.15 représente le code OWL de l'instance "High\_Glucose" de la classe "Glucose" et sa valeur de probabilité.

```

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#High_Glucose">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Glucose"/>
  <untitled-ontology-86:Has_ProbDistribution rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution_High_Glucose"/>
  <untitled-ontology-86:Glucose_value rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal" 110 /untitled-ontology-86:Glucose_value>
  <untitled-ontology-86:Has_Evidence High /untitled-ontology-86:Has_Evidence>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution_High_Glucose">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution"/>
  <untitled-ontology-86:Has_PriorProbability rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Prior_Probability_High_Glucose"/>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Prior_Probability_High_Glucose">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Prior_Probability"/>
  <untitled-ontology-86:PriorProbability_Value rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal" 0.049756295694557264 /untitled-ontology-86:PriorProbability_Value>
</owl:NamedIndividual>

```

FIGURE 2.15 – Le code en OWL de l'instance "High\_Glucose".

La figure 2.16 représente le code OWL de classe probabiliste "IsDiabetic"

```

<owl:Class rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#IsDiabetic">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Antecedent_Risk_Factor"/>
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probabilistic"/>
</owl:Class>

```

FIGURE 2.16 – Le code en OWL de la classe "IsDiabetic".

La figure 2.17 représente le code OWL de l'instance "True\_Diabetes" de la classe "IsDiabetic" et sa table de probabilité conditionnelle.



```

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#True_Diabetes">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#IsDiabetic"/>
  <untitled-ontology-86:Has_Parent rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#High_Glucose"/>
  <untitled-ontology-86:Has_ProbDistribution rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution_True_Diabetes"/>
  <untitled-ontology-86:Has_Evidence rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean" true"/>
  <untitled-ontology-86:IsDiabetic_state rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean" true"/>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution_True_Diabetes">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probability_Distribution"/>
  <untitled-ontology-86:Has_ConditionalProbability rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Conditional_Probability_True_Diabetes"/>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Conditional_Probability_True_Diabetes">
  <rdf:type rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Conditional_Probability"/>
  <untitled-ontology-86:CP_Value>0.1 0.9 0.01293103448275862 0.9870689655172413 0.04545454545454546 0.9545454545454546 0.008474576271186441
  0.9915254237288136 0.01351351351351351 0.9864864864864865 0.0009276437847866419 0.9990723562152134 0.01041666666666667 0.9895833333333333
  0.004629629629629 0.9953703703703703 0.1666666666666667 0.8333333333333333 0.02173913043478261 0.9782608695652174 0.2083333333333333
  0.7916666666666666 0.1666666666666667 0.8333333333333333 0.875 0.125 0.8333333333333333 0.1666666666666667 0.8125 0.1875 0.6851851851851851
  0.3148148148148148</untitled-ontology-86:CP_Value>
</owl:NamedIndividual>

```

FIGURE 2.17 – Le code en OWL de l’instance "True\_Diabetes".

La figure 2.18 représente le code OWL de la classe "IsSuspectedToCHD".

```

<owl:Class rdf:about="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#IsSuspectedToCHD">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/yasmine/ontologies/2023/4/untitled-ontology-86#Probabilistic"/>
</owl:Class>

```

FIGURE 2.18 – Le code en OWL de la classe "IsSuspectedToCHD".

La figure 2.19 représente le code OWL de l’instance False\_CHD de la classe "IsSuspectedToCHD" et sa table de probabilité conditionnelle.



<b>Tâche</b>	<b>Input</b>	<b>Traitement</b>	<b>Output</b>
<b>Subsomption probabiliste</b>	Deux concept $C1$ et $C2$	Vérifiée si le concept $C1$ est plus général que le concept $C2$	Une valeur booléenne
<b>Satisfiabilité probabiliste</b>	Un concept $C$	Vérifiée si le concept $C$ possède au moins un individu.	Une valeur booléenne.
<b>Récupération d'instances</b>	Un concept $C$ .	Récupérer tous les individus du concept $C$ donnée.	Une liste de toutes les instances.
<b>Instanciation probabiliste</b>	Un concept $C$ et un individu $I$	Calculer les probabilités pour chaque état de la classe $C$ que l'individu $I$ appartient à cet état.	Une liste des couples $\langle Etat1, P(I \text{ ait } Etat1) \rangle$ , $\langle Etat2, P(I \text{ ait } Etat2) \rangle$ , ..., $\langle Etat n, P(I \text{ ait } Etat n) \rangle$ . où : <ul style="list-style-type: none"> <li>• <b><i>Etat n</i></b> : est un état de la classe <math>C</math> de type chaîne de caractères.</li> <li>• <b><math>P(I \text{ ait } Etat n) \in [0, 1]</math></b> : est la probabilité que l'individu <math>I</math> appartient à cet état.</li> </ul>
<b>L'incertitude existentielle</b>	Un individu $I$ .	Chercher l'individu $I$ et retourner une valeur booléenne.	Une valeur booléenne.
<b>L'incertitude liée aux propriétés d'objets</b>	La propriété d'objet $O$ , l'individu $I1$ et l'individu $I2$ .	Calculer la probabilité que les deux individus sont liés avec la relation d'objet $O$ .	La valeur du $P(\exists O(I1, I2))$ , avec $P \in [0, 1]$ .
<b>L'incertitude liée aux propriétés de données</b>	L'individu $I$ , la propriété de données $D$ et la valeur $V$ .	Calculer la probabilité que pour l'individu $I$ la propriété de données $D$ ait la valeur $V$ .	Un couple $\langle D(I, V), \exists P(D(I, V)) \rangle$ , avec $P \in [0, 1]$ .
<b>L'incertitude liée aux ensembles des propriétés</b>	L'individu $I$ et les listes des propriétés $E1$ et $E2$ .	Calculer la probabilité de ensemble de propriétés $E1$ sachent un autre ensemble de propriétés $E2$ .	Une liste des triples $\langle Pr_i, Range_i, p \rangle$ , avec $P \in [0, 1]$ , et $Pr_i \in E1$ , et $Range_i \in Pr_i$ .

TABLE 2.5 – Les tâches de raisonnement probabiliste des connaissances ontologiques.

## 2.12 Conclusion

Dans ce chapitre nous avons expliqué l'architecture globale de notre système, commençant par le traitement fait sur le data-set, passons aux étapes de création de notre ontologie et réseau bayésienne, puis l'obtention de notre ontologie probabiliste. Ensuite on a aussi présenté la modélisation des tâches du raisonneur probabiliste.

Au chapitre suivant nous allons implémenter la solution proposée de ce chapitre, tester notre raisonneur et valider les résultats trouvés.

## Chapitre 3

---

# IMPLEMENTATION ET TEST DU SYSTEME

---

### 3.1 Introduction

Dans le chapitre précédent, nous avons présenté les différentes étapes de construction de l'ontologie probabiliste et la conception du raisonneur.

Dans ce chapitre, nous allons décrire l'implémentation de notre raisonneur probabiliste. Nous incluons des captures d'écran des différentes interfaces du raisonneur, à la fin du chapitre, nous procéderons à un test pour valider notre application.

### 3.2 Langage de Programmation

**Python :** Est un langage de programmation open source connu par sa simplicité et sa facilité d'apprentissage, c'est un langage de programmation interprété, il supporte différents paradigmes de programmation, notamment la programmation procédurale et fonctionnelle, en plus de la programmation orientée objet. Il dispose d'une grande variété de bibliothèques et de modules prédéfinis pour un développement facile. Enfin, Python est portable, c'est-à-dire qu'il fonctionne sur de nombreuses variantes d'Unix, y compris Linux et macOS, et sur Windows.<sup>1</sup>

Dans notre projet, nous l'avons utilisé à la fois pour le développement de notre raisonneur et pour l'implémentation de notre application.

