

**UNIVERSITÉ SAAD DAHLEB DE BLIDA 1**

**Faculté des sciences**

Département d'informatique



**MÉMOIRE DE MASTER**

**En Informatique**

Option : Ingénierie Des Logiciels

**THÈME :**

**Prédiction des fuites dans les réseaux  
d'alimentation en eau potable en  
utilisant les ontologies probabilistes**

Réalisé par

KESSI Manal

HAMITOUCHE Chakib

Mme. ABED Présidente

Mme. BOUMAHDI Examinatrice

Mme. FAREH Promotrice

23 Septembre 2020



# Remerciements

Ce travail est le résultat d'un dur labeur et de beaucoup de sacrifices ; nos remerciements vont d'abord à ALLAH, créateur de l'univers qui nous a doté d'intelligence, et nous a maintenu en santé mentale et physique pour mener à bien cette année d'étude, à qui nous profondément et sincèrement reconnaissants.

Nous tenons à remercier toutes les personnes commençons par nos parents, nos familles par leur sacrifices pour nous fournir les meilleurs conditions de travail, ainsi nos amis, qui ont contribué au succès de notre travail et qui nous ont aidé lors de la rédaction de ce mémoire.

Nous voudrions remercier, notre promotrice Mme. M. FAREH, professeur à l'université de Saad Dahleb Blida -1-, qui fut la première à nous faire découvrir le sujet qui a guidé notre mémoire pour avoir accepté de nous guider par sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.

Nous désirons exprimer notre reconnaissance envers Mr. I. RIALI, qui nous a apporté son support moral et intellectuel tout au long de notre démarche, et pour avoir eu la patience de répondre à nos innombrables questions.

Nous tenons exprimer notre reconnaissance envers M. ROMMEL Carvalho, Data Science Manager, Auction and Delivery, Marketing Science RD chez Facebook, qui nous a aidé durant l'utilisation de son outils qui nous a servi de base de développement.

Nous tenons à exprimer notre gratitude à l'ensemble des employés et responsables de la société SEAAL Tipaza, pour leur contribution et leur disponibilité pour l'apport technique concernant le domaine de travail.

Nous remercions également tous les membres du jury pour nous avoir honorés par leur présence et pour avoir accepté d'évaluer notre travail.



## Résumé

De nos jours, la détection des fuites d'eau se fait tardivement qu'après leurs apparitions, et l'impact qu'elles causent sur les utilisateurs et leur environnement ce qui causent des problèmes majeurs pour les municipalités algériennes. Les coupures d'eau, les inondations, l'apparition des maladies liées à l'eau. La mise en œuvre d'un système de prédiction des fuites devient une nécessité, un système qui permet de faciliter aux services publics et aux entreprises concernées la prédiction des fuites dans un réseau d'AEP.

Pour ce faire, nous avons consacré ce travail à l'utilisation des ontologies probabilistes, pour le but de construire un système informatique dédié à la prédiction des fuites dans les conduites des réseaux d'AEP. L'ontologie probabiliste est modélisée selon la méthodologie UMP-ST, basée sur un réseau bayésien multi-entités. La construction de ce dernier se réalise par la définition de l'ensemble de ses composants MFragments et ses différents types de nœuds, ainsi que les distributions de probabilité liées aux nœuds. L'utilisation du système s'effectue par l'introduction des données liées à une conduite, pour avoir un résultat concernant les probabilités sur la présence ou l'absence d'une fuite dans cette conduite.

### **Mots clés :**

Ontologies probabilistes, réseaux bayésiens, AEP, prédiction des fuites, Réseau Bayésien Multi Entité, inférence.

## **Abstract**

Nowadays, the detection of water leaks is done belatedly after their appearance and the impact they cause on users and their environment, which causes major problems for Algerian municipalities. Water cuts, floods, the onset of water-related diseases. The implementation of a leak prediction system is becoming a necessity, a system allowing the public services and companies concerned to better predict leaks in an AEP network.

Therefore, we have dedicated this work to the use of probabilistic ontologies, for the purpose of building a computer system dedicated to the prediction of leaks in the pipes of AEP networks. The probabilistic ontology is modeled according to the UMP-ST methodology, based on a Multi-Entity Bayesian Network. The construction of the latter is carried out by defining all of its MFragments components and its different types of nodes, as well as the probability distributions related to the nodes. The use of the system is carried out by the introduction of data related to a pipe, to obtain a result concerning the probabilities of the presence or absence of a leak in this pipe.

### **Keywords :**

Probabilistic ontologies, bayesian networks, AEP, leak prediction, Multi Entity Bayesian Network, Inference.

## ملخص

في الوقت الحاضر يتم الكشف عن تسريبات المياه في وقت متأخر بعد ظهورها وتأثيرها على المستخدمين وبيئتهم مما يسبب مشاكل كبيرة للبلديات الجزائرية. انقطاع المياه ، الفيضانات ، تفشي الأمراض المرتبطة بالمياه. أصبح تنفيذ نظام التنبؤ بالتسرب ضرورة ، وهو نظام يسهل على المرافق والشركات المعنية التنبؤ بالتسريبات في شبكة AEP للقيام بذلك ، كرسنا هذا العمل لاستخدام الأنطولوجيا الاحتمالية ، من أجل بناء نظام كمبيوتر شخصي للتنبؤ بالتسريبات في أنابيب شبكات AEP

تم نمذجة الأنطولوجيا الاحتمالية وفقاً لمنهجية UMP-ST ، بناءً على شبكة بايزي متعددة الكيانات . يتم تنفيذ هذا الأخير من خلال تحديد جميع مكونات MFragments وأنواع العقد المختلفة ، بالإضافة إلى توزيعات الاحتمالات المتعلقة بالعقد.

يتم استخدام النظام عن طريق إدخال البيانات المتعلقة بالأنبوب ، للحصول على نتيجة تتعلق باحتمالات وجود أو عدم وجود تسرب في هذا الأنبوب.

الكلمات الدالة :

الأنطولوجيا الاحتمالية ، شبكات بايز ، AEP ، تنبؤ التسرب ، شبكة بايزي متعددة الكيانات ، الاستدلال.

---

# Table des matières

---

<b>Table des figures</b>	<b>11</b>
<b>Liste des tableaux</b>	<b>13</b>
<b>Liste des abréviations</b>	<b>15</b>
<b>Introduction Générale</b>	<b>17</b>
Contexte de travail . . . . .	17
Problématique . . . . .	17
Objectifs du travail . . . . .	18
Organisation du mémoire . . . . .	19
<b>1 Ontologies classiques et probabilistes</b>	<b>21</b>
1.1 Introduction . . . . .	21
1.2 Ontologies classiques . . . . .	21
1.2.1 Définition . . . . .	21
1.2.2 Composants de l'ontologie . . . . .	22
1.2.3 Rôle de l'ontologie . . . . .	22
1.2.4 Construction des ontologies classiques . . . . .	23
1.2.5 Langages des ontologies classiques . . . . .	24
1.2.6 Limite des ontologies classiques . . . . .	25
1.3 Ontologies Probabilistes . . . . .	25
1.3.1 Définition . . . . .	26
1.3.2 Langages des ontologies probabilistes . . . . .	26
1.3.3 Avantages du langage PR-OWL . . . . .	27
1.3.4 Méthode UMP-ST . . . . .	27
1.4 Conclusion . . . . .	28
<b>2 Réseaux bayésiens standard et multi-entités</b>	<b>29</b>
2.1 Introduction . . . . .	29

2.2	Réseaux bayésiens . . . . .	29
2.2.1	Définition . . . . .	29
2.2.2	Raisons d'utilisation des réseaux bayésiens . . . . .	30
2.2.3	Apprentissage de structure et des paramètres . . . . .	31
2.2.4	Algorithme EM . . . . .	32
2.2.5	Inférence dans les réseaux bayésiens . . . . .	32
2.3	Réseaux bayésiens multi entités (MEBN) . . . . .	34
2.3.1	Fragment MEBN (MFrag) . . . . .	34
2.3.2	Théorie MEBN (MTheory) . . . . .	35
2.3.3	Inférence dans les réseaux bayésiens multi-entités . . . . .	35
2.4	Conclusion . . . . .	36
<b>3</b>	<b>Présentation des Réseaux d'AEP</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Définition d'AEP . . . . .	37
3.3	Caractéristiques des AEP . . . . .	37
3.4	Classifications des réseaux d'AEP . . . . .	39
3.4.1	Réseaux ramifiés . . . . .	39
3.4.2	Réseaux maillés . . . . .	40
3.4.3	Réseaux étagés . . . . .	40
3.4.4	Réseaux à alimentation distinctes . . . . .	40
3.5	Risques de défaillance sur les AEP . . . . .	41
3.5.1	Fuites . . . . .	41
3.5.2	Casses . . . . .	41
3.6	Facteurs générateurs des risques . . . . .	42
3.7	Méthodes de détection existantes . . . . .	43
3.7.1	Mesure des vibrations . . . . .	43
3.7.2	Utilisation des réseaux bayésiens de croyance (BBN) . . . . .	43
3.7.3	Traitement des valeurs de pression d'eau . . . . .	44
3.7.4	Technique informatisée d'analyse de la pression d'eau . . . . .	44
3.8	Comparaison des méthodes . . . . .	46
3.9	Analyse horizontale . . . . .	46
3.10	Conclusion . . . . .	47
<b>4</b>	<b>Conception du système de prédiction des fuites</b>	<b>49</b>
4.1	Introduction . . . . .	49
4.2	Problématique . . . . .	49
4.3	Motivation du travail . . . . .	49
4.4	SEAAL . . . . .	50
4.5	Dataset . . . . .	50

4.6	Scénario global . . . . .	50
4.7	Création de l'ontologie probabiliste . . . . .	52
4.7.1	Exigences . . . . .	52
4.7.2	Besoins (Goals) . . . . .	52
4.7.3	Analyse et Conception . . . . .	54
4.7.4	Implémentation . . . . .	59
4.7.5	Test . . . . .	63
4.7.6	Résultat final . . . . .	66
4.8	Conclusion . . . . .	67
<b>5</b>	<b>Implémentation du système de prédiction des fuites</b>	<b>69</b>
5.1	Introduction . . . . .	69
5.2	Outils utilisés pour le développement du système . . . . .	69
5.3	Mise en œuvre . . . . .	71
5.4	Base de connaissances . . . . .	71
5.5	Interfaces . . . . .	72
5.5.1	Interface de l'accueil . . . . .	72
5.5.2	Interface d'ajout d'une conduite . . . . .	73
5.5.3	Interfaces de la prédiction . . . . .	74
5.6	Validation du système . . . . .	76
5.7	Conclusion . . . . .	77
<b>6</b>	<b>Conclusion générale</b>	<b>79</b>
	<b>Bibliographie</b>	<b>81</b>

---

# Table des figures

---

1.1	Phases de construction des ontologies . . . . .	24
1.2	Composition d'un MTheory de PR-OWL (Costa, 2005) . . . . .	26
1.3	Principales différences entre OWL et PR-OWL (Costa, 2005). . . . .	27
1.4	Cycle de modélisation de l'ontologie probabiliste (POMC) (Carvalho et al., 2016) . . . . .	28
2.1	Exemple simple d'un réseau bayésien (Ben Mrad, 2015) . . . . .	30
3.1	Réseau d'AEP "ramifié" (ramifié, 2020) . . . . .	39
3.2	Réseau d'AEP"en maillés" (en mailles, 2020) . . . . .	40
3.3	Une fuite (leaking2020). . . . .	41
3.4	Une casse (iStock, 2016). . . . .	42
4.1	Schéma global du système. . . . .	51
4.2	Les entités. . . . .	54
4.3	Dataset préparé. . . . .	56
4.4	Structure du réseau bayésien. . . . .	57
4.5	Paramètres de l'algorithme EM. . . . .	58
4.6	Réseau bayésien. . . . .	58
4.7	Diagramme de classe. . . . .	59
4.8	MFrag des facteurs. . . . .	60
4.9	MFrag des fuites. . . . .	61
4.10	MFrag des conduites. . . . .	61
4.11	LPD du nœud "TypeS" (MFrag Facteur). . . . .	62
4.12	LPD du nœud "Corrosivité" (MFrag Facteur). . . . .	62
4.13	LPD du nœud "Matériaux" (MFrag Facteur). . . . .	63
4.14	LPD du nœud "presenceFuite" (MFrag Fuite). . . . .	63
4.15	SSBN du nœud "presenceFuite". . . . .	64
4.16	Résultat du SSBN généré suivant le scénario n01. . . . .	65
4.17	Résultat du SSBN généré suivant le scénario n02. . . . .	66

5.1	La Base de connaissance . . . . .	72
5.2	Interface "Accueil". . . . .	73
5.3	Interface "Supprimer". . . . .	73
5.4	Interface "Ajouter Conduite". . . . .	74
5.5	Interface "Prédiction". . . . .	75
5.6	Interface "Probabilités de présence de la fuite". . . . .	75
5.7	Interface "Probabilités d'absence de la fuite". . . . .	76
5.8	Interface "Informations sur le système". . . . .	77

---

# Liste des tableaux

---

3.1	Tableau de comparaison des méthodes de détection existantes . . . . .	45
4.1	Tableau explicatif du schéma global du système . . . . .	52
4.2	Propriétés de l'entité Conduite . . . . .	55
4.3	Propriétés de l'entité Facteur . . . . .	55
4.4	Propriétés de l'entité Fuite . . . . .	55
4.5	Catégorisation des facteurs. . . . .	56



---

# Liste des abréviations

---

**AEP** Alimentation en Eau Potable. 5, 6, 9, 11, 17–19, 37, 39–41, 43, 44, 47, 50, 52, 79, 80

**ELM** Extreme Learning Machine. 44, 45

**MEBN** Multi-Entity Bayesian Network. 9, 26, 34, 35, 50, 52, 55, 59–63, 70

**NRMSE** Normalize Root Mean Square Error. 44

**OP** Ontologie probabiliste. 55

**RB** Réseau bayésien. 31, 57

**SSBN** Situation-Specific Bayesian Network. 11, 35, 51, 52, 62, 64–66

**SVM** Support Vector Machines. 44, 45

**UP** Processus unifié. 27



---

# Introduction générale

---

## Contexte de travail

De nos jours, la société moderne exige une alimentation constante et stable en matière d'eau potable, pour satisfaire aux besoins quotidiens vu que c'est un composant vital utilisé dans plusieurs domaines, cette alimentation doit être stable en matière de présence, satisfaisante en matière de quantité, et propre en matière d'eau conduite aux robinets.

Pour ce fait, les réseaux d'Alimentation en Eau Potable (AEP) ont été conçus, un ensemble de composants hydrauliques et technologiques assemblés d'une manière structurée pour but d'acheminer l'eau à partir de la source ou le lieu de stockage (barrages, châteaux d'eaux, baches à eau), jusqu'aux robinets des utilisateurs (individus ou usines) en utilisant des conduites liées d'une manière ferme sans aucune pollution.

Ces réseaux peuvent être structurés sur selon quatre (04) structures : ramifiés, maillés, étagés, à alimentation distinctes), cette structure est basée selon la nature de la région ou le réseau est déployé, les régions rurales nécessitent l'utilisation des réseaux ramifiés alors les régions urbaines nécessitent des réseaux maillés.

Le but de ces structures est d'assurer une bonne constante alimentation, ce qui nécessite la gestion des conduites des réseaux quotidiennement afin d'éviter les problèmes d'alimentation, et surtout le problème majeur, qui est l'apparition soudaine des fuites dans les conduites, un dilemme résultant de plusieurs facteurs et qui cause une grande perturbation de la distribution avec des pertes d'eau plus ou moins importantes et des ennuis aux consommateurs.

## Problématique

Les fuites consistent de gros soucis sur plusieurs plans, sur le plan économique, avoir une fuite est synonyme de perte d'eau et aussi des frais supplémentaires pour les réparations, les

réparations nuisent à l'image esthétique sur la surface de la partie touchée, les eaux perdues se stagnent dans des lacs qui deviennent avec le temps un milieu de culture pour les insectes et aux maladies transmissibles par l'eau, ce qui constitue un grand problème sanitaire pour la société.

Ces problèmes donnent impérativement une cause pour le développement de méthodes de prévention, plusieurs méthodes ont été développées pour détecter les fuites dans les conduites, entre méthodes purement hydrauliques jusqu'aux méthodes informatiques passant par les méthodes acoustiques, nous citons cependant la méthode de mesure des vibrations qui est une méthode acoustique ou les vibrations des conduites font matière d'étude de la méthode ainsi que la méthode basée sur l'utilisation des réseaux bayésiens de croyance (BBN), qui est à son tour une méthode informatique et aussi la technique informatisée d'analyse de la pression d'eau, ces méthodes permettaient seulement la détection et la localisation des fuites mais qu'après leur apparition, donc on ne peut pas s'échapper des dégâts qui peuvent être causés au moment d'apparition de la fuite et de sa réparation.

## **Objectifs du travail**

A cet effet, pour minimiser les risques liés aux fuites, il est très intéressant de prédire celles-ci avant même de leur apparition dans le réseau, cela permettra d'intervenir dans les meilleurs délais.

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

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

L'objectif de ce projet est d'utiliser les ontologies probabilistes pour la modélisation et la représentation de la connaissance incertaine afin de prédire les fuites dans les réseaux d'AEP, et ce, pour aider le gestionnaire du réseau à prendre la bonne décision.

La visée de ce travail a été mise principalement pour :

- Comprendre la notion des ontologies.

- Introduire la notion des probabilités dans une ontologie.
- Assimiler le domaine des réseaux d'AEP.
- Réaliser un système de prédiction des fuites.

## **Organisation du mémoire**

Afin d'assurer une recherche bien organisée, nous avons suivi la méthode de travail mentionnée ci-dessous :

- Chapitre 1 : Introduira le travail avec une présentation des ontologies et leurs types pour assurer une bonne introduction compréhensive dans le thème.
- Chapitre 2 : Énoncera la notion des réseaux bayésiens.
- Chapitre 3 : Présentera le domaine du travail, les réseaux d'AEP.
- Chapitre 4 : Fournira la conception du système.
- Chapitre 5 : Portera sur la mise en œuvre du système en expliquant son implémentation.

Ce mémoire sera conclu par une conclusion générale qui va servir de récapitulatif des notions abordées précédemment.



## Chapitre 1

---

# Ontologies classiques et probabilistes

---

## 1.1 Introduction

Les connaissances étaient depuis toujours une entité d'une grande importance pour les différents domaines, ces connaissances sont souvent abstraites et non structurables sous une forme d'idées et de pensées partagées sur un ou plusieurs de personnes, donc leur existence et leur bonne exploitation dépendent de la présence des connaisseurs, d'où vient le besoin de la structuration des connaissances afin de leur donner une image structurée et compréhensible et transmissible, les auteurs ont cherché et ont pu développer la technique des ontologies, le terminal du travail des auteurs était la définition de deux (02) nouveaux concepts, les ontologies classiques et les ontologies probabilistes.

Les ontologies classiques ont l'incapacité de représenter et de raisonner sur l'incertitude, ce qui nous amène dans ce chapitre à représenter deux (02) types des ontologies, l'ontologie classique et probabiliste, leurs différents composants, rôles, et langages de représentation et la définition de la méthodologie UMP-ST qui permet la construction de l'ontologie probabiliste.

## 1.2 Ontologies classiques

L'ontologie représente des connaissances pour modéliser un problème en respectant les exigences du problème.

### 1.2.1 Définition

Neches et al. ([Neches et al., 1991](#)) furent les premiers à en proposer une définition : « 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 ».

La définition de (Gruber, 1993) est la plus citée dans la littérature : « une ontologie est une spécification explicite d'une conceptualisation ».

Elle a été légèrement modifiée par Borts (Schinkel et al., 1997) : « une spécification explicite et formelle d'une conceptualisation partagée ».

Celle-ci s'explique ainsi (Studer et al., 1998) : explicite signifie que le type des concepts et les contraintes sur leurs utilisations sont explicitement définies, formelle se réfère au fait que la spécification doit être lisible par une machine, partagée se rapporte à la notion selon laquelle une ontologie capture la connaissance con-sensuelle, qui n'est pas propre à un individu mais validée par un groupe, conceptualisation se réfère à un modèle abstrait d'un certain phénomène du monde basé sur l'identification des concepts pertinents de ce phénomène (Hammache and Ahmed-Ouamer, 2006).

## 1.2.2 Composants de l'ontologie

Les connaissances dans les ontologies sont principalement formalisées en utilisant cinq types de composants (Staab and Maedche, 2000) :

1. **Les concepts (ou classes) :** Un concept est un constituant de la pensée (un principe, une idée, une notion abstraite) sémantiquement évaluable et communicable. L'ensemble des propriétés d'un concept constitue sa compréhension ou son intention et l'ensemble des êtres qu'il englobe son extension.
2. **Les relations (ou propriétés) :** Elles représentent des interactions entre les concepts, elles permettent de construire des représentations complexes de la connaissance du domaine, elles établissent des liens sémantiques binaires, organisables hiérarchiquement.
3. **Les fonctions :** Elles présentent des cas particuliers de relations dans lesquelles le nième élément de la relation est unique pour les  $n-1$  éléments précédents. Formellement, les fonctions sont définies telles que :  $F : c_1 * c_2 * ... * c_{n-1} * c_n$ .
4. **Les axiomes (ou règles) :** Les axiomes sont des expressions qui sont toujours vraies. Ils ont pour but de définir dans un langage logique la description des concepts et des relations permettant de représenter leur sémantique. Ils représentent les intentions des concepts et des relations du domaine et, de manière générale, les connaissances n'ayant pas un caractère strictement terminologique.
5. **Les instances (ou individus) :** Elles constituent la définition extensionnelle de l'ontologie, elles sont utilisées pour représenter des éléments dans un domaine donné.

## 1.2.3 Rôle de l'ontologie

Plusieurs chercheurs se sont intéressés à la finalité des exploitations des ontologies. Nous précisons dans ce qui suit les rôles et l'intérêt des ontologies au sein des systèmes à base de connaissances (SBC) et du Web Sémantique (Kolli, 2011) :

- **Connaissances du domaine d'un SBC** : Les ontologies servent à représenter les connaissances du domaine d'un SBC. En particulier, elles servent de squelette à la représentation des connaissances du domaine dans la mesure où elles décrivent les objets, leurs propriétés et la façon dont ils peuvent se combiner pour constituer des connaissances du domaine complètes.
- **Communication** : Les ontologies peuvent intervenir dans la communication entre personnes, organisations et logiciels.  
En effet, les ontologies servent par exemple, à créer au sein d'un groupe ou d'une organisation un « vocabulaire conceptuel commun ». Dans ce cas, on est plutôt dans le cadre d'une ontologie informelle. Dans le cas de la communication entre personnes et systèmes, l'ontologie est formelle et sert en général une tâche précise dans le SBC ou le système d'information. L'ontologie est un puissant moyen pour lever les ambiguïtés dans les échanges ([Uschold et al., 1996](#)).
- **Interopérabilité** : Le développement et l'implantation d'une représentation explicite d'une compréhension partagée dans un domaine donné, peut améliorer la communication, qui à son tour permet une plus grande réutilisation, un partage plus large et une interopérabilité plus étendue ([Uschold et al., 1996](#)). L'interopérabilité est donc une spécialisation de la communication qui permet de répertorier les concepts que des applications peuvent s'échanger même si elles sont distantes et développées sur des bases différentes.
- **Aide à la spécification de systèmes** : La plupart des logiciels conventionnels sont construits avec une conceptualisation implicite et que la nouvelle génération des systèmes utilisant les travaux en intelligence artificielle devrait être basée sur une conceptualisation explicitement représentée ([Mizoguchi et al., 1996](#)).  
En effet, l'ontologie fournit une classification des objets que doit manipuler le système.
- **Indexation et la recherche d'information** : Dans le Web Sémantique, les ontologies y sont utilisées pour déterminer les index conceptuels décrivant les ressources sur le Web.

#### 1.2.4 Construction des ontologies classiques

Les ontologies étant destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel.

En particulier, les ontologies doivent être considérées comme des objets techniques évolutifs et possédants un cycle de vie qui nécessite d'être spécifié. Les activités liées aux ontologies sont d'une part des activités de gestion de projet (planification, contrôle, assurance qualité) et d'autre part des activités de développement (spécification, conceptualisation, formalisation) ; s'y ajoutent des activités transversales de support telles que l'évaluation, la documentation, la gestion de la configuration ([Baneyx, 2007](#)).

Un cycle de vie inspiré du génie logiciel est proposé dans. Il comprend une étape initiale d'évaluation des besoins, une étape de construction, une étape de diffusion, et une étape d'utilisation. Après chaque utilisation significative, l'ontologie et les besoins sont réévalués et l'ontologie peut être étendue et, si nécessaire, en partie reconstruite. La phase de construction peut être décomposée en 3 étapes : conceptualisation, ontologisation, opérationnalisation (Fig. 1.1).

- La conceptualisation : identification des connaissances contenues dans un corpus représentatif du domaine.
- L'ontologisation : formalisation, autant que possible, du modèle conceptuel obtenu à l'étape précédente.
- L'opérationnalisation : transcription de l'ontologie dans un langage formel et opérationnel de représentation de connaissances.

L'étape d'ontologisation peut être complétée d'une étape d'intégration au cours de laquelle une ou plusieurs ontologies vont être importées dans l'ontologie à construire. Dans ce travail, nous s'intéressons à l'intégration des ontologies qui sont décrites en LDs.

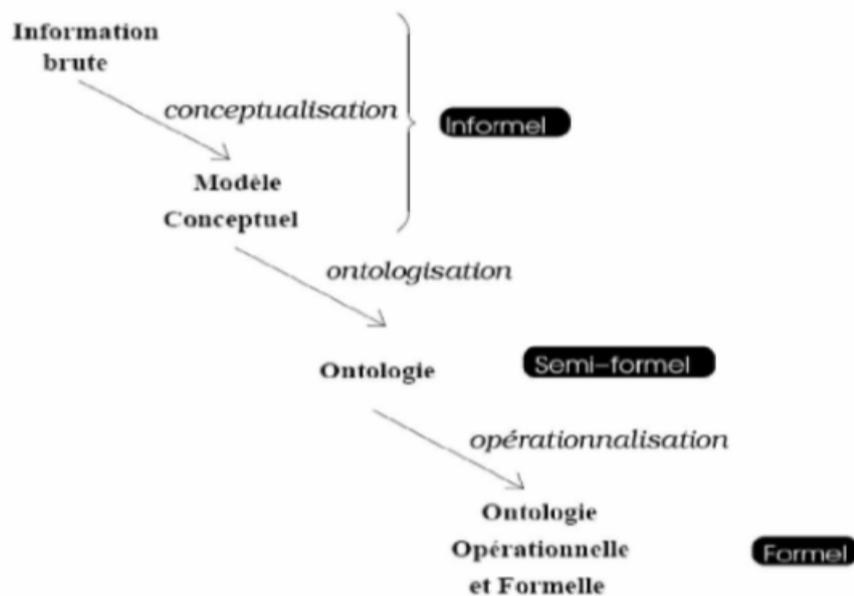


FIGURE 1.1 – Phases de construction des ontologies (Kolli, 2011)

### 1.2.5 Langages des ontologies classiques

— **RDF et RDFS** : Resource Description Framework (RDF) est le langage de base du Web sémantique, publié par le consortium W3C le 10 février 2004.

C'est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs méta-données, de façon à permettre le traitement automatique des telles descriptions.

RDF a pour rôle de relier chaque concept de base à d'autres définitions, afin de lui donner un sens.

Les documents RDF peuvent être écrits en différentes syntaxes, y compris en XML.

Il est possible d'avoir recours à d'autres syntaxes pour exprimer les triplets, ci dessous présentés.

RDF est simplement une structure de données constituée de nœuds et organisée en graphe. Bien que RDF/XML (la version XML proposée par W3C) ne soit qu'une sérialisation du modèle, elle est souvent appelée RDF.

Un document structuré en RDF est un ensemble de triplets.

Un triplet RDF est une association : <Sujet, objet, prédicat> - Le sujet représente la ressource à décrire.

- L'objet représente une donnée ou une autre ressource.

- Le prédicat représente un type de propriété applicable à cette ressource ([Ghorbel et al., 2008](#)).

— **OWL** : Ontologie Web Language (OWL) est, comme RDF, un langage profitant de l'universalité syntaxique de XML.

Fondé sur la syntaxe de RDF/XML, OWL offre un moyen de décrire des ontologies web : Il est justement un langage d'ontologies.

Si RDF et RDFS apportent à l'utilisateur la capacité de décrire des classes (avec des constructeurs) et des propriétés, OWL intègre, en sus, des outils de comparaison des propriétés et des classes : identité, équivalence, contraire, cardinalité, symétrie, transitivité, disjonction, etc. Ainsi, OWL offre aux machines une plus grande capacité d'interprétation du contenu web grâce à un vocabulaire plus large et à une vraie sémantique formelle.

OWL se compose de trois sous langages d'expressivité croissante : OWL Lite, OWL DL et OWL Full. - Toute ontologie OWL Lite correcte est une ontologie OWL DL légale. - Toute ontologie OWL DL correcte est une ontologie OWL Full légale. - Toute conclusion OWL Lite valide est une conclusion OWL DL valide. - Toute conclusion OWL DL valide est une conclusion OWL Full valide ([Ghorbel et al., 2008](#)).

## 1.2.6 Limite des ontologies classiques

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.

## 1.3 Ontologies Probabilistes

Plus les données sont grandes plus l'incertitude règne entre les liens, les ontologies probabilistes traitent l'incertitude dans l'environnement informatique.

### 1.3.1 Définition

D'après (Da Costa et al., 2006), 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 par exemple le modèle graphique dirigé Réseau Bayésien que nous allons utiliser dans la suite de notre travail, Ceux-ci joignent dans la même représentation la théorie des probabilités et la théorie des graphes. (Hlel et al., 2014).

### 1.3.2 Langages des ontologies probabilistes

PR-OWL : Étant une extension d'OWL, compatible avec la version 2.0 d'OWL, ayant la capacité de faire des raisonnements probabilistes basés sur les réseaux bayésiens avec entités multiples (MEBN : Multi Entity Bayesian Network).

**MEBN** : Est un langage probabiliste de premier ordre, il représente les concepts du monde comme des entités ayant des relations entre elles.

PR-OWL fait donc une représentation formelle et explicite des connaissances du domaine étudié. Outre les différentes caractéristiques d'OWL, PR-OWL tient compte des régularités statistiques des champs et conclut sur de nouvelles formes de connaissance par inférence tout en tenant compte de l'incertitude autour de ces connaissances (Da Costa et al., 2006).

PR-OWL utilise une autre forme de combinaison de classes, au lieu de (< sujet > < prédicat > < objet >) d'OWL, et les relations qui les relient ensemble. (Fig. 1.2) présente les classes dans un ovale et les relations qu'elles entretiennent ensemble sont indiquées par des flèches.

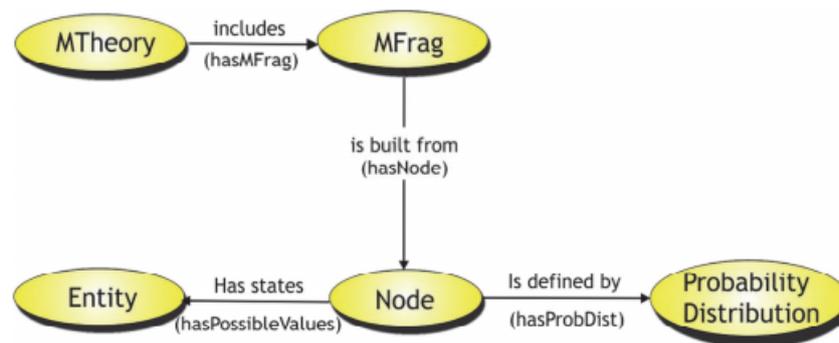


FIGURE 1.2 – Composition d'un MTheory de PR-OWL (Costa, 2005)

De manière générale, la classe MTheory représente un concept du monde, pour lequel l'ontologie va être construite.

Un MTheory est un regroupement de MFrag qui correspond à des sous-classes ou des propriétés de classes destinées à acquérir de nouvelles connaissances. Un MFrag regroupe les

variables aléatoires (Nodes) d'une classe représentées sous forme de nœuds résidents (Resident Node) et de nœuds de contexte (Context Node) où le MTheory en est le réseau.

Chaque nœud a des états qui lui sont propres appelé des entités (Entity);

Les nœuds sont définis par une table de distribution des probabilités conditionnelles (TPC) qui définit la connaissance apriori du modèle.

Les MFrag calculent la probabilité jointe de chacune de ses variables aléatoires tant dis que PR-OWL utilise les informations encapsulées dans les MFrag pour répondre aux requêtes probabilistes (Laskey et al., 2011).

### 1.3.3 Avantages du langage PR-OWL

Il permet de gérer les connaissances incertaines ce qui rend les résultats des ontologies plus réalistes et corrects.

Les principales différences entre OWL et une ontologie probabiliste comme PR-OWL se résument dans le tableau (Fig. 1.3).

<b>OWL</b>	<b>PR-OWL</b>
<i>Standardisé</i>	<i>Non standardisé</i>
<i>Langage déterministe</i>	<i>Langage probabiliste</i>
<i>Ne tient pas compte de l'incertain</i>	<i>Tient compte de l'incertain</i>
<i>Base logique</i>	<i>Réseau bayésien</i>
<i>Compatible : XML, RDF</i>	<i>Compatible : XML, RDF</i>
<i>Formalisme : Sujet, prédicat, objet</i>	<i>MTheory, MFrag</i>

FIGURE 1.3 – Principales différences entre OWL et PR-OWL (Costa, 2005).

### 1.3.4 Méthode UMP-ST

L'UMP-ST est un processus itératif et incrémental, basé sur l'UP, pour concevoir une ontologie probabiliste, tandis que UP sert de point de départ, UMP-ST s'appuie sur et est compatible avec l'ingénierie de l'ontologie et les processus d'élicitation des probabilités.

L'UMP-ST comprend toutes les phases de l'UP, cette méthodologie (Fig. 1.4) est applicable à tout problème nécessitant le développement d'une ontologie probabiliste, dans laquelle PR-OWL a été choisi comme langage ontologique en raison de son traitement complet de l'incertitude (Carvalho et al., 2016).