### 3.3 Outils de développement

Nous allons vous présenter la liste des outils de développement qui sont utilisés pour notre système :

---

1. <https://docs.python.org/3/faq/general.html#generalinformation>

1. **Protégé** : Est une plate-forme gratuite et open source produit par l'Université de Stanford. Qui permet la création, l'importation, la modification, et l'enregistrement des ontologies. Elle prend en charge différents formats de représentation des ontologies, notamment OWL et RDF. <sup>2</sup>
2. **Génie smile** : GeNIe Modeler est une interface utilisateur graphique (GUI) pour SMILE Engine, permet la construction et l'apprentissage de modèles interactifs. Il a été minutieusement testé sur le terrain depuis 1998, Il est écrit pour l'environnement Windows mais peut également être utilisé sur macOS et Linux. <sup>3</sup>
3. **Visual Studio Code** : est un éditeur de code source gratuit et open source développée par Microsoft, VS Code est disponible sur pour Windows, macOS et Linux, et il prend en charge de nombreux langages de programmation populaires tels que JavaScript, Python, C++, Java et bien d'autres. <sup>4</sup>

## 3.4 API Utilisés

1. **Owlread2y** : est un package permettant de manipuler les ontologies OWL 2.0 en Python. Il peut charger, modifier, sauvegarder des ontologies, et il supporte le raisonnement via Hermit (inclus). Owlready permet un accès transparent aux ontologies OWL. <sup>5</sup>
2. **PyQt5** : Un ensemble de liaisons Python pour les bibliothèques et les fonctionnalités de Qt, ce qui permet aux développeurs Python de créer des applications graphiques riches et interactives. <sup>6</sup>
3. **Imbalanced-learn** : est une bibliothèque open source, sous licence MIT, qui s'appuie sur scikit-learn et fournit des outils pour traiter la classification avec des classes déséquilibrées. <sup>7</sup>
4. **PySide2** : est un projet de liaisons Python pour Qt, offrant un accès complet au framework Qt 5.12+ ainsi qu'à des outils de génération de liaisons rapides pour toutes les bibliothèques. C++ <sup>8</sup>

## 3.5 Les interfaces graphiques

Dans cette section, nous allons présenter quelques interfaces de notre application.

---

2. <https://protege.stanford.edu/>

3. <https://www.bayesfusion.com/genie/>

4. <https://code.visualstudio.com/docs>

5. <https://owlready2.readthedocs.io/en/latest/intro.html>

6. <https://www.riverbankcomputing.com/static/Docs/PyQt5/introduction.html>

7. <https://imbalanced-learn.org/dev/index.html>

8. <https://pypi.org/project/PySide2/>

## 1. L'interface d'accueil

La figure 3.1 représente la première interface de notre système, elle contient quatre boutons. Le premier est le bouton 'Open ontology' qui permet à l'utilisateur d'ouvrir le fichier de l'ontologie classique. Le deuxième est le bouton 'Open bayesian network' qui permet à l'utilisateur d'ouvrir le fichier du réseau bayésien. Le troisième est 'Raisonneur' qui permet d'accéder à une interface contenant toutes les tâches de notre raisonneur. Le dernier bouton 'Exit' permet de fermer le système.



FIGURE 3.1 – Interface d'accueil.

## 2. L'interface de raisonneur

Sur l'interface du raisonneur, vous trouverez dix boutons. Un bouton permet de quitter l'application ('Exit') et un autre permet de revenir en arrière. Les huit autres boutons représentent les différentes fonctionnalités de notre raisonneur, à savoir :

- Subsumption probabiliste.
- Satisfiabilité probabiliste.
- Récupération d'instances.
- Incertitude existentielle.
- Incertitude de type individuel.



- Incertitude des propriétés de données.
- Incertitude des propriétés d'objet.
- Incertitude liée aux ensembles des propriétés.

La figure 3.2 illustre la représentation visuelle de l'interface du raisonneur :

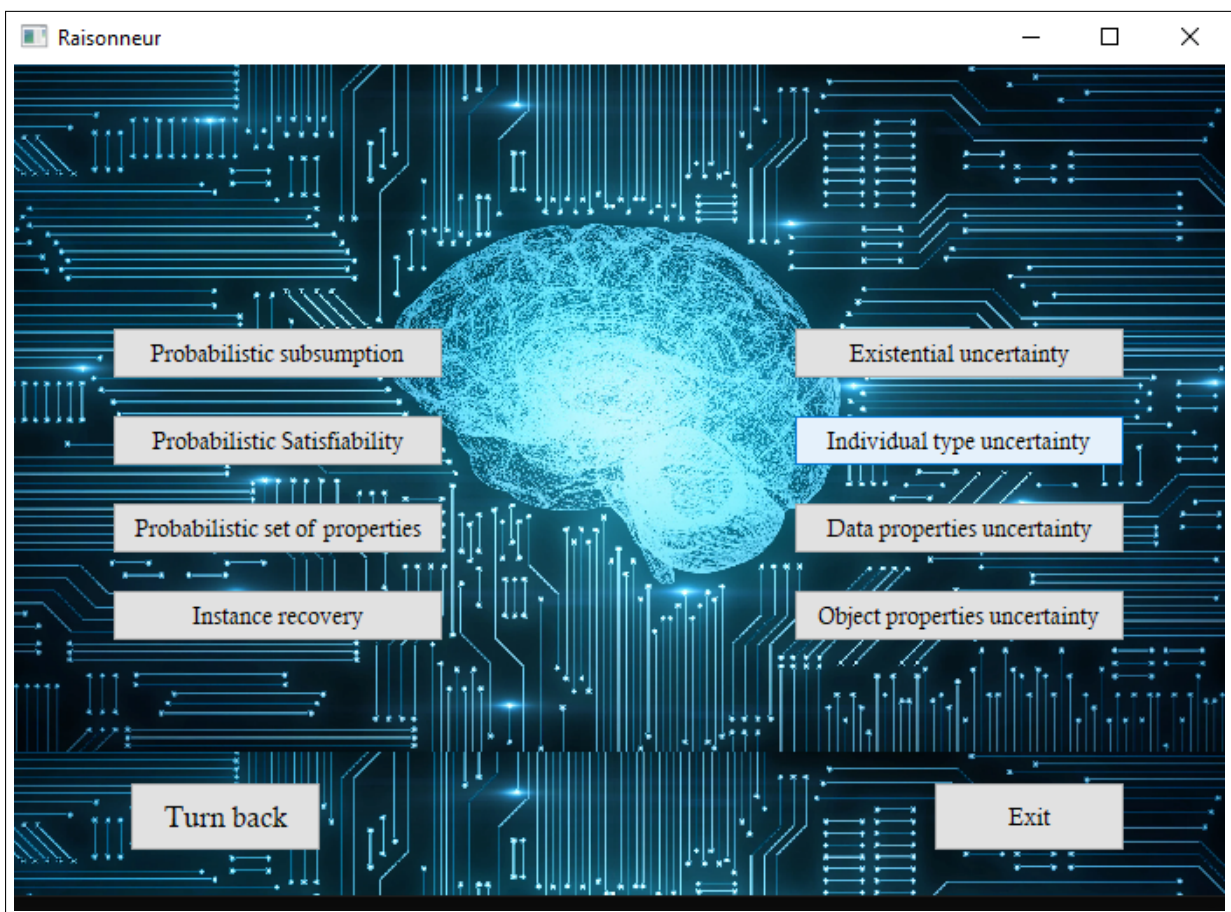


FIGURE 3.2 – Interface de raisonneur.