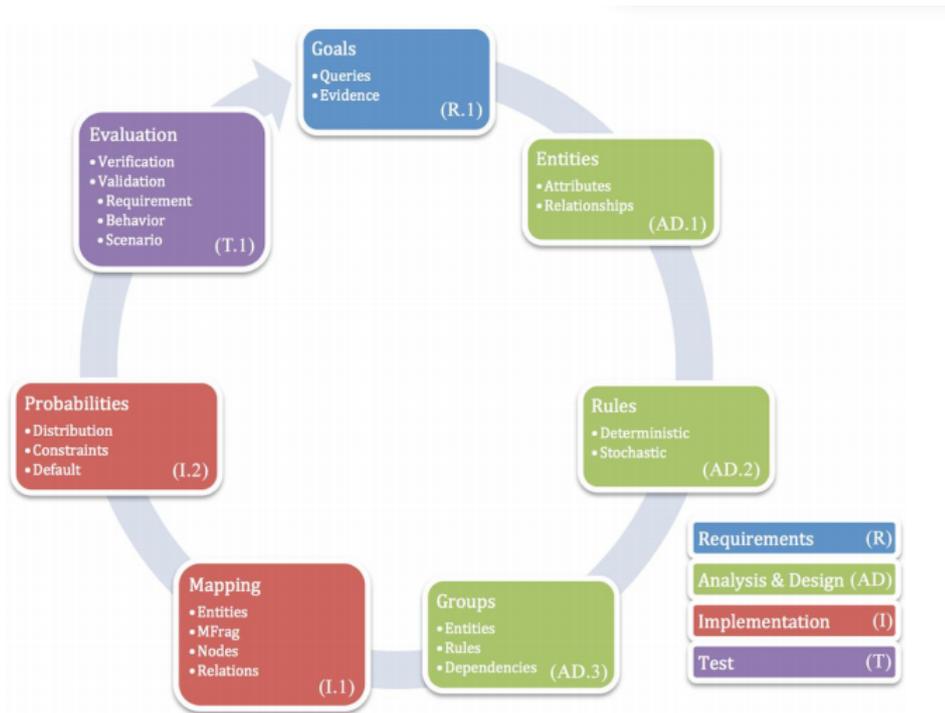


FIGURE 1.4 – Cycle de modélisation de l’ontologie probabiliste (POMC) (Carvalho et al., 2016)

## 1.4 Conclusion

Dans ce chapitre, nous avons vu la définition de l’ontologie ainsi que ses composants, son usage et ses étapes de construction en suivant son langage, nous avons montré que cette notion présente une limite, qui a donné naissance à une deuxième notion qui est l’ontologie probabiliste, nous avons abordé cette dernière durant le court de ce chapitre en mentionnant sa définition et ses langages, et ses avantages en citant la méthode UMP-ST et la différence entre les deux notions.

La suite logique de cette partie sera la technique d’introduction des probabilités dans une ontologie ce qui fera l’élément du prochain chapitre.

## Chapitre 2

---

# Réseaux bayésiens standard et multi-entités

---

## 2.1 Introduction

Après que nous avons vu l'importance de la représentation des informations dans le précédent chapitre, et nous avons vu l'apport que la notion de probabilité rapporte lors de son introduction dans les ontologies, mais une question apparaît dans ce point, comment les probabilités peuvent être introduites et utilisées dans les ontologies.

L'objet de ce chapitre va être la réponse à cette question ou nous allons montrer le modèle des réseaux bayésiens, leur utilité dans la modélisation des probabilités et leur inférence qui va obliger les auteurs à penser à une amélioration concrétisée dans les réseaux bayésien multi-entités.

## 2.2 Réseaux bayésiens

Le réseau bayésien est un système qui représente les informations incertaines avec une manière très lisible et claire.

### 2.2.1 Définition

De manière formelle ([Bouzaieni, 2014](#)), un réseau bayésien est un modèle graphique probabiliste représentant un ensemble de variables aléatoires représentées par des nœuds, liées par des arcs orientés et accompagnées de leurs indépendances conditionnelles. D'une façon formelle, un réseau bayésien est défini par :

- Un graphe acyclique orienté  $G = (V, E)$ , où  $V$  est l'ensemble des nœuds de  $G$ , et  $E$  est l'ensemble des arcs de  $G$ ;
- Un espace probabilisé fini  $(\Omega, Z, p)$ ;
- Un ensemble de variables aléatoires associées aux nœuds du graphe et définies sur  $(\Omega, Z, p)$ ,

telles que :

$$p(v_1, v_2, \dots, v_n) = \prod_{i=1}^n p(v_i | C(v_i))$$

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

**Exemple Simple :** Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison (Ben Mrad, 2015). Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette un coup d'œil sur la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide. Il en déduit qu'il a probablement plu et décide de partir au travail sans vérifier son arroseur automatique. La représentation graphique du modèle causal utilisé par M. Holmes est composé de 4 nœuds : Arroseur (A), Pluie (P), Jardin humide (J), Jardin du voisin humide (W), (Fig. 2.1).

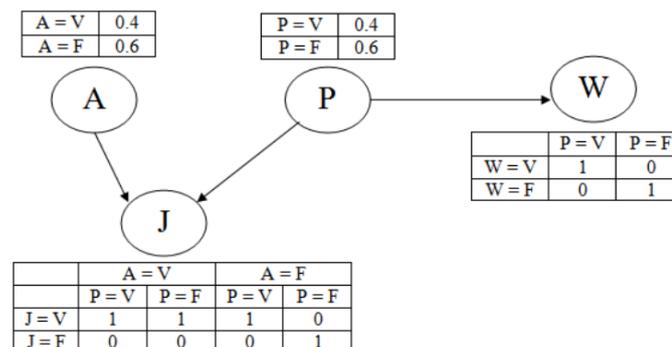


FIGURE 2.1 – Exemple simple d'un réseau bayésien (Ben Mrad, 2015)

## 2.2.2 Raisons d'utilisation des réseaux bayésiens

Les réseaux bayésiens rapportent un plus considérable pour la représentation des informations, vu que ça aide d'une manière cruciales à illustrer des informations abstraites, leurs utilisations apporte (Naïm et al., 1999) :

1. Une acquisition des connaissances : La possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle.
2. Une représentation des connaissances : La représentation graphique d'un réseau bayésien est explicite, intuitive et compréhensible par un non spécialiste, ce qui facilite à la fois la validation du modèle, ses évolutions éventuelles et surtout son utilisation.

3. Une utilisation des connaissances : Le réseau bayésien est polyvalent : on peut se servir du même modèle pour évaluer, prévoir, diagnostiquer, ou optimiser des décisions, ce qui contribue à rentabiliser l'effort de construction du réseau bayésien.
4. Une qualité de l'offre en matière de logiciels : Il existe aujourd'hui de nombreux logiciels pour saisir et traiter des réseaux bayésiens (Matlab, Bayesia, Hugin, Netica, Elvira).

### 2.2.3 Apprentissage de structure et des paramètres

La construction d'un réseau bayésien consiste à trouver une structure ou un graphe et estimer les paramètres (probabilités conditionnelles) (Zaabot, 2012). Cependant devant une très grande base de données, personne ne peut extraire seule la structure adaptée à une telle quantité de données. C'est ici qu'intervient l'apprentissage artificiel. Donc L'apprentissage d'un RB à partir de données se décompose en deux (02) phases :

#### 1. Apprentissage de la structure :

Trouver la meilleure structure par apprentissage automatique à partir de données observées. L'objectif de l'apprentissage de la structure (Ben Mrad, 2015) est de trouver une structure du graphe à partir des données disponibles et qui représente le mieux un problème.

Une solution naïve pour trouver la meilleure structure d'un réseau bayésien, est de parcourir tous les graphes possibles, de leur associer un score, puis de choisir le graphe qui a le score le plus élevé. Cependant, le nombre de structures différentes pour un réseau bayésien de  $n$  nœuds est super exponentiel. Il est donc impossible d'effectuer un parcours exhaustif en un temps raisonnable. Pour cette raison, la plupart des méthodes d'apprentissage de structure utilisent une heuristique de recherche dans l'espace des graphes acycliques dirigés. Les méthodes d'apprentissage de structure qui se basent sur un calcul de score maximisent le score de la structure  $G$  qui décrit le mieux les données  $D$ .

#### 2. Apprentissage des paramètres :

Estimer des paramètres à partir de données observées. Les méthodes d'apprentissage de réseaux bayésiens à partir de données font généralement l'hypothèse de fidélité.

Étant donnée une structure de RB,  $G$  et un ensemble de données  $D$ , le principe de l'algorithme d'apprentissage des paramètres est d'estimer les distributions de probabilités (paramètres).

$$\theta = \theta_i \text{ avec } \theta_i = P(X_i | Pa(X_i))$$

La table de probabilité conditionnelle constituée de :

$$\theta_{ijk} = P(X_i = x_k | Pa(X_i = x_j))$$

Cette estimation peut être effectuée différemment selon que les données disponibles soient complètes ou incomplètes (Nguyen, 2012).

## 2.2.4 Algorithme EM

L'algorithme EM — pour Expectation-Maximisation — est un algorithme itératif du à Dempster, Laird et Rubin (1977).

Il s'agit d'une méthode d'estimation paramétrique s'inscrivant dans le cadre général du maximum de vraisemblance. Lorsque les seules données dont on dispose ne permettent pas l'estimation des paramètres, et/ou que l'expression de la vraisemblance est analytiquement impossible à maximiser, l'algorithme EM peut être une solution. De manière grossière et vague, il vise à fournir un estimateur lorsque cette impossibilité provient de la présence de données cachées ou manquantes ; ou plutôt, lorsque la connaissance de ces données rendrait possible l'estimation des paramètres (Santos, 2015). L'algorithme EM tire son nom du fait qu'à chaque itération il opère deux (02) étapes distinctes :

1. **Phase Expectation** : Souvent désignée comme « l'étape E », elle procède comme son nom le laisse supposer à l'estimation des données inconnues, sachant les données observées et la valeur des paramètres déterminée à l'itération précédente ;
2. **Phase Maximisation** : Ou « étape M », procède donc à la maximisation de la vraisemblance, rendue désormais possible en utilisant l'estimation des données inconnues effectuée à l'étape précédente, et met à jour la valeur du ou des paramètre(s) pour la prochaine itération.

## 2.2.5 Inférence dans les réseaux bayésiens

L'inférence, ou la mise à jour des croyances, consiste à calculer la probabilité a posteriori au niveau de certaines variables en prenant en compte les informations sur d'autres variables appelées observations.

Les réseaux bayésiens permettent de mettre à jour les distributions de probabilités au niveau des variables, on commence par : fixer des observations, c'est à dire affecter des valeurs aux variables d'observation, puis on utilise un algorithme d'inférence pour calculer les probabilités a

posteriori des autres variables.

Pour un réseau bayésien dont l'ensemble des nœuds est  $X = X_1, X_2, \dots, X_n$ , les algorithmes d'inférence permettent de calculer les probabilités a posteriori  $P(X_i|Y)$  d'une variable  $X_i$  sachant les observations  $Y$  où  $Y \subset X$  et  $X_i \notin Y$ .

Aucun algorithme d'inférence n'applique directement le théorème de Bayes car cela conduit à des calculs très coûteux en temps.

Il existe une multitude de travaux dans le domaine de l'inférence dans les réseaux bayésiens. Nous en présentons ici un échantillon représentatif, deux (02) grandes familles d'algorithmes d'inférence se distinguent : Les algorithmes d'inférence exacte et les algorithmes d'inférence approchée (Mrad, 2015).

1. **Inférence exacte** : Parmi les algorithmes d'inférence exacte, on peut citer l'algorithme de propagation de messages, ainsi que les algorithmes de coupe ou de conditionnement qui exploitent la structure du réseau.

Judea Pearl, le père de l'inférence dans les réseaux bayésiens, a développé un algorithme d'envoi de messages (message-passing) (Pearl, 2014).

Le principe consiste, pour chaque nœud du réseau, à mettre à jour la distribution de probabilité de proche en proche avec ses voisins pour prendre en compte les observations sur un ou plusieurs nœuds.

Donc l'inférence consiste à calculer la probabilité a posteriori d'une (ou plusieurs) variable(s) ou requête(s) étant donné un événement observé autrement dit, une affectation de valeurs à un ensemble de variables d'observation.

-X représente la variable de requête ;

-E représente l'ensemble de variable d'observation  $E_1, \dots, E_m$  ;

-e est un événement observé particulier ;

-Y les variables non observées  $Y_1, \dots, Y_l$  (parfois nommées variables cachées).

D'où l'ensemble complet de variables  $X = X \cup E \cup Y$ . Une requête de type demande la distribution de probabilité à posteriori :  $P(X|e)$  (Fung and Del Favero, 1994).

2. **Inférence approchée** : Les algorithmes d'inférence approchée sont plus souvent utilisés dans les réseaux bayésiens complexes, avec plusieurs cycles. Ces méthodes donnent une estimation des probabilités a posteriori et sont non déterministes. Ils sont principalement de deux (02) types (Naïm et al., 2011) :

– Les algorithmes d'inférence approchée du premier type utilisent des méthodes exactes mais opèrent seulement sur une partie du graphe.

– Les algorithmes d’inférence approchée du deuxième type utilisent des méthodes de simulation stochastique. Les plus connus sont les algorithmes de Monte-Carlo ou probabilistic logic sampling ([Henrion, 1988](#)), ces algorithmes donnent une estimation des probabilités moyennant la génération d’échantillons d’instanciation du réseau ([Fung and Del Favero, 1994](#)).

Le problème soulevé par ces méthodes est dû au nombre important d’instanciations nécessaires pour calculer une approximation des probabilités a posteriori.

Pour des réseaux bayésiens de grande taille ou lorsque certaines variables ont des valeurs avec une probabilité proche de 0 ou 1, ces méthodes requièrent un très grand nombre d’échantillons.

## 2.3 Réseaux bayésiens multi entités (MEBN)

Les réseaux bayésiens multi-entités étendent le réseaux bayésien pour atteindre le haut niveau d’expressivité des Logique du premier ordre. Où la capacité de BN pour modéliser l’incertitude est combiné avec l’expressivité de la Logique du premier ordre, dans les connaissances MEBN sont représenté comme une collection de fragments MEBN (MFrag) et un ensemble de MFrag sont organisés en Théories MEBN (MTheories) ([Riali et al., 2017](#)).

### 2.3.1 Fragment MEBN (MFrag)

L’unité de base dans un MEBN est appelée M-Fragment, ces fragments définissent les différents morceaux de réseau bayésien, regroupés ensembles, constituent une MTheory.

Un MFrag est constitué d’un ensemble de nœuds groupés en trois catégories : les nœuds d’entrées, les nœuds résidents et les nœuds contextuels.

- **Les nœuds résidents** : Ce sont les variables aléatoires réelles qui forment le cœur du sujet d’un MFrag. La logique MEBN exige que la distribution probabiliste locale de chaque nœud résident doit être unique et explicitement défini dans son MFrag original. Les valeurs possibles d’un nœud résident peuvent être une entité existante ([Laskey et al., 2010](#)).
- **Les nœuds contextuels** : Ce sont des variables aléatoires booléennes représentant des conditions qui doivent être satisfaites pour faire une distribution dans un MFrag valide. Les nœuds de contexte peuvent représenter plusieurs types de modèles d’incertitude sophistiqués, tels que l’incertitude sur les relations entre les entités. Si on peut inférer de la base de connaissances (c’est à dire l’ontologie) que la valeur d’un nœud de contexte est vraie, la distribution de probabilité de l’MFrag sera appliquée dans le modèle d’inférence. Si cette valeur est fausse, une distribution par défaut sera utilisée. Si la valeur est inconnue, le nœud de contexte devient pratiquement un parent de tous les nœuds résidents dans le même MFrag ([Laskey et al., 2010](#)).

- **Les nœuds d'entrées :** Ces nœuds sont essentiellement des «pointeurs» faisant référence aux nœuds résidents d'un autre MFrag.

Les nœuds d'entrée fournissent également un mécanisme permettant la réutilisation des nœuds résidents entre les MFrag. Les nœuds d'entrée influencent la distribution de probabilité des nœuds résidents qui sont leurs enfants dans un MFrag donné, mais leurs propres distributions sont définies ailleurs (c'est à dire dans leurs propres MFrag d'origine). Dans une MTheory complète, chaque nœud d'entrée doit pointer vers un nœud résident dans un MFrag ([Laskey et al., 2010](#)).

### 2.3.2 Théorie MEBN (MTheory)

Un MTheory est un ensemble de MFrag qui satisfait collectivement des contraintes de cohérence assurant l'existence d'une distribution unique de probabilité conjointe sur les instances des variables aléatoires représentées dans chacune des MFrag dans l'ensemble ([Carvalho et al., 2016](#)).

### 2.3.3 Inférence dans les réseaux bayésiens multi-entités

L'inférence MEBN commence lorsqu'une requête est posée pour évaluer le degré de croyance en une variable aléatoire cible étant donné un ensemble de variables aléatoires de preuve. Nous commençons par un Générative MTheory, ajoutez un ensemble de MFrag de recherche représentant des informations spécifiques au problème et spécifiez les nœuds cibles pour notre requête.