### 3. L'interface de la fonctionnalité "Subsumption probabiliste"

L'interface a été conçue dans le but de faciliter les tests de subsumption. Vous y trouverez deux listes déroulantes vous permettant de sélectionner les concepts à raisonner. Une fois les concepts choisis, il vous suffit de cliquer sur le bouton "Start reasoning" pour lancer le processus de raisonnement. Si vous souhaitez revenir en arrière, un bouton "Turn back" est disponible pour vous faciliter la navigation. La figure 3.3 illustre la représentation visuelle de l'interface "Subsumption probabiliste".



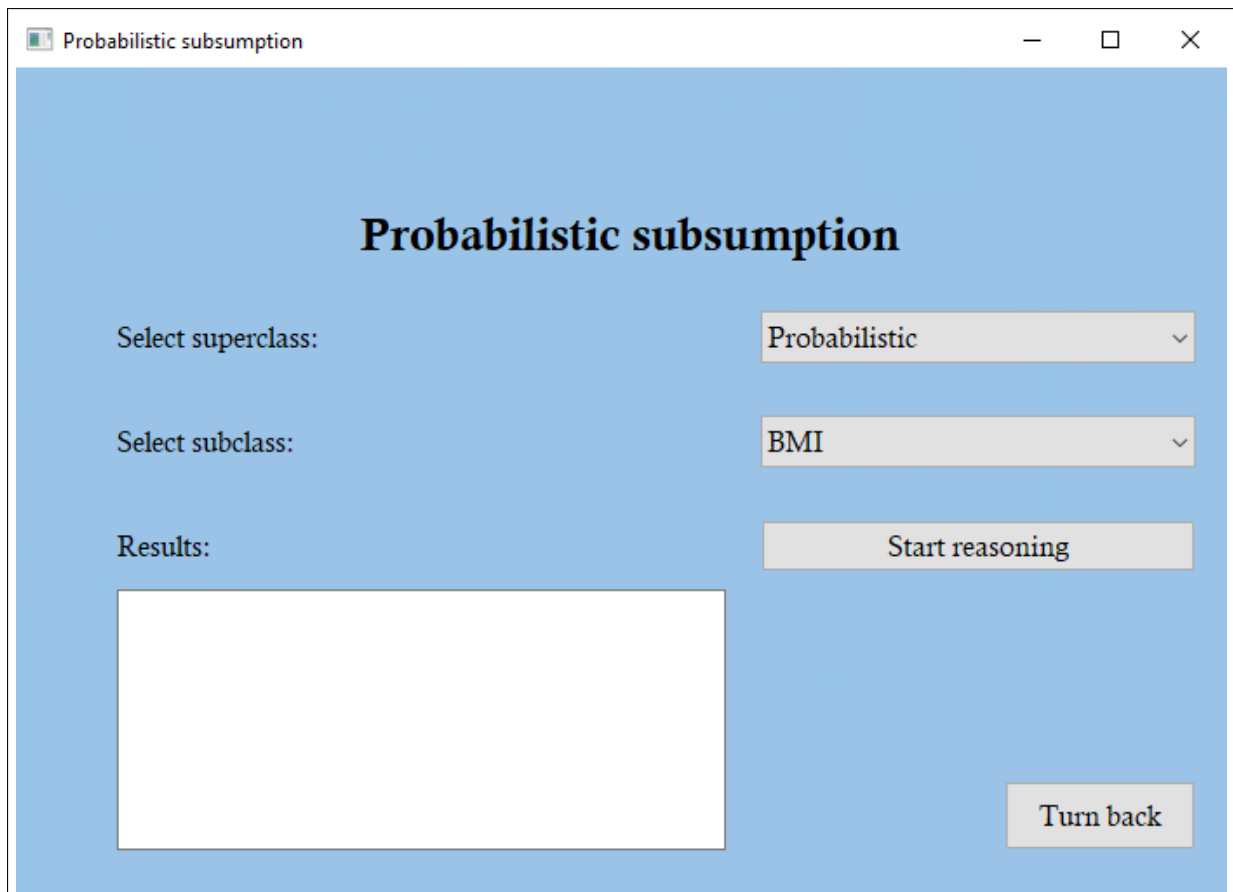


FIGURE 3.3 – Interface de la fonctionnalité "subsumption probabiliste".

#### 4. L'interface de la fonctionnalité "Satisfiabilité probabiliste"

L'interface a été spécialement conçue pour simplifier les tests de Satisfiabilité des concepts. Elle comprend une liste déroulante qui vous permet de sélectionner facilement le concept que vous souhaitez tester. Une fois le concept choisi, il vous suffit de cliquer sur le bouton "Start reasoning" pour lancer le processus. Si vous souhaitez revenir en arrière, vous avez également la possibilité d'utiliser le bouton "Turn back" pour faciliter votre navigation. La figure 3.4 illustre la représentation visuelle de l'interface "Satisfiabilité probabiliste".