La première étape de l'inférence MEBN consiste à construire un réseau bayésien spécifique à la situation (SSBN), qui est un réseau bayésien construit en créant et en combinant des instances des MFrag dans le Générative MTheory.

Lorsque chaque MFrag est instancié, des instances de ses variables aléatoires sont créées pour représenter les informations de base connues, les évidences observées et questions d'intérêt pour le décideur.

S'il y a des variables aléatoires avec des distributions non définies, alors l'algorithme procède en instanciant leur origine respective MFrag. Le processus de récupération et d'instanciation de MFrag se poursuit jusqu'à ce qu'il y ait aucune variable aléatoire restante ayant des distributions non définies ou valeurs inconnues. Un SSBN peut contenir n'importe quel nombre d'instances de chaque MFrag, selon le nombre d'entités et leurs interrelations, ensuite, l'algorithme d'inférence du réseau bayésien standard est appliqué.

Enfin, la réponse à la requête est obtenue en inspectant les probabilités postérieures des nœuds cibles ([Costa et al., 2005](#)).

## 2.4 Conclusion

Pour couronner ce chapitre ou nous avons abordé au début la définition des réseaux bayésiens avec leurs apports d'utilisation ; soutenu par la démonstration des différentes techniques d'apprentissage, passant par l'inférence de ces réseaux ; ce dernier point était la modalité pour laquelle une extension a été créée nommée les réseaux bayésiens multi-entités, que nous avons évoqué dans la suite en expliquant leurs composants et leur inférence. La continuité de cet écrit passera par la présentation dans le prochain chapitre de l'environnement de la mise en œuvre de ce travail.

## Chapitre 3

---

# Présentation des Réseaux d'AEP

---

### 3.1 Introduction

Nous savons tous que la vie humaine que ce soit dans un agglomération urbaine ou rurale, nécessite un groupe de commodités afin d'assurer la convivialité des personnes, dont la disponibilité et l'alimentation de l'eau potable dans les domiciles et aux endroits où l'utilisation de cette eau est indispensable, est pour assurer cette alimentation, il faut transmettre cette eau depuis les points de réserve jusqu'aux points de l'utilisation, cette activité est assurée par l'utilisation des AEP et nécessite des connaissances dans le domaine hydraulique et des conduites.

Ce qui fera le sujet de ce chapitre, où nous allons détailler la définition du système d'Alimentation en Eau Potable (AEP), ses caractéristiques, les différentes classifications possibles, sa vulnérabilité aux défaillances et les facteurs de risques qui peuvent les causer et au final les méthodes existantes de détection.

### 3.2 Définition d'AEP

Les réseaux d'AEP conditionnent l'activité et le développement des agglomérations. Afin d'appréhender leur fonctionnement l'étude des différentes composantes d'un réseau d'AEP est indispensable. Un réseau d'AEP a pour fonction principale de desservir en eau potable un ensemble de points tels que : Compteurs d'abonnés, Bouches de lavage, Poteaux d'incendie (Dupont, 1979).

### 3.3 Caractéristiques des AEP

1. **Conduites** : Les conduites permettent l'acheminement de l'eau d'un point à un autre point du réseau. Une conduite est un segment de tuyau ou canalisation délimitée par deux points de consommation d'eau appelés nœuds. Chaque conduite est caractérisée par : Un nœud

initial et un nœud final. Une longueur donnée  $L$ . Un diamètre  $d$ . Un coefficient de rugosité  $C$  traduisant la perte de charge. Un état : ouvert, fermé. L'écoulement de l'eau s'effectue du nœud disposant de la pression la plus élevée vers le nœud dont la pression est plus faible. La rugosité traduit la résistance de la conduite à l'écoulement de l'eau. Les parois internes des conduites au contact de l'eau créent un phénomène de friction qui s'accompagne de perte d'énergie due au frottement créant ainsi une perte de charge linéaire (Houichi, 2012).

2. **Pompes** : Une pompe est un dispositif permettant de fournir de l'énergie au liquide. Le fonctionnement de la pompe est relié généralement à un réservoir. Le démarrage et l'arrêt de la pompe sont fonction du niveau du réservoir ou de plages horaires spécifiques. Une pompe peut être caractérisée soit par une puissance constante (énergie), fournie à l'eau au cours du temps indépendamment du débit et de la hauteur de refoulement, ou par une courbe caractéristique qui décrit la relation entre la hauteur de refoulement et le débit fournie à l'aide d'une fonction  $H=f(Q)$  La pompe devra compenser les dissipations d'énergie dans les conduites tant d'aspiration que de refoulement ; dénommées pertes de charge elles sont couramment évaluées en hauteur d'eau exprimée en mètre.
3. **Puissance de la pompe** : La puissance de la pompe est égale au travail effectué pendant l'unité de temps pour élever le débit de charge en mètre.
4. **Vannes** : Certaines conduites de longueur fictive comporte des vannes qui permettent de limiter la pression ou le débit en des points précis du réseau. Les vannes sont caractérisées par : les nœuds d'entrée et de sortie ; le diamètre ; le diamètre de la consigne de fonctionnement et l'état de la vanne ; coefficient de perte de charge singulière. Une vanne peut être ouverte ou fermée, le fonctionnement de la vanne est fonction de consignes relatives à un nœud indexé généralement en rapport avec la pression. L'écoulement de l'eau à travers la vanne s'accompagne d'une perte d'énergie exprimée par la perte de charge singulière.
5. **Nœuds** : Les nœuds représentent des points de jonction entre les conduites. Ils correspondent à des points d'entrée ou de sortie d'eau. Il existe deux (02) catégories de nœuds :
  - **Nœuds à débit fixe** : Ces nœuds se caractérisent par une cote au sol connue et un débit connu (demande), l'inconnue est la pression au nœud qui doit être calculée. Ils correspondent à des points de consommation dans le réseau. Ces nœuds peuvent décrire la consommation d'un ou de plusieurs abonnés de même type. Nous distinguons entre les abonnés selon le type de consommation : domestique, industrielle, administration. La consommation au nœud exprimée par la demande peut être constante ou variable.

- **Nœuds à charge fixe** : Ce sont des nœuds où la charge est fixée ou dont la cote piézométrique de l'eau est connue. Il peut s'agir d'un réservoir dont le niveau d'eau varie en fonction du temps au sol ou sur tour, d'un poteau à incendie ou d'une bêche de pompage dont le niveau reste inchangé. Pour ces nœuds le débit doit être calculé. Selon la nature des abonnés, la demande est décrite par une courbe de consommation.

## 3.4 Classifications des réseaux d'AEP

On distingue plusieurs types de réseaux, à savoir :  
Les réseaux ramifiés, maillés, étagés et les réseaux à alimentation distinctes.

### 3.4.1 Réseaux ramifiés

Cette classification est basée sur l'extension de la conduite jusqu'à l'arrivée chez le client, une série d'extensions extraites à partir de la conduite principale, toute extension est similaire ou petite par rapport sa précédente, cette structure est très économique vu que l'eau circule dans un seul sens directement vers le compteur d'utilisation mais d'autre part elle présente une grave défaillance, en cas de panne sur la conduite principale, toutes les branches extraites seront privées de l'alimentation instantanément.

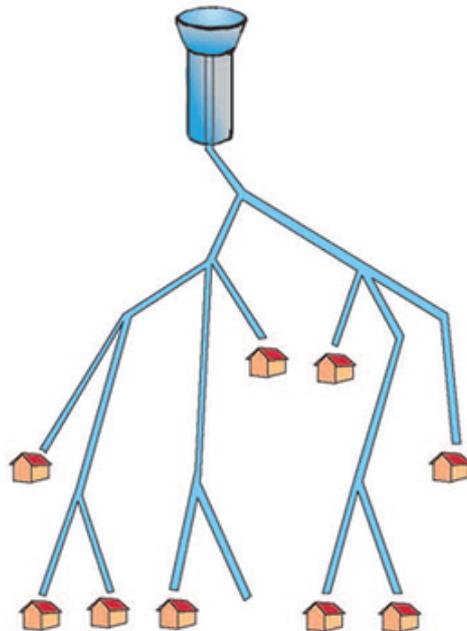


FIGURE 3.1 – Réseau d'AEP "ramifié" (ramifié, 2020)

### 3.4.2 Réseaux maillés

Cette classification est établit sur le principe des boucles, le réseau est composé de conduites étendues de la conduite principale suivant des contours fermés formant ainsi plusieurs mailles (boucles), l'installation de ce type est plus coûteuse que les réseaux ramifiés mais propose une alimentation en retour, ce qui permet la continuité de la disponibilité de l'eau en cas de pannes vu que l'eau peut atteindre un même point par plusieurs chemins. Cette structure est utilisée généralement dans les zones urbanisées.

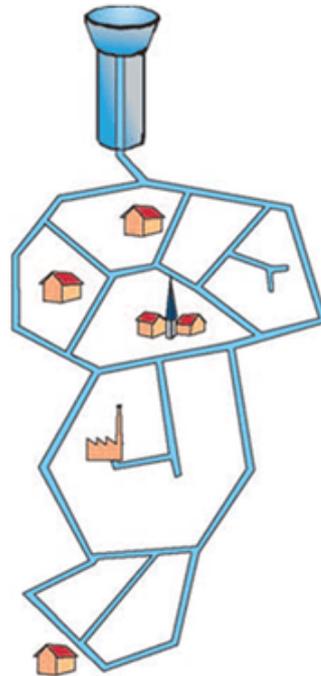


FIGURE 3.2 – Réseau d'AEP"en maillés" ([en mailles, 2020](#))

### 3.4.3 Réseaux étagés

Ce type de réseaux se distingue par l'importante différence de niveaux tout au long du réseau, ce qui assure une bonne alimentation de l'eau mais cause une forte pression sur les points de distribution les plus bas, pour cela, des réservoirs intermédiaires pour régulariser la pression dans le réseau.

### 3.4.4 Réseaux à alimentation distinctes

Cette classification est faite pour les réseaux qui assurent deux (02) types de distribution, le premier type est la distribution de l'eau potable pour les foyers domestiques, le deuxième type est la distribution de l'eau non potable destinée à l'utilisation industrielle et l'usage en agriculture,

cette structure n'est utilisée que dans les installations importantes ou les deux cas d'utilisation de l'eau sont présents dans la même structure.

Leur classification se fait en se basant sur la topographie du terrain (terrain accidenté, différence d'altitude importante) et l'occupation du sol. Ces réseaux peuvent être alimentés à partir d'un ou plusieurs réservoirs. Avec le réseau étagé, il est possible, de constituer des réseaux indépendants avec une pression limitée aux environs de 40 mètres d'eau (Dupont, 1979).

Les réseaux à alimentation distinctes ne se justifient que dans les installations extrêmement importantes.

## 3.5 Risques de défaillance sur les AEP

Les AEP comme tout composant technique utilisable est sous le risque de subir a des défaillances qui seront le résultat de un ou plusieurs facteurs ayant une relation avec l'environnement et l'utilisation de l'AEP, et pour cela, nous avons recensé les différents types de défaillances possibles :

### 3.5.1 Fuites

Ce sont de petites pertes d'eau comme montre la (Fig. 3.3) peu perturbantes du fonctionnement normal du réseau AEP, elles apparaissent fréquemment dans les vannes, les raccordements, les joints, les points de jonction, ou dans le corps même de la conduite.



FIGURE 3.3 – Une fuite (leaking2020).

### 3.5.2 Casses

Comme l'illustre la (Fig. 3.4), appelées aussi ruptures, les casses sont des dégradations importantes des conduites du réseau AEP, qui causent un arrêt temporaire de l'alimentation, la

durée dépend de la gravité de la casse et des moyens de réparation.

À noter qu'une rupture est considérée comme une fuite, mais une fuite n'est pas obligatoirement une rupture.



FIGURE 3.4 – Une casse (iStock, 2016).

## 3.6 Facteurs générateurs des risques

Comme tout composant physique utilisé sur terrain réel, les conduites subissent plusieurs facteurs, naturelles ou artificielles, pouvant consister des risques qui vont conduire à des défaillances (fuite ou casse), parmi ces facteurs nous citons les facteurs liés directement à la nature de la conduite, comme l'âge, la profondeur, la linéaire, la corrosion, le matériau de fabrication, le diamètre de la conduite, il existe aussi des facteurs externes comme le type du sol ou la conduite est déposée, le trafic routier entourant la conduite, la pression interne du liquide conduit à l'intérieur, ce dernier facteur peut causer un autre facteur appelé le coup de bélier.

### 1. Âge de la conduite

Le risque de fuites augmente parallèlement avec l'augmentation de l'âge de la conduite, autrement dit, plus la conduite est vieille plus le risque est grand à cause du vieillissement et l'usage des composants de cette conduite avec le temps.

### 2. Corrosion de la conduite

L'effet de corrosion dépend du matériau avec qui la conduite est faite ainsi qu'aux facteurs externes, de son côté, il influence directement sur la conduite en détruisant lentement sa structure ce qui crée des failles qui se développent en fuites.

### 3. Diamètre de la conduite

Vu sa forme circulaire, toute conduite a un diamètre, cette propriété peut causer des nuisances à la conduite.

#### 4. **Linéaire de la conduite**

Autrement dit la longueur de la conduite, la linéaire peut causer des fuites à cause de l'augmentation de la surface de contact de la conduite avec son environnement externe ainsi qu'à l'eau conduit à l'intérieur.

#### 5. **Matériaux de la conduite**

Nous distinguons quatre (04) types de matériaux utilisés pour la construction des conduites classés du plus au moins vulnérable Acier, BPAT, FD, FG, et PEHD.

#### 6. **Profondeur de la conduite**

La profondeur est un facteur qui crée la différence dans l'influence des autres, c'est-à-dire que si la profondeur de la conduite par rapport à la surface est grande, l'effet des facteurs liés à cette surface est moins et vice-versa.

#### 7. **Type du sol**

L'influence du type de sol se concrétise dans l'effet que les composants du sol apportent sur la conduite, nous pouvons citer l'exemple d'un sol acide, une propriété qui peut être naturelle ou artificielle pour la fertilisation des sols, ce type influencera directement sur la structure de la conduite et va l'user ce qui accélérera l'évolution d'une fuite.

### 3.7 **Méthodes de détection existantes**

#### 3.7.1 **Mesure des vibrations**

Basée sur le phénomène vibro-acoustique, cette méthode s'appuie sur l'utilisation des outils de surveillance des vibrations dans les réseaux d'AEP pour la détection d'éclatement de fuites, l'objectif de cette méthode est de diminuer les coûts de maintenance des AEP, pour ce fait ; l'équipe de recherche du Mr Alberto Martini qui fait partie de la société italienne Hera SpA, a utilisé des accéléromètres placés sur les conduites juste avant les compteurs d'eau des habitants pour surveiller les ondes des vibrations émises par les conduites, les résultats récupérés de la surveillance passent par un traitement effectué par un algorithme de distinction entre les vibrations normales et celles les vibrations émises par les fuites, la méthode a permis la détection et la localisation des fuites dès leurs début tout en étant de petites fuites, ce qui a permis d'effectuer des réparations précoces et d'éviter les pertes des grandes casses ([Alberto Martini, 2015](#)).

#### 3.7.2 **Utilisation des réseaux bayésiens de croyance (BBN)**

Cette méthode a été développée pour la réalisation d'une plate-forme efficace pour la surveillance du réseau d'AEP et la prise de décision en matière de gestion en temps réel, la raison de ce travail était pour vu de réduire le taux de contamination des maladies liées à l'eau ; des chercheurs à leur tête Mr Royce A. Francis, ont exploité un dataset contenant l'historique des

fuites de la division Mid-Atlantic durant la période 2010-2011, cette division inclut trois (03) états, New York, New Jersey et la Pennsylvanie, nous remarquons que ces dernières sont des états très peuplés ce qui donne une charge de plus au dataset et un réseau d'AEP de haute complexité vu le grand nombres de foyers liés par des AEP à destination de réaliser un système de prédiction des fuites, le résultat de ce travail était une plate-forme dynamique avec un réseau bayésien de croyance capable de prédire l'existence d'une fuite selon les valeurs des facteurs introduits ce qui leurs a permis de réduire le taux de risque de contamination ([Royce A. Francis, 2015](#)).

### **3.7.3 Traitement des valeurs de pression d'eau**

Basée sur la technique d'apprentissage SVM (Support Vector Machines) ou en français "Machines vectorielles de soutien", cette méthode a été réalisé pour but de localiser et prédire les fuites, et en ajoutant la taille probable de ces fuites pour permettre aux décideurs de juger la gravité de la situation et décider la nécessité et la taille de l'intervention, le développement de cette méthode a nécessité l'utilisation de l'historique des fuites de la région sud-est de la ville de Melbourne en Australie, cet historique consisté une base d'apprentissage qui contient une centaine d'exemples de fuites détectées précédemment pour le système en lui permettant d'apprendre la liaison entre les deux facteurs qui sont la pression et le débit d'eau de la conduite, et la présence d'une fuite, après l'accomplissement de l'entraînement et l'apprentissage, Mr John Mashford et son équipe sont arrivés à réaliser un système prédicteur de fuites selon la pression et le flux d'eau dans les réseaux d'AEP ([John Mashford and Burn, 2009](#)).

### **3.7.4 Technique informatisée d'analyse de la pression d'eau**

Fondée sur le principe de comparaison, cette technique développée Mme A.Ejah,U.S et son équipe consiste à l'utilisation de deux (02) techniques d'apprentissages, la technique du SVM et la technique ELM (Extreme Learning Machine) ou en français "Machine d'apprentissage extrême" et ensuite la comparaison entre les valeurs obtenues selon l'ampleur de l'erreur calculée selon NRMSE (Normalize Root Mean Square Error), plus la valeur NRMSE est basse, plus les résultats de prédiction sont exactes ; pour but de prévoir la magnitude autrement dit la taille de la fuite et sa localisation, pour ce fait, des données de pressions des conduites de la région Khayangan Park Makassar en Indonésie ont été utilisé pour l'apprentissage des deux structures, après la fin du processus d'apprentissage, la comparaison en NRMSE a montré que la technique SVM a donné une taille de fuite de 0,1508 à la position 0,4258, alors que la technique ELM a donné une taille de fuite de 0,1017 ce qui est plus bas que SVM à la position 0,2063 qui est à son tour plus bas aussi que le résultat de sa précédente, ce qui montre que la technique ELM est plus fiable pour la prédiction, la détection et la localisation des fuites ([A.Ejah Umraeni Salam1\\* and Maricar2, 2015](#)).

Méthode	Auteur	Région	Type	Technique utilisée	Données de test	Avantages	Inconvénients
Mesure des vibrations	Alberto MARTINI, Marco TRONCOSSI, Alessandro RIVOLA - 2015	Bologne, Italie	Détection, Localisation	Surveillance des vibrations	Vibrations collectées par les accéléromètres	Prévention de l'aggravation des fuites	Détection des fuites qu'après leur apparition.
Utilisation des réseaux bayésiens de croyance (BBN)	Royce A. FRANCIS, Seth D. GUIKEMA, Lucas HENNEMAN - 2015	Division Mid-Atlantic, États-Unis	Prédiction	Réseaux bayésiens de croyance (BBN)	Dataset des fuites de la division Mid-Atlantic	Maintenance de la conduite avant l'apparition de la fuite	La méthode ne prend pas en compte tous les facteurs.
Traitement des valeurs de pression d'eau pour la détection des fuites	John MASHFORD, Dharmika De SILVA, Donavan MARNEY, Stewart BURN - 2009	Sud-est de Melbourne, Australie	Prédiction, Localisation	SVM	Historique des fuites du sud-est de Melbourne	Prise de décisions préventives avant l'apparition de la fuite	La méthode ne prend pas en compte tous les facteurs.
Technique informatisée pour l'analyse de la pression de l'eau	Ejah A. UMRAENISALAM, Muh. TOLA, Mary SELINTUNG, Farouk MARICAR - 2015	Kayangan Park Makassar, Indonésie	Prédiction, Détection, Localisation	SVM & ELM	Données de pression des conduites de la région Kayangan Park Makassar	Prédiction précise de la fuite	La méthode ne prend pas en compte tous les facteurs.

TABLE 3.1 – Tableau de comparaison des méthodes de détection existantes

## 3.8 Comparaison des méthodes

Après que nous avons étudié les méthodes citées précédemment, nous avons trouvé que la méthode informatisée pour l'analyse de la pression de l'eau assure la prédiction, la détection et la localisation des fuites, alors que la méthode de mesure des vibrations et la méthode de traitement des valeurs de pression d'eau assurent toutes les deux (02) la prédiction et la localisation des fuites tandis que la méthode basée sur l'utilisation des réseaux bayésiens de croyance (BBN) ne fait que la prédiction, d'autre part, la technique informatisée pour l'analyse de la pression de l'eau est considérée comme une amélioration de la méthode de traitement des valeurs de pression en ajoutant l'utilisation du ELM ce qui donne une meilleure précision pour elle cependant l'utilisation des réseaux bayésiens de croyance (BBN) a mis en place une technologie différente mais toujours dans le même concept d'apprentissage alors que la technique de mesure de vibration est considérée comme une méthode acoustique vu son principe basé sur les données des vibrations.

## 3.9 Analyse horizontale

En comparant entre les méthodes proposées, le résultat est comme suit :

- La première méthode est détection basée sur la mesure de vibration est utilisée pour la localisation et détection des fuites, développée par les messieurs Alberto Martini, Marco Troncosi et Alessandro Rivola en 2015, elle consiste à suivre les vibrations apparues au niveau des conduites en utilisant des accéléromètres, ce qui rend la localisation de la fuite facile, d'autre part, cette méthode joue un rôle majeur dans la réduction des coûts de maintenance vu qu'elle permet d'éviter l'aggravation des fuites mais elle ne les traite pas qu'après leur apparition dans le réseau ;
- La deuxième méthode est bien la détection basée sur l'utilisation des réseaux bayésiens de croyance (BBN), contrairement à sa précédente, réalisée par l'équipe de Mr Royce A. Francis qui se compose des Mrs Seth D. Guikema et Lucas Henneman et publiée en 2015, cette technique permet seulement la prédiction des fuites grâce aux données du système de distribution d'eau potable et au réseaux bayésiens de croyance (BBN), cette technique offre aux équipes techniques la possibilité de l'anticipation des fuites mais elle fait face à un grand inconvénient qui est la non prise totale des facteurs ;
- La troisième méthode est le traitement des valeurs de pression d'eau pour la prédiction et la localisation des fuites, proposée par les Mrs John Mashford, Dhammika De Silva, Donavan Marney et Stewart Burn en 2009, cette méthode permet la prédiction et la localisation des fuites par l'utilisation des Support Vector Machine (SVM), cette méthode est confrontée à son tour au problème de non prise totale des facteurs ;

- La quatrième et dernière méthode est la technique informatisée d'analyse de la pression d'eau, développée par l'équipe de Mme Ejah Umraeni Salam qui est composée de Mme Mary Selintung et les Mrs Muh Tola et Farouk Maricar en 2015, cette méthode permet de la prédiction, la détection et la localisation des fuites en comparant les résultats des deux (02) techniques utilisées, les Support Vector Machine (SVM) et les Extreme Machine Learning (ELM), cette méthode est aussi pas prenante de l'ensemble total des facteurs.

### **3.10 Conclusion**

Ce chapitre a fait objet d'une représentation détaillée de l'environnement de travail de ce thème, ou nous avons abordé la définition des réseaux d'AEP avec leurs caractéristiques et leurs classifications, suivi par les risques de défaillance avec leurs facteurs et en clôturant par les méthodes existantes pour la prévention de ces risques. Après le passage sur l'ensemble du thème durant les trois (03) premiers chapitres, nous allons aborder la partie technique du projet à partir du prochain chapitre sur la conception de notre système.



## *Chapitre 4*

---

# **Conception du système de prédiction des fuites**

---

## **4.1 Introduction**

Après la compréhension du domaine de travail, les composants et les facteurs de fuite, nous sommes arrivés à l'étape de conception où nous allons concevoir et modéliser le savoir que nous avons acquis afin de préparer la base de la programmation de notre système.

Nous allons aborder dans ce chapitre la raison pour laquelle ce projet a été établi, suivi par la présentation de la société SEAAL et leur contribution dans ce travail, nous allons expliquer le cycle de conception commençant par la préparation du dataset, après le scénario global du système et la création de l'ontologie probabiliste.

## **4.2 Problématique**

Les casses et les fuites au niveau sont devenues un véritable dilemme pour la société SEAAL ainsi que les autres sociétés similaires, chose qui l'a poussé à penser à une solution optimale à ce fléau pour éviter aux utilisateurs les problèmes sanitaires et écologiques et aussi pour épargner la société des pertes économiques importantes.

## **4.3 Motivation du travail**

Les fuites constituent un problème constant, d'une part pour l'utilisateur qui peut-être un habitant, un commerçant ou bien un industriel, et d'une autre part du fournisseur qui est l'entreprise responsable de la distribution de l'eau, ce qui rend que l'existence d'une solution à ce problème indispensable est nécessaire pour les deux cotés afin de leur faciliter leur travail et aussi leur permettre les meilleures conditions de vie, les fuites infectent négativement aussi sur

l'écologie vu que les eaux perdues causent la pollution de son entourage et aussi, elles causeront des flaques d'eaux qui sont synonymes de milieu de culture pour les maladies et les insectes.

Notre système à comme objectif, l'aide des entreprises chargées des ressources en eau pour la détection et la prédiction des fuites afin de permettre la diminution et l'élimination des pertes causées par les fuites.

## 4.4 SEAAL

La Société des Eaux et de l'Assainissement d'Alger « SEAAL », est une Société Publique par Actions, détenue à 70 par l'Algérienne Des Eaux « ADE » et à 30 par l'Office National de l'Assainissement « ONA » (SEAAL, 2020).

La principale mission de SEAAL est de produire et de desservir en eau potable, puis de collecter et traiter les eaux usées sur le périmètre des wilayas d'Alger et de Tipasa. Elle dessert ainsi 3,8 millions d'habitants soit environ 10 de la population Nationale (SEAAL, 2020).

SEAAL assure le service auprès de 761 824 clients (ménages, administrations, commerces, industriels et sites touristiques) et le nombre de ses salariés s'élève à 6.293 (SEAAL, 2020).

## 4.5 Dataset

Le dataset que nous allons utilisé nous a été fourni par la société SEAAL, pour but de faire apprendre au système les différents cas de présence ou absence des fuites selon la combinaison des valeurs des facteurs de risques, et pour cela, ce dataset contient l'historique des fuites d'un ensemble de conduites avec les valeurs de sept (07) facteurs, que nous avons cité précédemment, le diamètre, la longueur, la corrosivité, l'âge, le type du matériau de la conduite, le trafic routier et le type du sol, clôturé avec une colonne sur la présence d'une fuite.

## 4.6 Scénario global

Notre conception consiste à faire les prédictions des fuites dans les réseaux d'AEP en utilisant une ontologie probabiliste, qui est à la base une ontologie classique modélisée par la méthodologie UMP-ST.

Pour pouvoir utiliser notre système, il faut passer par deux (02) grandes étapes :

1. La première consiste à la construction de notre ontologie probabiliste avec PR-OWL, qui se base sur la construction du MEBN, ce dernier nécessite l'apprentissage de la structure pour la création des nœuds qui composent le MEBN, et l'apprentissage de paramètres pour remplir les LPD de l'ontologie probabiliste.

2. La suivante étape a pour but de permettre a l'expert du domaine d'envoyer des requêtes sur les conduites qui souhaite diagnostiquer pour lesquelles un SSBN (Situation-Spécifique Bayesian Networks) est géré en sorte qu'il correspond a la requête demandée. Le résultat de l'inférence est une probabilité qui indique la présence ou l'absence d'une fuite.

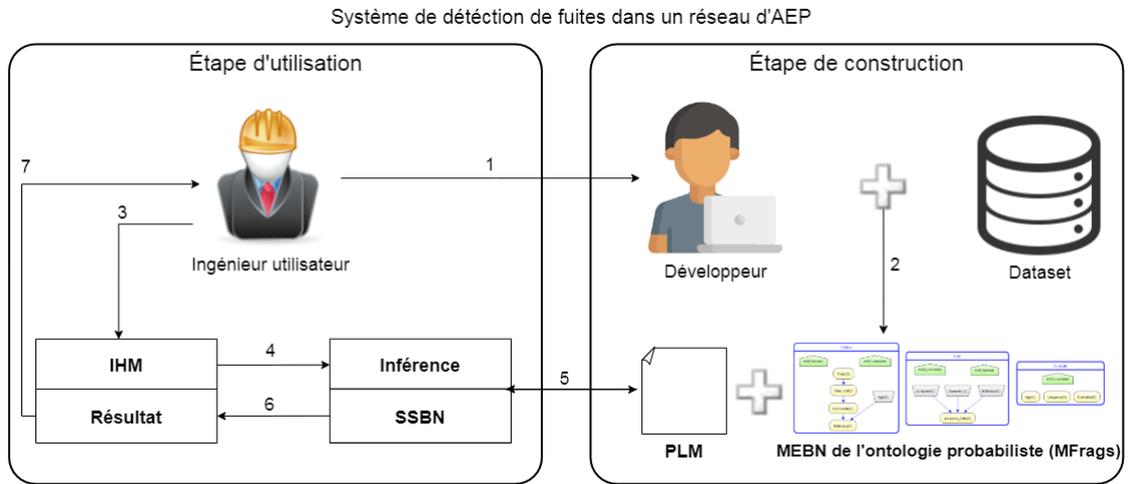


FIGURE 4.1 – Schéma global du système.

Arc	Description
1	L'ingénieur explique son savoir et son expertise concernant les fuites et l'influence des facteurs au développeur d'une manière abstraite.
2	Le développeur projette le savoir transmis sur un dataset et le traduit en un réseau bayésien multi-entités en suivant la méthodologie UMP-ST exprimant le savoir acquis d'une manière explicite et compréhensible par la machine.
3	L'ingénieur ouvre l'interface et introduit les valeurs des facteurs selon son choix juste en sélectionnant des dropdowns et clique sur OK pour lancer l'opération.
4	Le système lit les valeurs introduites des facteurs et lance une inférence basée sur ces valeurs.
5	Le système fait appel à l'ontologie probabiliste et le MEBN pour tester les valeurs selon les règles de probabilités prédéfinies, en retour un SSBN est généré effectuant un calcul des probabilités de la fuite selon les valeurs déjà introduites.
6	Le système affiche le résultat de l'opération qui est la probabilité de la fuite calculée selon le SSBN.
7	L'ingénieur observe un résultat numérique bien illustré sous forme de deux (02) nombres définissant les probabilités de présence et absence de la fuite successivement et l'interprète facilement selon son savoir.

TABLE 4.1 – Tableau explicatif du schéma global du système

## 4.7 Création de l'ontologie probabiliste

L'objectif de la création de l'ontologie probabiliste est de faire la modélisation et la représentation des connaissances incertaines, afin de prédire la présence des fuites dans les réseaux d'AEP, pour but d'aider le gestionnaire du réseau à prendre les bonnes décisions.

### 4.7.1 Exigences

Le cycle de modélisation de l'ontologie probabiliste (POMC) commence par la discipline des exigences. La discipline des exigences définit les objectifs qui doivent être atteints en représentant et en raisonnant avec une représentation calculable de la sémantique du domaine. Pour cette discipline, il est important de définir les questions auxquelles le modèle est censé répondre, c'est-à-dire les questions à poser le système en cours de conception. Pour chaque question, un ensemble d'éléments d'information qui pourraient aider à répondre à la question (évidence) doit être défini (Carvalho et al., 2016).

### 4.7.2 Besoins (Goals)

Les besoins (Goals), veut dire les différents types de questions ou requêtes que l'utilisateur veut que le système réponde via le raisonnement de l'ontologie probabiliste (OP) (Carvalho et al., 2016). Voici l'ensemble des besoins, des requêtes et des preuves (évidences) concernant notre domaine de raisonnement pour qu'une conduite connaisse une fuite :

- Besoin (Goal) :
  - Identifier si la conduite a une fuite?
- Requête (Query) :
  - La conduite a-t-elle les facteurs de risque de fuite ?
- Évidence :
  - Le diamètre de la conduite.
  - La longueur de la conduite.
  - L'âge de la conduite.
  - Le trafic routier sur la conduite.
  - La corrosivité de la conduite.
  - Le matériau de la conduite.
  - Le type du sol.

### 4.7.3 Analyse et Conception

Cette étape décrit les classes d'entités, leurs attributs, les relations entre eux et les règles qui sont applicables dans notre domaine. Cette étape est indépendante du langage utilisé pour implémenter le modèle (Carvalho et al., 2016).

1. **Les entités** : Description des classes d'entités, leurs attributs, les relations entre eux, pour que nous puissions établir les règles qui sont applicables dans notre domaine.
  - **Les entités et leurs attributs** : les entités : ce sont les classes de l'ontologie :
  - Conduite (Table 4.2)
  - Facteur (Table 4.3)
  - Fuite (Table 4.4)

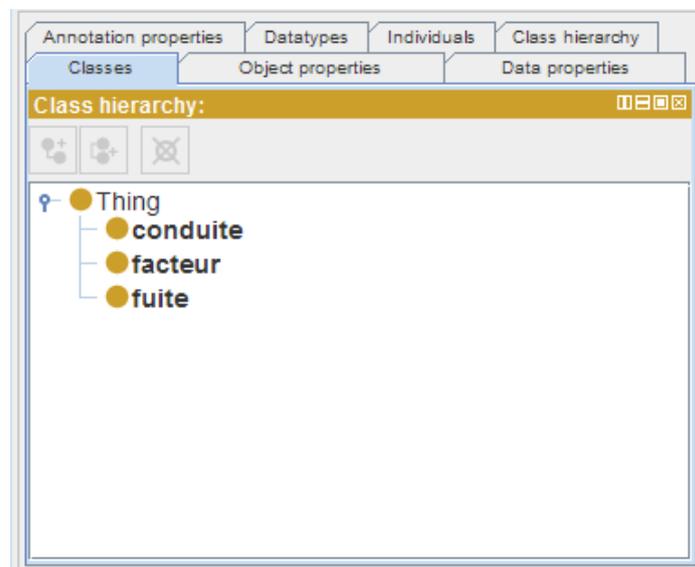


FIGURE 4.2 – Les entités.