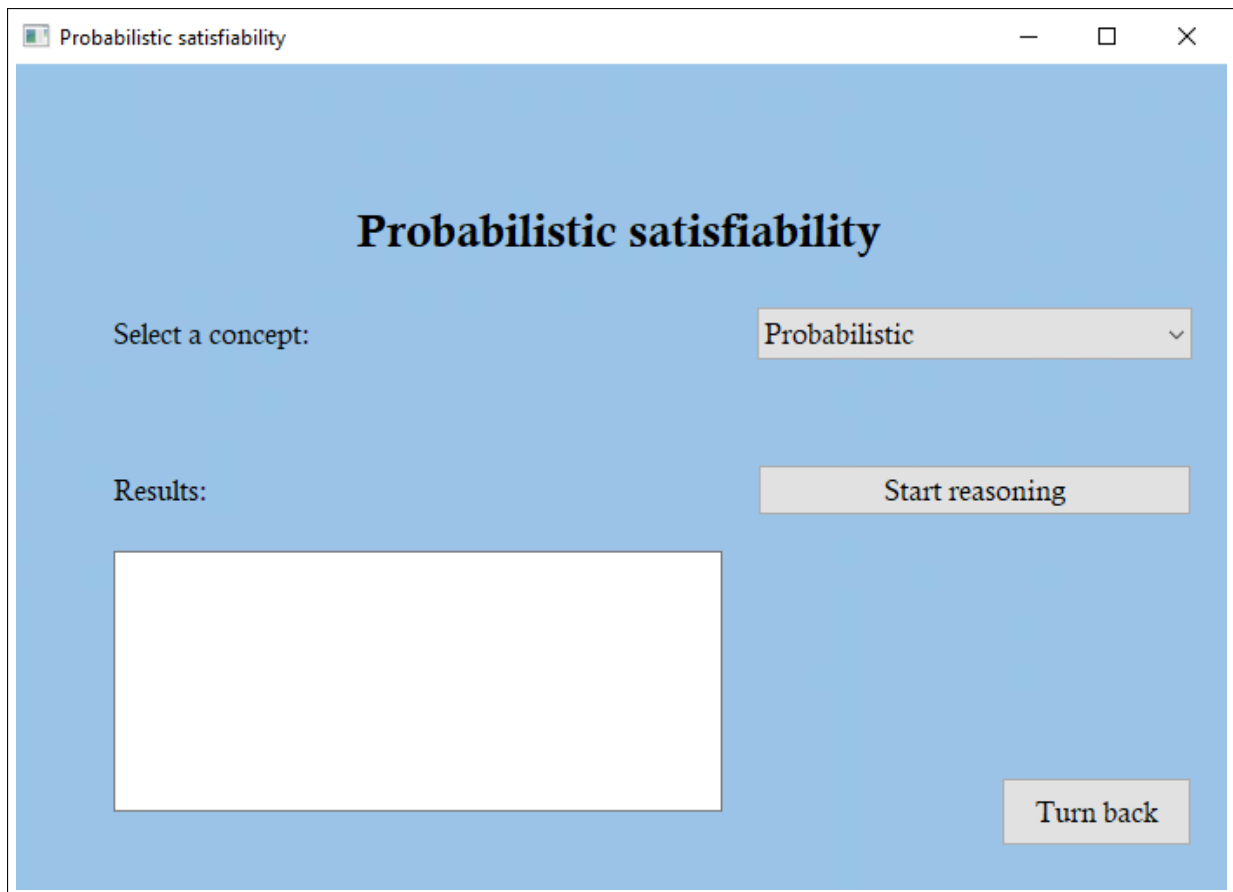


FIGURE 3.4 – Interface de la fonctionnalité "Satisfiabilité probabiliste".

##### 5. L'interface de la fonctionnalité "Récupération d'instance"

L'interface a été spécifiquement conçue pour simplifier la récupération des instances d'un concept donné. Vous y trouverez une liste déroulante vous permettant de sélectionner facilement les concepts. Une fois les concepts choisis, il vous suffit de cliquer sur le bouton "Start reasoning" pour lancer le processus. Si vous souhaitez revenir en arrière, un bouton "Turn back" est disponible pour faciliter votre navigation. La figure 3.5 illustre la représentation visuelle de l'interface "Récupération d'instance".

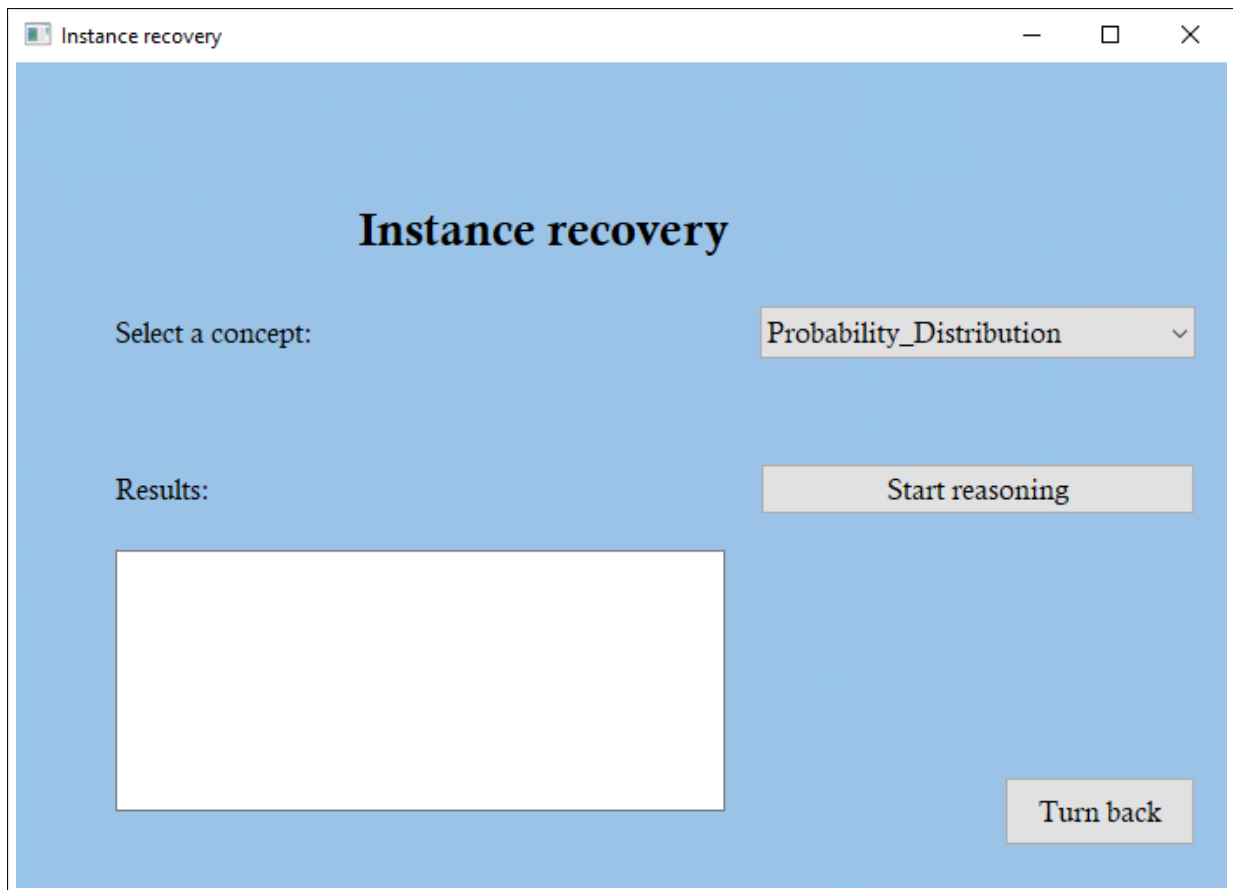


FIGURE 3.5 – Interface de la fonctionnalité "Récupération d'instance".

#### 6. L'interface de la fonctionnalité "Incertitude existentielle"

L'interface a été spécifiquement conçue pour simplifier la recherche d'une instance donnée. Vous y trouverez une zone où vous pouvez saisir le nom de l'instance que vous souhaitez rechercher. Une fois la saisie terminée, il vous suffit de cliquer sur le bouton "Start reasoning" pour lancer le processus. Si vous souhaitez revenir en arrière, un bouton "Turn back" est disponible pour faciliter votre navigation. La figure 3.6 illustre la représentation visuelle de l'interface "Incertitude existentielle".

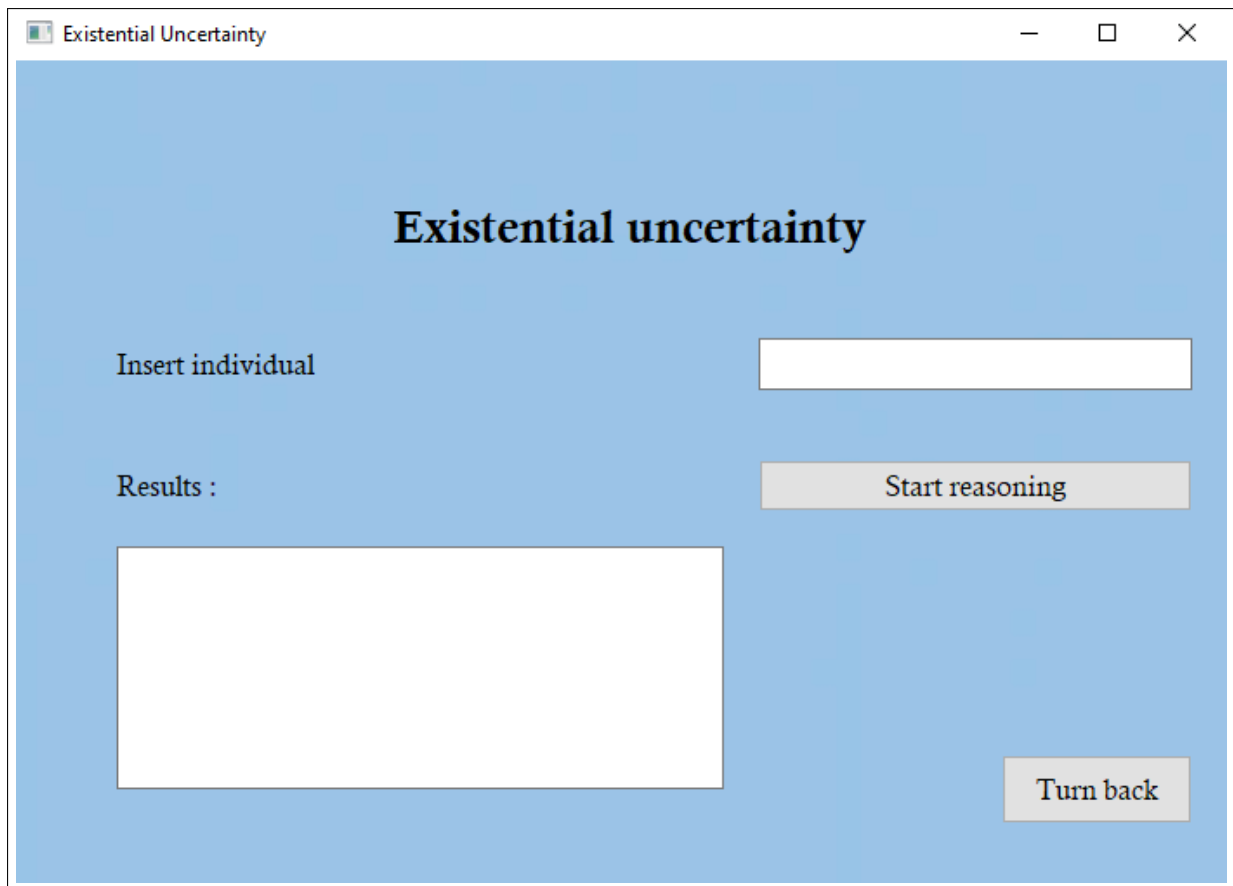


FIGURE 3.6 – Interface de la fonctionnalité "Incertitude existentielle".

#### 7. L'interface de la fonctionnalité "Incertitude de type d'individu"

L'interface a été spécifiquement conçue pour simplifier la détermination de la probabilité d'appartenance d'un individu à une classe. Vous y trouverez deux listes déroulantes, l'une pour les individus et l'autre pour les concepts probabilistes. Dès qu'un individu est sélectionné, ses évidences apparaissent. Ensuite, vous pouvez cliquer sur le bouton "Start reasoning" pour lancer le processus de raisonnement, et le résultat s'affiche dans des barres de progression circulaires. Si vous souhaitez revenir en arrière, un bouton "Turn back" est disponible pour faciliter votre navigation. La figure 3.7 illustre la représentation visuelle de l'interface "Incertitude de type d'individu".

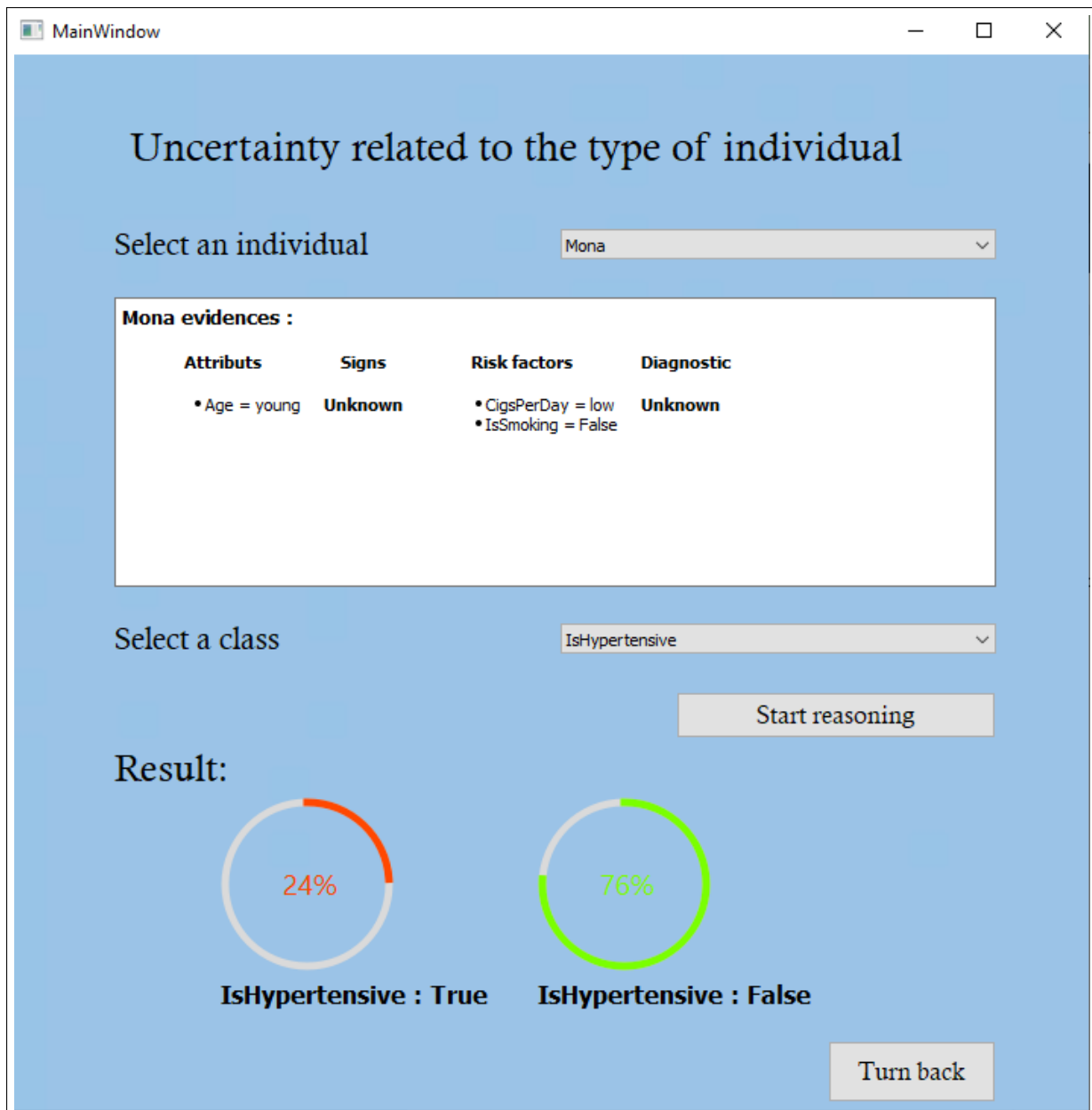


FIGURE 3.7 – Interface de la fonctionnalité "Incertitude de type d'individu".

## 3.6 Validation du système

Le processus de validation de notre système s'est déroulé en deux phases :

- La validation en consultant l'expert du domaine.
- La validation par les mesures de performance.