Propriété	Domain	Range
IDC	Conduite	String
Âge	Conduite	Double
Diamètre	Conduite	Double
Longueur	Conduite	Double
afacteur	Conduite	Facteur
afuite	Conduite	Fuite

TABLE 4.2 – Propriétés de l’entité Conduite

Propriété	Domain	Range
Corrosivité	Facteur	String
Matériaux	Facteur	String
Type de sol	Facteur	String
Trafic routier	Facteur	String

TABLE 4.3 – Propriétés de l’entité Facteur

Propriété	Domain	Range
ID	Fuite	Int
Présence (présence d’une fuite)	Fuite	Boolean

TABLE 4.4 – Propriétés de l’entité Fuite

2. **Les règles** : la mise en œuvre d’une OP nécessite de spécifier des probabilités numériques, ces valeurs peuvent être obtenue auprès d’un expert du domaine ou tirées de l’observation. dans notre travail les valeurs de probabilité sont tirées du dataset en trois (03) étapes : Afin d’utiliser le réseau bayésien multi-entité (MEBN) qui est a la base du PR-OWL, nous devons avoir des données de type nominal, notre dataset contient des valeurs manquantes ce qui nous amène à utiliser l’algorithme EM pour donner une estimation maximale à ces valeurs, puis nous allons calculer les valeurs de probabilité de chaque nœud du MEBN.

a) **Étape 1** : Catégorisation des attributs numérique

Après la collecte des données près de l’expert du domaine, nous avons pu déterminer l’ensemble des intervalles pour chaque attribut.

Cette étape est nécessaire pour la construction du MEBN, nous avons traduit les valeurs numériques des facteurs (âge, diamètre, longueur) (Table 4.5), qui existent dans le dataset a des catégories qui sont définit par des intervalles avec l’aide de l’expert du domaine selon son expertise, le reste des attributs est déjà sous une forme normalisée.

<b>Facteur</b>	<b>Intervalle</b>	
Âge	QAA [0-10] QAB [10-20] QAC [20-30] QAD [>30]	en année
Diamètre	QDA [0-80] QDB [80-200] QDC [200-300] QDD [>300]	en millimètre
Longueur	QLA [0-100] QLB [100-200] QLC [200-300] QLD [>300]	en mètre

TABLE 4.5 – Catégorisation des facteurs.

La figure ci-dessous (Fig. 4.3) représente le Data set obtenu après la définition des catégories pour chaque facteur à nominaliser.

ID_ARC	Diametre	Materiaux	Longueur	Corrosivite	Age	Type	Trafic	Fuite
1024580	QDB	PEHD	QLA	FAIBLE	QAD	argileux	MOYEN	non
1024576	QDB	FD	QLB	FAIBLE	QAD	argileux	MOYEN	non
1013808	QDD	FD	QLA	FORT	QAD	sableux	FORT	oui
1013807	QDD	FD	QLA	FORT	QAD	argileux	FAIBLE	oui
1013804	QDB	FD	QLA	MOYEN	ARCHIVES	limoneux	FAIBLE	non
1013803	QDB	PEHD	QLA	FORT	QAD	sableux	FORT	oui
1013801	QDA	FD	QLA	MOYEN	QAD	limoneux	FORT	oui
1013799	QDB	FD	QLA	MOYEN	QAD	sableux	FORT	oui
1013753	QDB	FD	QLA	MOYEN	QAD	limoneux	FORT	oui
1013752	QDD	PEHD	QLA	FORT	QAD	sableux	FORT	oui
1013749	QDB	PEHD	QLA	MOYEN	QAD	argileux	FORT	oui
1008539	QDC	FD	QLA	FAIBLE	QAD	sableux	MOYEN	non
1008535	QDB	FD	QLA	FAIBLE	QAD	sableux	FORT	oui
993995	QDB	FD	QLA	FORT	QAD	argileux	MOYEN	oui
993993	QDD	FD	QLA	FORT	QAD	argileux	MOYEN	oui
987076	QDC	FD	QLA	MOYEN	QAD	sableux	MOYEN	non
987059	QDC	FD	QLB	FORT	QAD	sableux	FORT	oui
986878	QDB	FD	QLA	FORT	QAD	sableux	MOYEN	oui
986877	QDB	FD	QLA	FORT	QAD	sableux	MOYEN	oui
986863	QDB	FD	QLA	FORT	QAD	sableux	MOYEN	oui
967625	QDA	FD	QLA	FAIBLE	QAD	limoneux	FAIBLE	non
967618	QDB	FD	QLA	FAIBLE	QAD	argileux	MOYEN	non
967617	QDB	FD	QLA	FAIBLE	QAD	limoneux	FAIBLE	non
967616	QDB	FD	QLA	FAIBLE	QAD	sableux	MOYEN	non
966769	QDD	FD	QLA	FORT	QAD	sableux	MOYEN	oui
966766	QDD	FD	QLD	FORT	QAD	sableux	MOYEN	oui
953198	QDB	FD	QLA	FAIBLE	QAD	argileux	FORT	oui
911936	QDB	FD	QLA	FAIBLE	QAD	argileux	FORT	oui
911930	QDB	FD	QLA	FAIBLE	QAD	argileux	FAIBLE	non
911912	QDB	FD	QLA	FAIBLE	QAD	sableux	FAIBLE	non

FIGURE 4.3 – Dataset préparé.

b) **Étape 2 : Construction de réseau bayésien (structure et paramètres)**

Après de nombreuses discussions avec l'expert du domaine nous avons pu construire le RB de notre système, la (Fig. 4.4) représente la structure du réseau bayésien.

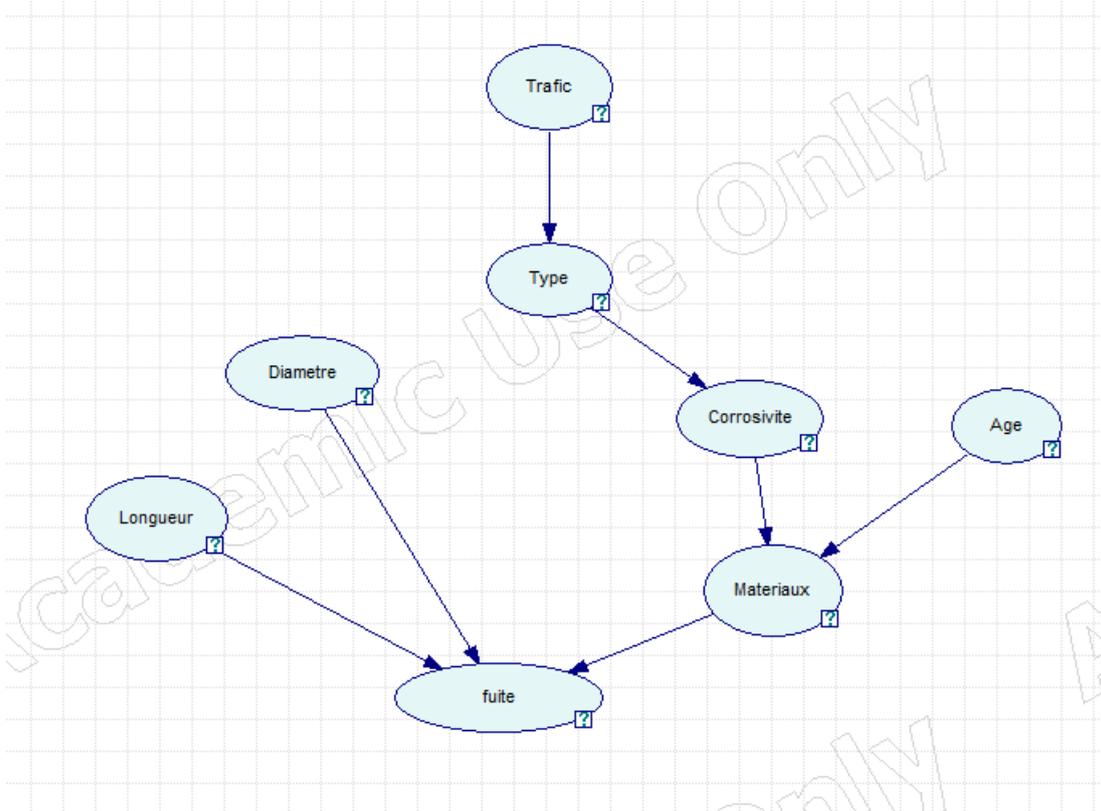


FIGURE 4.4 – Structure du réseau bayésien.

c) **Étape 3 : Application de l'algorithme EM et calcul de probabilité**

Nous avons appliqué l'algorithme EM sur notre dataset afin de donner une estimation maximale pour les valeurs de données manquantes et préparer les paramètres de notre RB pour calculer les probabilités de chaque nœud.

L'application de l'algorithme EM se fait sur le logiciel Genie SMILE, après la construction du réseau bayésien, nous avons chargé le dataset dont les données ont été catégorisées mais toujours avec quelques unes manquantes, ensuite en cliquant sur le bouton "Learning" suivi par le bouton "Learn Parameters", une fenêtre s'affiche pour le choix des paramètres de l'algorithme, nous avons choisi "Uniformized" afin que l'algorithme apprenne les paramètres d'une manière uniforme et estime les valeurs manquantes comme est affichée dans la figure (Fig. 4.5).

La (Fig. 4.6) montre les résultats obtenus des probabilités pour chaque nœud.

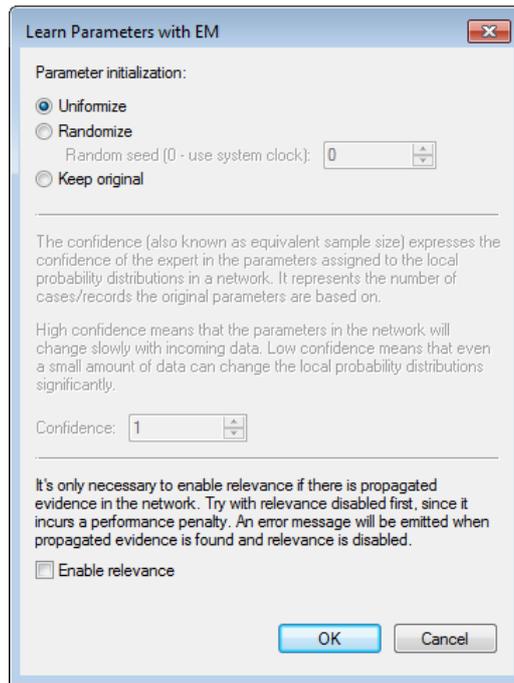


FIGURE 4.5 – Paramètres de l’algorithme EM.

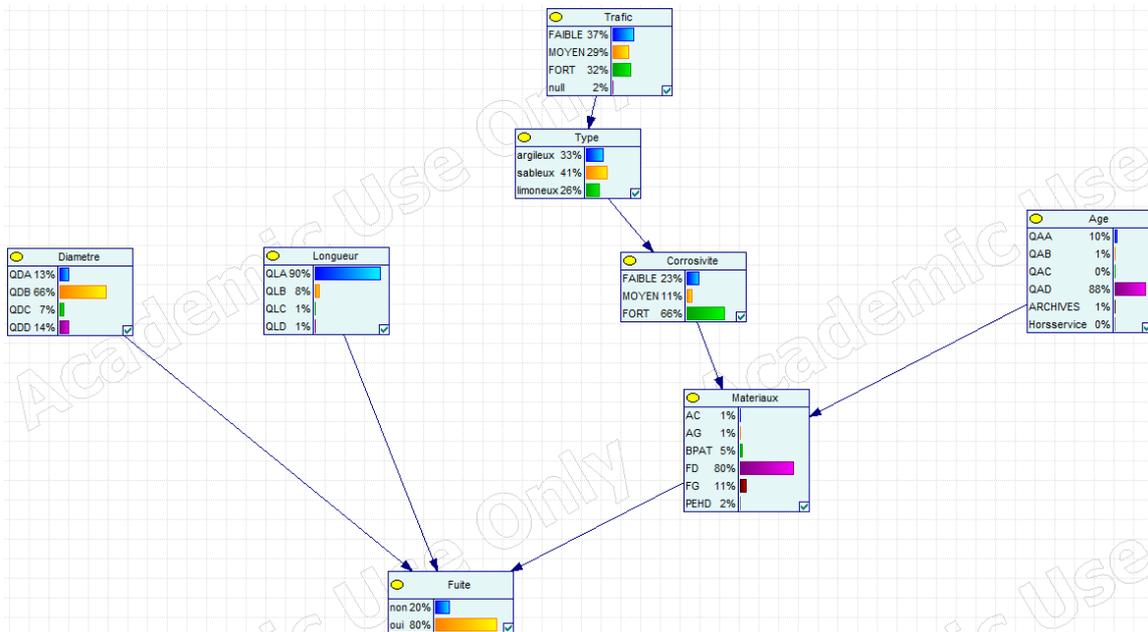


FIGURE 4.6 – Réseau bayésien.

3. **Les groupes :** Dans cette partie, nous allons identifier les entités principales, les attributs et les relations pour notre OP, et cela en utilisant le diagramme de classe UML.

La figure (Fig. 4.7) montre le diagramme de classe de notre système :

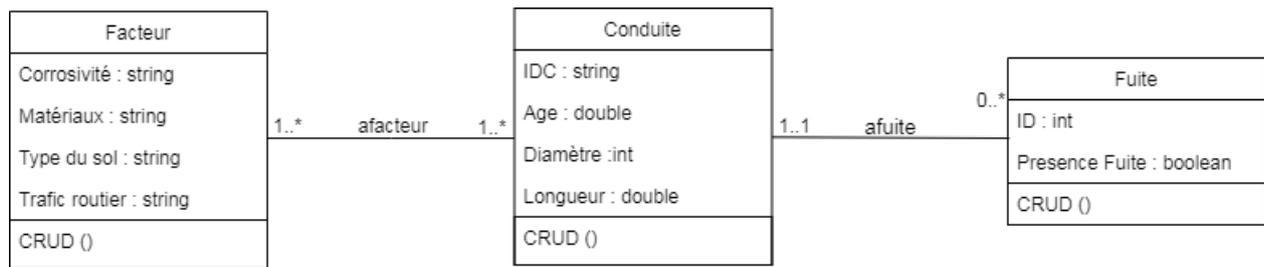


FIGURE 4.7 – Diagramme de classe.

#### 4.7.4 Implémentation

Après l’accomplissement de l’étape de la conception, ce qui nous a donné en résultat l’ontologie classique de notre système, cette dernière va nous servir de base pour la conception de l’ontologie probabiliste sur UnBBayes.

**Mapping :** C’est une méthode de modélisation des ontologies pour la représentation d’information en utilisant un langage de description qui est dans notre cas PR-OWL, qui sera accompagnée d’un MEBN afin de combiner entre le principe de l’ontologie et les probabilités ; Et pareil à toute modélisation, le mapping a un ensemble de composants de modélisation que nous allons citer ci-dessous :

1. MTheory : Considérée comme la base, une MTheory est le conteneur de toutes les composants du MEBN, elle englobe les MFrag dans un seul ensemble afin d’assurer l’unicité des probabilités de leurs contenus, une MTheory se crée automatiquement au moment de l’importation d’une ontologie sur UnBBayes.
2. MFrag : Deuxième niveau de la modélisation après les MTheory, MFrag est une représentation traductive des entités de l’ontologie classique, nous pouvons considérée dans un certain cas ou le MEBN réalisé est exactement similaire à l’ontologie que les MFrag sont les entités de l’ontologie en MEBN ;  
Un MFrag représente une distribution de probabilité conditionnelle pour les cas de ses variables aléatoires résidentielles données à leurs parents dans le graphique des fragments et les nœuds contextuels (Carvalho et al., 2016).
3. Variable ordinaire (Ordinary variable) : Première du contenu des MFrag, elle permet la déclaration de variables dans le MEBN est de la typée selon les classes de l’ontologie déjà chargée sur UnBBayes afin d’identifier tout les composants du MFrag et pour utiliser ces attributs.
4. Nœud résident (Resident node) : Noyau du MFrag, le nœud résident comme son nom l’indique permet aux différents propriétés de l’ontologie de résider dans un MFrag afin d’effectuer l’opération pour laquelle le MEBN a été conçu.  
À noter qu’un nœud résident peut être créé indépendamment de l’ontologie directement sur l’éditeur.

5. Nœud d'entrée (Input node) : Ce sont les nœuds qui permettent l'utilisation des nœuds résidents dans d'autres MFrag en leurs faisant appel afin d'intégrer leurs données dans le calcul des probabilités du nœud résident, cette intégration se fait par un arc sortant du nœud d'entrée vers le nœud résident.
6. Nœud de contexte (Context node) : Ce sont les nœuds chargés de la représentation des conditions permettant l'obtention du résultat avec ses probabilités dans le nœud résident, ces nœuds doivent être de type booléen pour différencier la présence de la condition requise ou non, ce qui signifie un point très important dans le calcul des probabilités vu leurs dépendances des conditions externes.
7. Arcs : UnBBayes propose deux (02) types d'arcs entre les composants.
  - Relation de causalité entre les nœuds d'entrées avec des nœuds résidents : dans le cas où un nœud d'entrée influence les probabilités du nœud résident.
  - Relation de causalité entre les nœuds résidents avec d'autres nœuds résidents : dans le cas où un nœud résident influence un autre nœud résident.

La traduction de notre ontologie classique en MEBN sur UnBBayes avec l'utilisation des composants mentionnés précédemment nous a permis de construire un MEBN composé de :

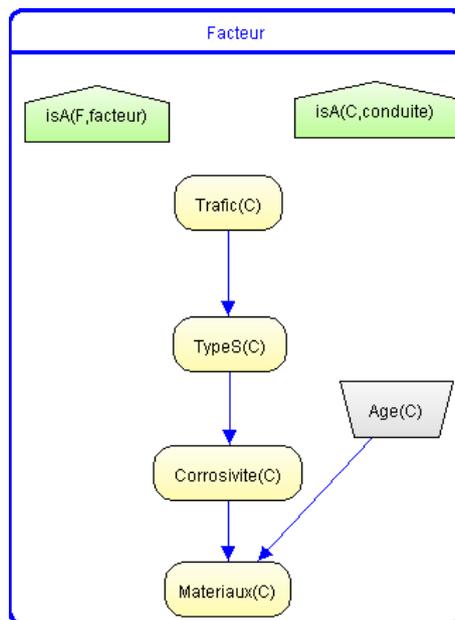


FIGURE 4.8 – MFrag des facteurs.

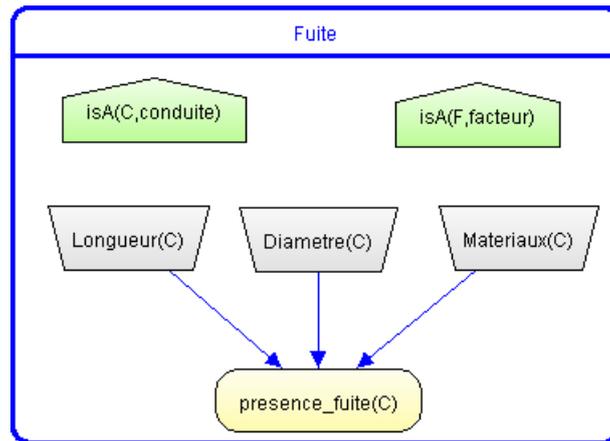


FIGURE 4.9 – MFragment des fuites.

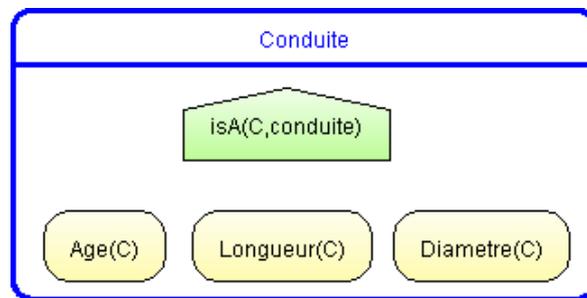


FIGURE 4.10 – MFragment des conduites.

**Probabilités :** Après avoir structuré le MEBN suivant les données, nous sommes arrivés à la dernière étape qui représente la différence pour laquelle nous avons utilisé les ontologies probabilistes, les Distributions de Probabilité Locales (LPD) pour toutes les variables aléatoires (VR) résidents.

Chaque MFragment définit les LPD pour ses nœuds résidentiels, présentés sous forme d’ovales jaunes. Ces distributions sont conditionnées par la satisfaction des contraintes exprimées par les VR contextuels, présentés sous forme de pentagones verts. Les distributions locales peuvent dépendre des valeurs des VR récurrents, présentés sous forme de trapèzes gris, dont les distributions sont définies dans les MFrags dans lesquelles ils résident (Carvalho et al., 2016).

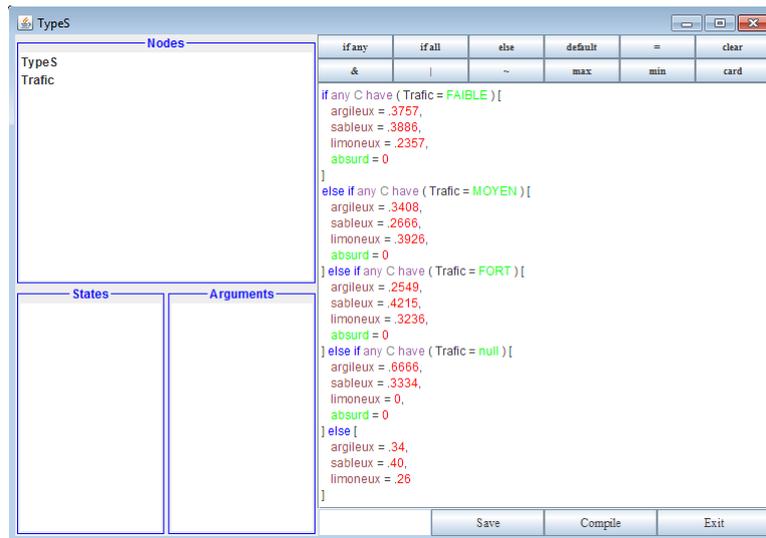


FIGURE 4.11 – LPD du nœud "TypeS" (MFRag Facteur).

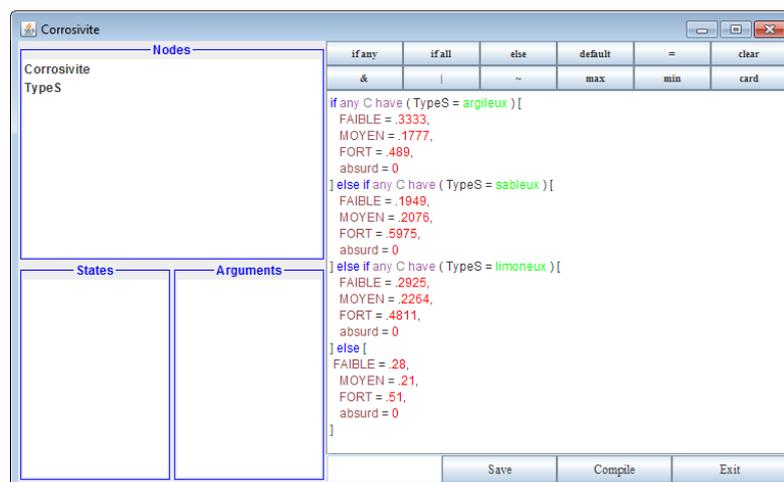


FIGURE 4.12 – LPD du nœud "Corrosivité" (MFRag Facteur).

La distribution de probabilités locale (LPD) va permettre au MEBN de couvrir les différents cas possibles de présence de fuites, selon les valeurs d'état des facteurs, cette distribution est introduite sous forme de requête extraite du dataset ou nous ayant les différentes combinaisons qui peuvent donner une forte probabilité à la présence d'une fuite. L'exécution de ces commandes crée un réseau bayésien spécifique à la situation (SSBN), voici le résultat de l'exécution de notre système, qui donne les probabilités de la présence et l'absence d'une fuite dans un cas standard.

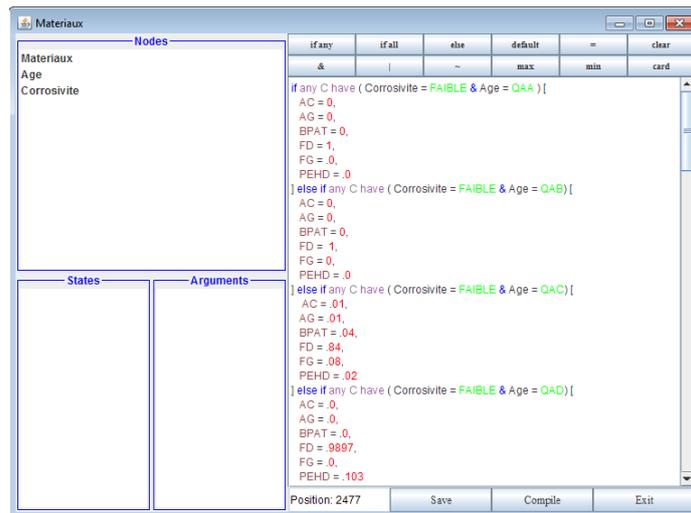


FIGURE 4.13 – LPD du nœud "Matériaux" (MFRag Facteur).

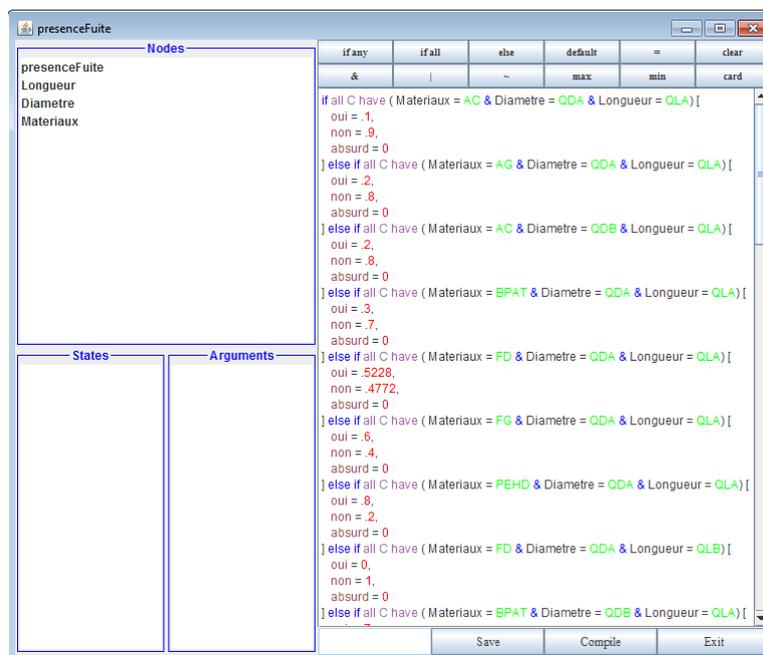


FIGURE 4.14 – LPD du nœud "presenceFuite" (MFRag Fuite).

## 4.7.5 Test

Après la conception du MEBN et l'introduction des requêtes des différents cas de fuites, il est temps présentement de tester la structure pour s'assurer de sa fiabilité.

Comme dans toute technique de modélisation, le modèle conçu doit assurer les exigences de l'utilisateur tout en respectant les règles de conception, pour cela nous allons effectuer des tests d'évaluation afin de repérer les défaillances et les corriger, la littérature nous propose deux (02) méthodes de test d'évaluation ; la vérification et la validation, l'objectif de la méthode de

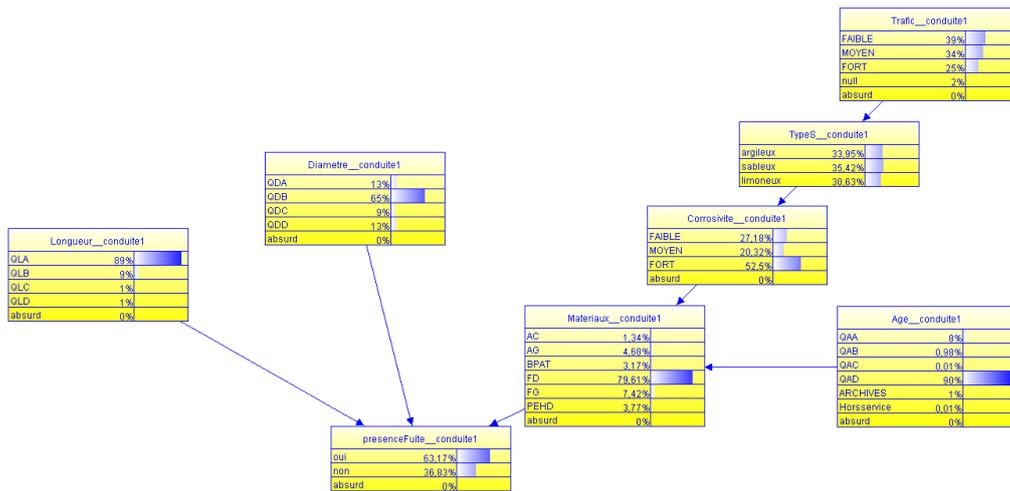


FIGURE 4.15 – SSBN du nœud "presenceFuite".

vérification est l'assurance de la satisfaction des besoins du client, tant-dis que la méthode de validation a pour but de s'assurer du bon déroulement du système sur la machine qui donnera au final des résultats réalistes et corrects.

1. La vérification :

Notre système a été conçu pour assurer la fonctionnalité de la prédiction des fuites, cette assurance se traduit techniquement par l'affichage de la probabilité de la présence d'une fuite et de son absence, notre système assure cette demande par le calcul des probabilités du nœud 'presenceFuite' résultantes de l'influence des probabilités des nœuds entrants exprimant les facteurs influant la conduite.

Donc notre système satisfait la demande de l'utilisateur exprimée dans le besoin de savoir la présence d'une fuite selon les règles de modélisation.

2. La validation :

Après la réussite de la vérification, il reste l'évaluation des résultats donnés par notre système, pour cela, nous allons effectuer 2 scénarios de tests totalement différents afin de toucher à de différents cas ;

Les scénarios seront en forme d'hypothèse de cas, suivi par le résultat attendu ensuite les évidences du cas, conclu par l'affichage du résultat obtenu après l'introduction du scénario.

— Scénario n01 :

— Hypothèse :

presenceFuite(C), C en référence à conduite.

— Résultat désiré :

Présence de la fuite.

— Évidences : C en référence à conduite.

Diametre (C) = QDB.

Materiaux (C) = FG.

Longueur (C) = QLB.

Corrosivite (C) = FORT.

Age (C) = QAD.

TypeS (C) = sableux.

Trafic (C) = MOYEN.

Et voici le résultat obtenu (Fig. 4.16) :

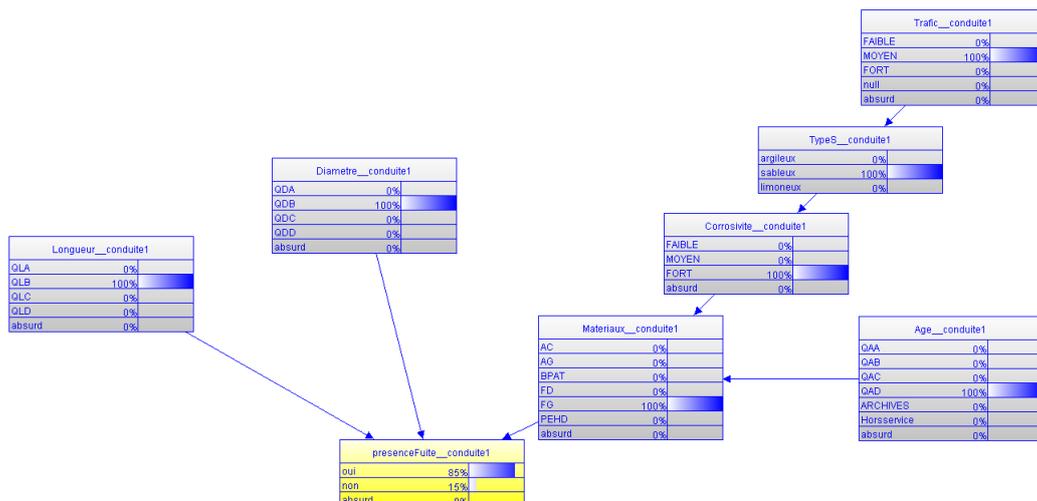


FIGURE 4.16 – Résultat du SSBN généré suivant le scénario n01.

Nous remarquons que le résultat voulu est obtenu, place au deuxième scénario qui va tester un autre cas d'absence de fuite.

- Scénario n02 :
    - Hypothèse :  
presenceFuite(C), C en référence à conduite.
    - Résultat désiré :  
Absence de la fuite.
    - Évidences : C en référence à conduite.
- Diametre (C) = QDC.  
 Materiaux (C) = FD.  
 Longueur (C) = QLB.  
 Corrosivite (C) = MOYEN.  
 Age (C) = QAD.  
 TypeS (C) = limoneux.  
 Trafic (C) = MOYEN.

Et voici le résultat obtenu (Fig. 4.17) :

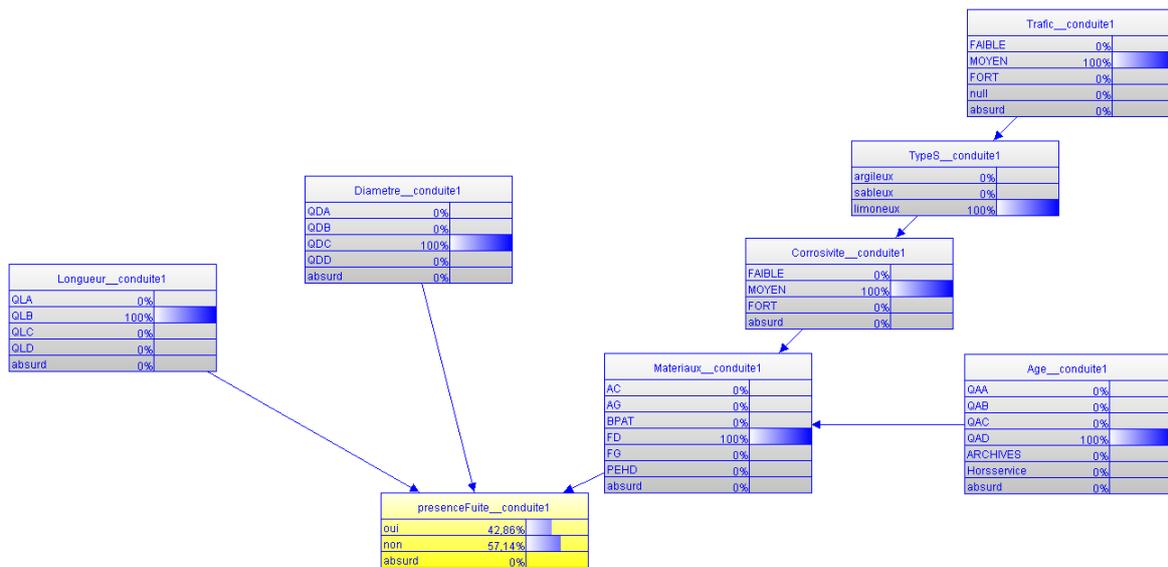


FIGURE 4.17 – Résultat du SSBN généré suivant le scénario n02.