### 3.6.1 Validation du système par l'expert du domaine

La validation de notre système repose sur l'évaluation des résultats de différentes tâches de raisonnement sur des scénarios de diagnostic médical dans le domaine de la cardiologie. Selon

l'expertise du cardiologue, les tâches à tester sont les suivantes :

### 1. Instanciation probabiliste :

- **Hypothèse** :Être âgé, hypertendu et avoir un niveau de cholestérol élevé augmente le risque de développer une maladie coronarienne.
- **Scénario** :Le patient "*Patient1*" est âgé et hypertendu, sans présenter d'autres facteurs de risque favorisant la maladie.
- **Évidences** :Le tableau 3.1 représente des évidences du patient "*Patient1*".

<b>Patient 1</b>	<b>Sex = Female</b>	<b>Age = Old</b>	<b>diagnostic = Unk- now</b>
<b>Signs</b>	<b>Values</b>	<b>Risk factors</b>	<b>Values</b>
BMI	Normal	Obesity	False
DiastolicBP	Normal	IsHypertensive	True
SystolicBP	Normal	IsDiabetic	False
HeartRate	Medium	IsTouchedbyStroke	False
Glucose	Normal	CigsPerDay	Low
TotalCholesterol	High	IsSmoking	False

TABLE 3.1 – Les évidences du "*Patient1*".

- **Requêtes de raisonnement** : Calculer la probabilité que le patient a un risque de développer une maladie coronarienne.
- **Résultat** :La probabilité de développer une maladie coronarienne dans dix ans chez le "*Patient1*" est 89%.

### 2. Raisonnement sur L'incertitude liée aux propriétés de données :

- **Hypothèse** :Les hommes ont un risque plus élevé de développer des maladies cornariennes par rapport aux femmes.
- **Scénario** :Le sexe du patient "*Patient2*" est inconnu et il a un diagnostic positif d'avoir une maladie coronarienne sur une période de 10 ans.
- **Évidences** :Le tableau 3.2 représente des évidences du patient "*Patient2*".

<b>Patient 2</b>	<b>Sex = Unknow</b>	<b>Age = Old</b>	<b>diagnostic = True</b>
<b>Signs</b>	<b>Values</b>	<b>Risk factors</b>	<b>Values</b>
BMI	Normal	Obesity	False
DiastolicBP	Optimal	IsHypertensive	False
SystolicBP	Optimal	IsDiabetic	False
HeartRate	Medium	IsTouchedbyStroke	False
Glucose	Normal	CigsPerDay	Low
TotalCholesterol	Normal	IsSmoking	False

TABLE 3.2 – Les évidences du "Patient2".

- **Requêtes de raisonnement** : Calculer la probabilité que le patient soit une femme puis calculer la probabilité qu'il soit un homme.
- **Résultat** :Le patient "Patient2" est une femme avec une probabilité de 40% et il est homme avec une probabilité de 60%.

### 3. Raisonnement sur L'incertitude liée aux propriétés d'objets :

- **Hypothèse** :La principale cause des maladies cardiovasculaires est d'être fumeur.
- **Scénario** :Le patient "Patient3" est hypertendu et a un diagnostic positif d'avoir une maladie coronarienne sur une période de 10 ans. Cependant, l'information concernant sa consommation de tabac est indéterminée.
- **Évidences** :Le tableau 3.3 représente des évidences du patient "Patient3".

<b>Patient 3</b>	<b>Sex = Male</b>	<b>Age = Young</b>	<b>diagnostic = True</b>
<b>Signs</b>	<b>Values</b>	<b>Risk factors</b>	<b>Values</b>
BMI	Normal	Obesity	False
DiastolicBP	Mild hypertension	IsHypertensive	True
SystolicBP	High	IsDiabetic	False
HeartRate	Medium	IsTouchedbyStroke	False
Glucose	Normal	CigsPerDay	Unknow
TotalCholesterol	Normal	IsSmoking	Unknow

TABLE 3.3 – Les évidences de "Patient3".

- **Requêtes de raisonnement** : Calculer la probabilité que le patient "Patient3" fume.
- **Résultat** :Le patient "Patient3" est fumeur avec une probabilité de 56%.

### 4. Raisonnement sur L'incertitude liée aux ensembles des propriétés :

- **Hypothèse** :Effectuer les analyses demandées par le médecin permet d'obtenir un diagnostic plus fiable et précis.

- **Scénario :**Le patient "*Patient4*" souffre d'obésité morbide et présente des niveaux de tension artérielle systolique et diastolique indiquant une hypertension légère. il est donc recommandé de réaliser des analyses de glucose et de cholestérol pour évaluer la présence éventuelle d'autres maladies chroniques.
- **Évidences :**Le tableau 3.4 représente des évidences du patient "*Patient4*".

<b>Patient 4</b>	<b>Sex = Male</b>	<b>Age = Young</b>	<b>diagnostic = Unknow</b>
<b>Signs</b>	<b>Values</b>	<b>Risk factors</b>	<b>Values</b>
BMI	Morbide obesity	Obesity	True
DiastolicBP	Mild hypertension	IsHypertensive	Unknow
SystolicBP	Mild hypertension	IsDiabetic	Unknow
HeartRate	Medium	IsTouchedbyStroke	False
Glucose	Unknow	CigsPerDay	Low
TotalCholesterol	Unknow	IsSmoking	False

TABLE 3.4 – Les évidences de "*Patient4*".

- **Requêtes de raisonnement :** Calculer les probabilités de diabète chez le patient "*Patient4*" et son risque de développer une maladie coronarienne avant puis après les analyses.
- **Résultat :**Avant les résultats des analyses, le patient "*Patient4*" avait un risque de diabète de 6% et un risque de maladie coronarienne dans les dix prochaines années de 65%. Cependant, après avoir reçu des résultats indiquant un taux de glucose très élevé et un niveau de cholestérol élevé, la probabilité de diabète augmente à 88% et le risque de développer une maladie coronarienne atteint désormais 90%.

### 3.6.2 Validation du système par les mesures de performance

L'objectif de cette section est de valider le système en évaluant son efficacité, sa fiabilité et sa précision par rapport aux données du dataset. Pour réaliser la validation de notre système, nous avons utilisé la matrice de confusion qui permet de récapituler les prédictions correctes et incorrectes du système.

Nous avons utilisé la validation croisée k-fold qu'est une technique d'évaluation de modèles prédictifs. Elle consiste à diviser l'échantillon initial en k sous-échantillons de taille égale. Un sous-échantillon est choisi comme ensemble de validation, tandis que les autres servent d'ensemble d'entraînement. Ce processus est répété k fois, en utilisant chaque sous-échantillon une fois comme ensemble de validation.

Le tableau 3.5 présente la structure générale de la matrice de confusion.



	<b>Predicted :True</b>	<b>Predicted :False</b>
<b>Actual :True</b>	TP	FN
<b>Actual :False</b>	FP	TN

TABLE 3.5 – La matrice de confusion.

- **Vrai positif (TP)**, Le nombre de patients où la prédiction ainsi que la valeur actuelle sont toutes deux positives.
- **Vrai négatif (TN)**, Le nombre de patients où la prédiction ainsi que la valeur actuelle sont toutes deux négatives.
- **Faux positif (FN)**, Le nombre de patients où la prédiction négative alors que la valeur actuelle positive (La prédiction a été fausse).
- **Faux négatif (FP)**, Le nombre de patients où la prédiction positive alors que la valeur actuelle négative (La prédiction a été fausse).

Pour notre cas, nous avons utilisé une validation croisée à 10 plis (K=10) et avons obtenu le tableau 3.6 qui présente la matrice de confusion de notre système.

	<b>Predicted :True</b>	<b>Predicted :False</b>
<b>Actual :True</b>	411	172
<b>Actual :False</b>	73	574

TABLE 3.6 – La matrice de confusion de notre système.

Nous procéderons ensuite au calcul des mesures suivantes :

1. **Precision** également appelée valeur prédictive positive, est une mesure statistique qui représente la probabilité qu'un résultat positif d'un test ou d'une prédiction soit réellement positif. Dans notre cas, la précision répond à la question : parmi tous les patients étiquetés comme positifs, combien sont réellement positifs ?

La precision est calculée en utilisant la formule suivante :

$$Precision = TP / (TP+FP).$$

2. **Recall** est une mesure statistique utilisée pour évaluer la capacité d'un modèle de classification ou d'un système de prédiction à identifier correctement les exemples positifs parmi tous les exemples réellement positifs. Dans notre cas permet de répondre à la question : parmi tous Recall les patients vraiment positifs, combien ont été identifiés ?

le recall est calculé en utilisant la formule suivante :

$$Recall = TP / (TP+FN).$$

3. **F-score** est une mesure statistique utilisée pour évaluer la précision d'un modèle de classification ou d'un système de prédiction. Il combine à la fois la précision (capacité du modèle à identifier correctement les exemples positifs) et le rappel (capacité du modèle à retrouver tous les exemples positifs) en une seule valeur.

Le F-score est calculé en utilisant la formule suivante :

$$F\text{-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}).$$

4. **Accuracy** est une mesure statistique utilisée pour évaluer la performance d'un modèle de classification ou d'un système de prédiction. Elle représente le pourcentage d'exemples correctement classés par rapport au nombre total d'exemples.

Accuracy est calculé en utilisant la formule suivante :

$$\text{Accuracy} = (TP + TN) / (TP + FN + TN + FP).$$

Les résultats obtenus : **Precision** = 0.85, **Recall** = 0.70, **F-score** = 0.77, **Accuracy** = 0.80.

La figure 3.8 représente les résultats des mesures de performance obtenus par notre système.

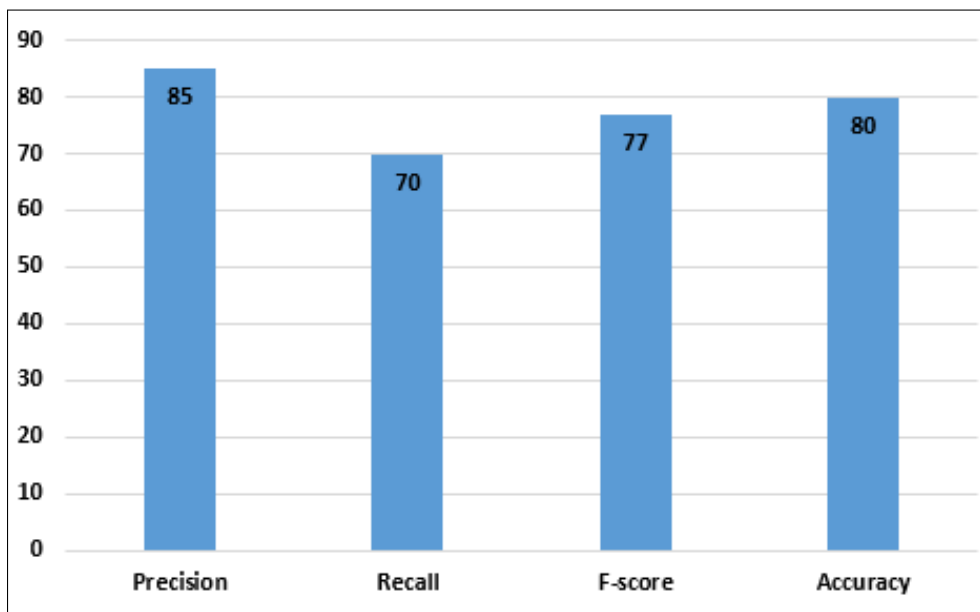


FIGURE 3.8 – Histogramme 1 : Precision, Recall, F-score, Accuracy.

### 3.7 Interprétation des résultats

Le raisonneur a montré des performances prometteuses, avec une précision élevée de 85% et une précision globale de 80%. Cependant, un rappel de 70% suggère qu'il peut y avoir des cas positifs qui ne sont pas détectés, indiquant un potentiel d'amélioration. Il est essentiel de revoir attentivement le dataset utilisé pour l'apprentissage afin de s'assurer de sa représentativité et de l'inclusion adéquate des cas positifs.

Pour améliorer le rappel du modèle, l'expert du domaine nous a recommandé d'ajouter des attributs supplémentaires pertinents pour le diagnostic des maladies coronariennes, tels que les antécédents familiaux et les symptômes.

### **3.8 Conclusion :**

Dans ce dernier chapitre, nous avons exposé notre raisonneur et présenté les différents outils, langages et API utilisés dans notre travail. Enfin, nous avons validé la performance du raisonneur et interprété les résultats obtenus.

---

# Conclusion et perspectives

---

Ce travail s'inscrit dans le domaine de la modélisation et du raisonnement sur des connaissances incertaines dans les ontologies. Dans cette étude, nous avons développé un raisonneur probabiliste spécifiquement conçu pour traiter les connaissances ontologiques incertaines.

Les ontologies jouent un rôle essentiel dans la représentation formelle des connaissances d'un domaine spécifique. Cependant, elles ne sont pas adaptées pour modéliser et raisonner sur des connaissances incertaines. Afin de relever ce défi, nous avons adopté une approche basée sur les réseaux bayésiens, qui offre une puissante méthodologie pour traiter l'incertitude.

Dans la première partie de ce mémoire, nous avons examiné les ontologies en tant que méthode de représentation et raisonnement des connaissances. Cependant, nous avons constaté que les ontologies sont limitées dans leur capacité à traiter l'incertitude. Nous avons donc opté pour les réseaux bayésiens, une approche qui permet de modéliser et de raisonner sur les connaissances incertaines de manière plus robuste. Par conséquent, afin de bénéficier des avantages des deux approches, nous nous sommes tournés vers les ontologies probabilistes. Ces ontologies combinent les caractéristiques bénéfiques des ontologies classiques et des réseaux bayésiens.

La deuxième partie a été consacrée à la conception de notre système en choisissant un dataset pertinent pour notre système et effectuant les prétraitements nécessaires afin qu'il soit adaptée à nos besoins. Ainsi, nous avons procédé à la construction de notre ontologie classique et du notre réseau bayésien en collaborant avec l'expert du domaine pour la modélisation de ses structures. Par la suite, les paramètres de réseau bayésien ont été appris en utilisant l'algorithme EM. Ensuite, nous avons créé un méta-modèle ontologique pour représenter les réseaux bayésiens, dans le but de l'utiliser dans l'algorithme de l'intégration de l'ontologie classique avec le réseau bayésien afin d'obtenir l'ontologie probabiliste finale, qui puisse être utilisée dans notre raisonneur. De plus, nous avons procédé à la conception de notre raisonneur probabiliste en modifiant ses différentes tâches.

Pour concrétiser notre conception, nous avons mis en œuvre notre raisonneur en présentant ses tâches à travers des interfaces graphiques. Ensuite, nous l'avons soumis à des tests avec l'expert du domaine et l'évalué avec des mesures de performance afin de valider et de démontrer l'efficacité de notre système.

Nos perspectives consistent à améliorer notre dataset en ajoutant des attributs supplémentaires pertinents pour le diagnostic et en remplaçant les instances du dataset par des cas réels de la population algérienne, afin d'obtenir un raisonnement adéquat pour cette démographie. Il est important de veiller à ce que les données acquises soient équilibrées pour tous les attributs des maladies chroniques présents dans le dataset, afin d'évaluer le raisonnement pour ces cas.

De plus, l'algorithme d'intégration de l'ontologie classique avec le réseau bayésien peut être amélioré en définissant d'autres méta-modèles ontologiques pour différents types de réseaux bayésiens. Cela permettrait de généraliser notre approche et de l'appliquer à d'autres types de réseaux bayésiens.

---

# Bibliographie

---

- [1] F. Amardeilh. *Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle*. PhD thesis, Université de Nanterre-Paris X, 2007.
- [2] F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, D. Nardi, et al. *The description logic handbook : Theory, implementation and applications*. Cambridge university press, 2003.
- [3] M. Baziz. *Indexation conceptuelle guidée par ontologie pour la recherche d'information*. PhD thesis, Toulouse 3, 2005.
- [4] J. Charlet, B. Bachimont, and R. Troncy. Ontologies pour le web sémantique. *Revue I3, numéro Hors Série «Web sémantique*, pages 43–63, 2004.
- [5] P. C. G. Da Costa. *Bayesian semantics for the Semantic Web*. George Mason University, 2005.
- [6] P. C. G. Da Costa, K. B. Laskey, and K. J. Laskey. Pr-owl : A bayesian ontology language for the semantic web. In *Uncertainty reasoning for the semantic web I : ISWC international workshops, URSW 2005-2007, revised selected and invited papers*, pages 88–107. Springer, 2008.
- [7] C. Dantas, R. Nunes, A. Canuto, and J. Xavier-Júnior. Instance hardness as a decision criterion on dynamic ensemble structure. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 108–113. IEEE, 2019.
- [8] D. De Nizza, J. Ortiz, H. Meurisse, and P.-Y. Schobbens. Formalisation et construction d'une ontologie dans le domaine des infections orthopédiques. In *IC-24èmes Journées francophones d'Ingénierie des Connaissances*, 2013.
- [9] J.-B. Denis and M. Scutari. Réseaux bayésiens avec r. In *Réseaux bayésiens avec R*. Edp sciences, 2021.

- [10] K. Dramé. *Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical*. PhD thesis, Bordeaux, 2014.
- [11] O. François. *De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes*. PhD thesis, INSA de Rouen, 2006.
- [12] O. François and P. Leray. Etude comparative d'algorithmes d'apprentissage de structure dans les réseaux bayésiens. *Rencontres des Jeunes Chercheurs en IA*, 2003.
- [13] F. Fürst. L'ingénierie ontologique-rapport de recherche. Technical report, Technical report, IRIN-Université de Nantes, 2002. <http://www.sciences...>, 2002.
- [14] A. Gómez-Pérez and R. Benjamins. Overview of knowledge sharing and reuse components : Ontologies and problem-solving methods. IJCAI and the Scandinavian AI Societies. CEUR Workshop Proceedings, 1999.
- [15] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2) :199–220, 1993.
- [16] N. Hai, D. Gong, and S. Liu. Ontology knowledge base combined with bayesian networks for integrated corridor risk warning. *Computer Communications*, 174 :190–204, 2021.
- [17] N. J. Hernandez. *Ontologies de domaine pour la modélisation du contexte en recherche d'information*. PhD thesis, Université Paul Sabatier-Toulouse III, 2005.
- [18] E. Hlel, S. Jamoussi, and A. B. Hamadou. Intégration d'un réseau bayésien dans une ontologie. In *IC-25èmes Journées francophones d'Ingénierie des Connaissances*, pages 295–297, 2014.
- [19] R. Ishak. *Modélisation De La Connaissance Incertaine Dans Le Web Sémantique*. PhD thesis, Université Saad Dahleb - Blida, 2020.
- [20] F. Kerdjoudj. *Gestion de l'incertitude dans le processus d'extraction de connaissances à partir de textes*. PhD thesis, Paris Est, 2015.
- [21] A. Le Pham. *De l'optimisation à la décomposition de l'ontologique dans la logique de description*. PhD thesis, Université Nice Sophia Antipolis, 2008.
- [22] P. Leray. *Réseaux bayésiens : Apprentissage et diagnostic de systemes complexes*. PhD thesis, Université de Rouen, 2006.
- [23] J. A. Lopo and K. D. Hartomo. Evaluating sampling techniques for healthcare insurance fraud detection in imbalanced dataset. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(2) :223–238, 2023.

- [24] K. Manel. Intégration d'ontologies dans le cadre du web sémantique : une détection des relations sémantiques basée sur le rāpc. *Doctorat en Sciences en Informatique, Université Mentouri Constantine*, 2011.
- [25] D. L. McGuinness, R. Fikes, J. Hendler, and L. A. Stein. Daml+ oil : an ontology language for the semantic web. *IEEE Intelligent Systems*, 17(5) :72–80, 2002.
- [26] M. Mezati. *UNE NOUVELLE APPROCHE POUR LA SÉMANTIQUE DES ENVIRONNEMENTS VIRTUELS*. PhD thesis, Université Mohamed Khider-Biskra, 2017.
- [27] A. B. Mrad. *Observations probabilistes dans les réseaux bayésiens*. PhD thesis, Université de Valenciennes et du Hainaut-Cambresis ; École nationale d . . . , 2015.
- [28] R. Neches, R. E. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W. R. Swartout. Enabling technology for knowledge sharing. *AI magazine*, 12(3) :36–36, 1991.
- [29] H.-T. Nguyen. *Réseaux bayésiens et apprentissage ensembliste pour l'étude différentielle de réseaux de régulation génétique*. PhD thesis, Université de Nantes, 2012.
- [30] R. Peñaloza. Introduction to probabilistic ontologies. In *Reasoning Web International Summer School*, pages 1–35. Springer, 2020.
- [31] F. Petiet. *Réseau bayésien dynamique hybride : application à la modélisation de la fiabilité de systèmes à espaces d'états discrets*. PhD thesis, Université Paris-Est, 2019.
- [32] D. Poole, C. Smyth, and R. Sharma. Semantic science : Ontologies, data and probabilistic theories. In *Uncertainty Reasoning for the Semantic Web I : ISWC International Workshops, URSW 2005-2007, Revised Selected and Invited Papers*, pages 26–40. Springer, 2008.
- [33] F. A. Setiawan, E. K. Budiardjo, and W. C. Wibowo. Bynowlife : a novel framework for owl and bayesian network integration. *Information*, 10(3) :95, 2019.
- [34] S. Staab and A. Maedche. *Axioms are objects, too : Ontology engineering beyond the modeling of concepts and relations*. Univ., 2000.
- [35] R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering : principles and methods. *Data & knowledge engineering*, 25(1-2) :161–197, 1998.
- [36] S. Zermani. *Implémentation sur SoC des réseaux Bayésiens pour l'état de santé et la décision dans le cadre de missions de véhicules autonomes*. PhD thesis, Université de Bretagne occidentale-Brest, 2017.
- [37] S. Zhang, Y. Sun, Y. Peng, X. Wang, et al. Bayesowl : A prototype system for uncertainty in semantic web. In *Proceedings of the International Conference on Artificial Intelligence*, 2009.