Le système a réussi à générer les résultats attendus d'une manière précise pour les deux scénarios testés, ce qui traduit la bonne structure du système donc la réussite du test de la validation.

#### 4.7.6 Résultat final

Le processus de modélisation de l'ontologie probabiliste nous a permis de transformer notre ontologie classique à une ontologie probabiliste en PR-OWL, capable d'exprimer l'état d'une

conduite avec les probabilités de présence des fuites basée sur l'influence des facteurs et leurs probabilités.

## **4.8 Conclusion**

Nous avons abordé durant ce chapitre le processus de création de l'ontologie probabiliste de notre système, en suivant le processus UMP-ST afin de pouvoir détecter la présence d'une fuite selon les valeurs des facteurs influents. Ce résultat va nous servir de matière à utiliser pour la réalisation de notre système de détection de fuites.



## Chapitre 5

---

# Implémentation du système de prédiction des fuites

---

## 5.1 Introduction

Dans la continuation du développement de notre système, après que nous avons fini la conception il est temps d'implémenter notre ontologie probabiliste sur un système informatique dédié à la prédiction des fuites.

Nous allons aborder durant ce chapitre les outils et les étapes de l'implémentation de notre système.

## 5.2 Outils utilisés pour le développement du système

Le développement a nécessité l'utilisation d'un ensemble d'outils que nous citons ci-dessous :

### 1. SMILE

SMILE est un moteur de raisonnement et d'apprentissage / découverte causale pour les modèles graphiques, tels que les réseaux bayésiens, les diagrammes d'influence et les modèles d'équations structurelles. Techniquement, il s'agit d'une bibliothèque de classes C++ qui peut être intégrée dans un logiciel utilisateur existant via son API, améliorant ainsi les produits utilisateur avec des capacités de modélisation de décision. SMILE est entièrement portable et disponible pour la plupart des plates-formes informatiques - du centre de données à embarqué ([Fusion, 2020](#)).

Les développeurs de cet outil proposent des wrappers pour SMILE qui permettent de l'utiliser à partir de Java, Python, R, .NET et d'autres environnements de développement.

## 2. **PROTÉGÉ**

Un éditeur d'ontologie, gratuit et open source et un framework pour la création de systèmes intelligents ([Protégé, 2020](#)).

## 3. **JAVA**

Java est un langage de programmation orienté objet et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Il a été officiellement présenté le 23 mai 1995 au SunWorld.

La société Oracle racheta alors la société Sun en 2009, ce qui explique pourquoi ce langage appartient désormais à Oracle.

La technologie Java est exploitée dans le monde entier pour développer et fournir des applications mobiles et imbriquées, des jeux, du contenu Web et des logiciels d'entreprise. Utilisée par plus de 9 millions de développeurs dans le monde, elle permet de développer, de déployer et d'utiliser efficacement des applications et des services fascinants ([Java, 2020](#)).

## 4. **Eclipse IDE**

Eclipse IDE est un environnement de développement intégré libre (le terme Eclipse désigne également le projet correspondant, lancé par IBM) extensible, universel et polyvalent, permettant potentiellement de créer des projets de développement mettant en œuvre n'importe quel langage de programmation.

Eclipse IDE est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions.

La spécificité d'Eclipse IDE vient du fait de son architecture totalement développée autour de la notion de plug-in (en conformité avec la norme OSGi) : toutes les fonctionnalités de cet atelier logiciel sont développées en tant que plug-in.

Plusieurs logiciels commerciaux sont basés sur ce logiciel libre, comme par exemple IBM Lotus Notes 8, IBM Symphony ou Websphere Studio Application Developer ([Eclipse-IDE, 2020](#)).

## 5. **UnBBayes**

UnBBayes est un framework de réseau probabiliste écrit en Java. Il a à la fois une interface graphique et une API avec inférence, échantillonnage, apprentissage et évaluation. Il prend en charge les réseaux bayésiens, les diagrammes d'influence, MSBN, OOBN, HBN, MEBN / PR-OWL, PRM, la structure, les paramètres et l'apprentissage incrémental ([UnBBayes, 2020](#)).

## 6. **OWL-API**

La nécessité d'avoir un outil fiable pour manipuler des ontologies, notamment dans un environnement dynamique, a poussé les informaticiens à développer OWL-API. Il s'agit d'un ensemble d'interfaces riches en fonctionnalité, permettant une manipulation flexible

d'ontologies. De nombreux projets utilisent OWL-API comme outil de développement. A titre d'exemple, on peut citer Protégé-4, Swoop et OntoTrack ([OWL-API, 2020](#)).

#### 7. Protégé OWL-API

Est une bibliothèque Java open source pour la Web Ontology Language (OWL) et RDF (S). L'API fournit des classes et des méthodes pour charger et enregistrer des fichiers OWL, pour interroger et manipuler des modèles de données OWL, et pour effectuer un raisonnement basé sur les moteurs Description Logic. En outre, l'API est optimisée pour la mise en œuvre d'interfaces utilisateur graphiques ([Protege-wiki, 2020](#)).

### 5.3 Mise en œuvre

La suite du processus de développement du système est l'exploitation des ressources obtenues, après la traduction de notre ontologie format OWL en ontologie probabiliste format PR-OWL, grâce au réseau bayésien multi-entités qu'il nous a donné un MTheory, contenant trois (03) MFrag ou dans chaque MFrag des nœuds extraits des classes de l'ontologie, englobé dans un fichier sous format UBF, ce fichier va permettre la génération d'une base de connaissances sous format PLM et qui aura comme rôle la sauvegarde des données manipulées par le système, c'est trois (03) fichiers définissent l'ensemble des ressources que nous allons utiliser dans le développement du système, et cette utilisation nécessite des outils, que nous allons mentionner dans le point suivant.

### 5.4 Base de connaissances

Pour assurer la pertinence et la durabilité des informations du système, nous avons utilisé un fichier format PLM comme base de connaissances pour la sauvegarde des données des conduites introduites dans le système afin que l'utilisateur pourra les utiliser en cas de besoin au futur.

A noter que le format du fichier (format PLM) est utilisé par UnBBayes pour la création des bases de connaissances.

```

|;;; -*- Mode: Lisp; Package: STELLA; Syntax: COMMON-LISP; Base: 10 -*-
(CL:IN-PACKAGE "STELLA")

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

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

(IN-DIALECT :KIF)

(ASSERT (FACTEUR_LABEL FACTEUR1))
(ASSERT (CONDUITE_LABEL CONDUITE1))
(ASSERT (CONDUITE_LABEL CONDUITE2))
(ASSERT (CONDUITE_LABEL CONDUITE3))
(ASSERT (FUITE_LABEL OUIFUITE))
(ASSERT (FUITE_LABEL NONFUITE))
(ASSERT (THING_LABEL QAA))
(ASSERT (THING_LABEL QAB))
(ASSERT (THING_LABEL AG))
(ASSERT (THING_LABEL QAC))
(ASSERT (THING_LABEL QAD))
(ASSERT (THING_LABEL FAIBLE))
(ASSERT (THING_LABEL AC))

```

FIGURE 5.1 – La Base de connaissance

## 5.5 Interfaces

### 5.5.1 Interface de l'accueil

Voici la première interface qui s'affiche à l'utilisateur, une interface simple et lisible et qui propose les fonctionnalités principales de notre système d'une manière homogène.

Les boutons affichés ont pour mission d'accéder aux différentes fonctions du système (Fig. 5.2).

1. **Bouton "Afficher"** : Permet d'afficher la liste des conduites sauvegardées.
2. **Bouton "Rechercher"** : Permet de rechercher une conduite en introduisant son nom dans le champ du texte au-dessus.
3. **Bouton "Prédiction"** : Permet d'ouvrir la fenêtre de la prédiction pour afficher les probabilités de présence et absence de fuite dans la conduite sélectionnée.
4. **Bouton "Ajouter conduite"** : Permet d'ouvrir la fenêtre de l'ajout pour permettre à l'utilisateur d'introduire une nouvelle conduite avec ses facteurs et la sauvegarder dans la base de connaissance.
5. **Bouton "Supprimer"** : Permet de supprimer une conduite avec ses facteurs de la base de connaissances (Fig. 5.3).
6. **Bouton "Informations sur le système"** : Permet d'afficher les informations concernant les performances du système.

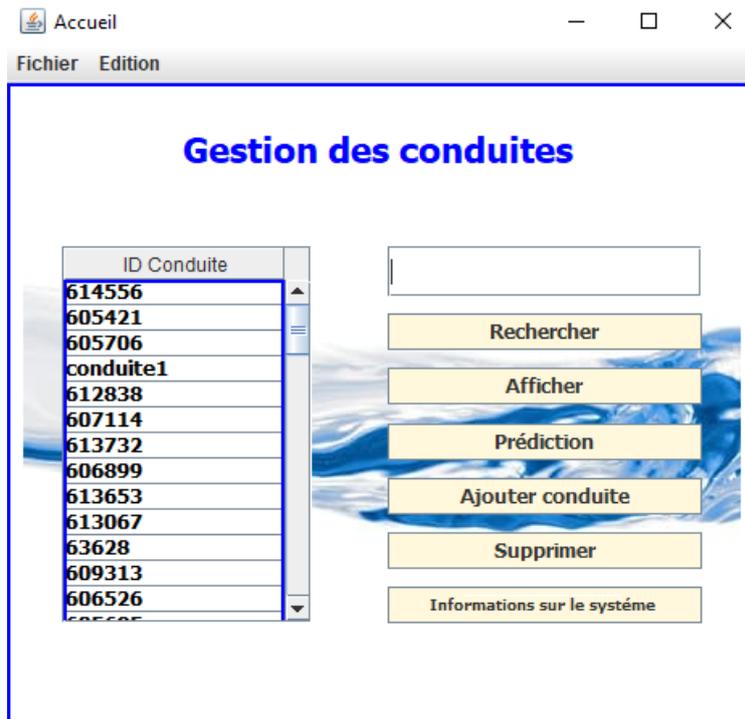


FIGURE 5.2 – Interface "Accueil".

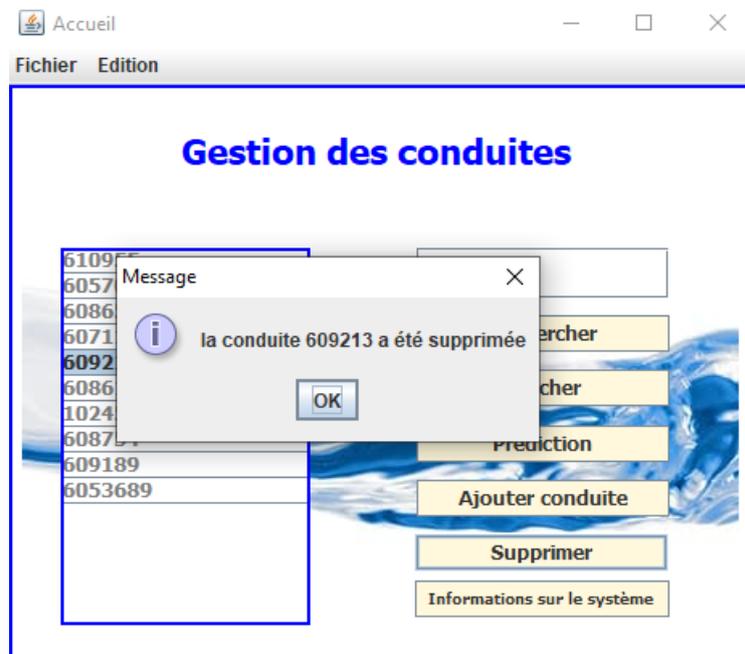


FIGURE 5.3 – Interface "Supprimer".

## 5.5.2 Interface d'ajout d'une conduite

Cette interface s'affiche après avoir cliqué sur le bouton "Ajouter conduite" de l'interface "Accueil", elle permet à l'utilisateur d'introduire le nom de la conduite avec les valeurs des

facteurs influents sur elle d'une manière simple et facile juste en sélectionnant une valeur des files déroulantes, afin de permettre au système d'utiliser ces valeurs pour la prédiction, il confirme en cliquant sur le bouton "Ajouter" pour la sauvegarder sur la base de connaissance, et au final il clique sur "Retour" pour revenir à l'accueil (Fig. 5.4).

Paramètre	Valeur
ID Conduite	635447
Trafic routier	FAIBLE
Type du sol	Sableux
Matériaux	AC
Corrosivité	MOYEN
Longueur	QLA
Diamètre	QDA
Age	QAA

FIGURE 5.4 – Interface "Ajouter Conduite".

### 5.5.3 Interfaces de la prédiction

#### 1. Interface prédiction :

Après avoir ajouté la conduite, l'utilisateur retourne à l'accueil pour lancer la prédiction, il choisit la conduite voulue de la file et clique sur le bouton "Prédiction", l'interface de la prédiction s'affichera où il aura la main de choisir la manière de voir le résultat de la prédiction, la première en cliquant sur le bouton "Afficher résultat" ou il a la possibilité de voir le résultat sous une forme de liste qui affiche en en-tête les probabilités de présence et de l'absence de la fuite, l'autre alternative est en cliquant sur le bouton "Vision graphique" pour voir le résultat sous forme graphique (Fig. 5.5).

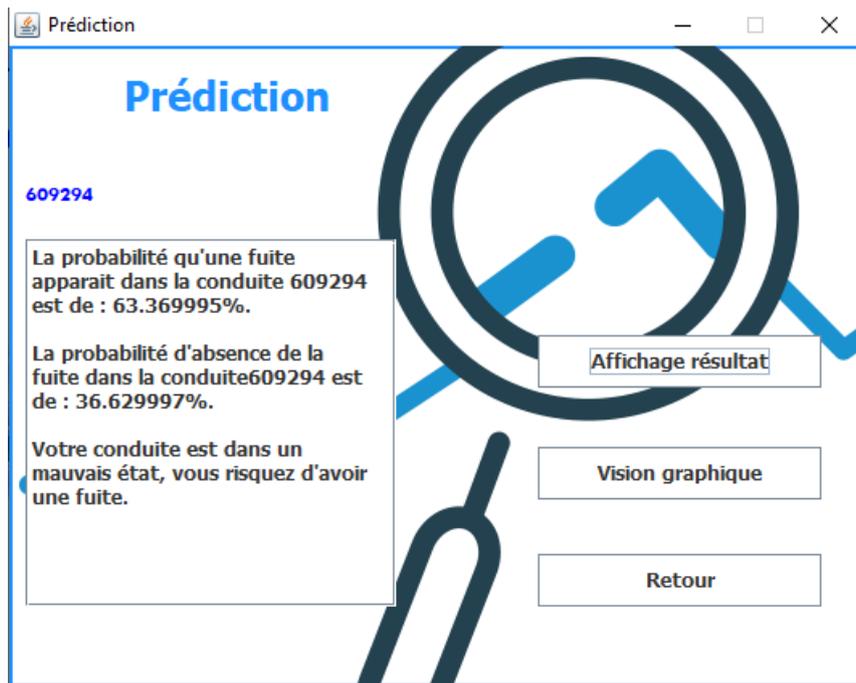


FIGURE 5.5 – Interface "Prédiction".

2. Interface des probabilités de présence de la fuite :

Elle affiche un graphe en anneau qui réfère à la probabilités de présence d'une fuite dans la conduite d'une manière lisible et claire (Fig. 5.6).

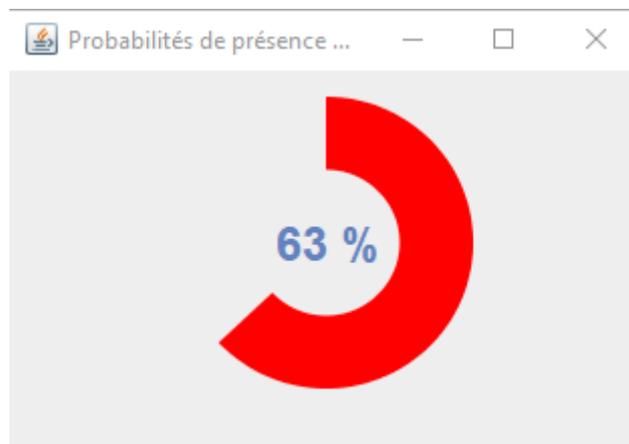


FIGURE 5.6 – Interface "Probabilités de présence de la fuite".

### 3. Interface des probabilités d'absence de la fuite :

Cette interface affiche à l'opposé de sa précédente les probabilités d'absence de la fuite (Fig. 5.7).

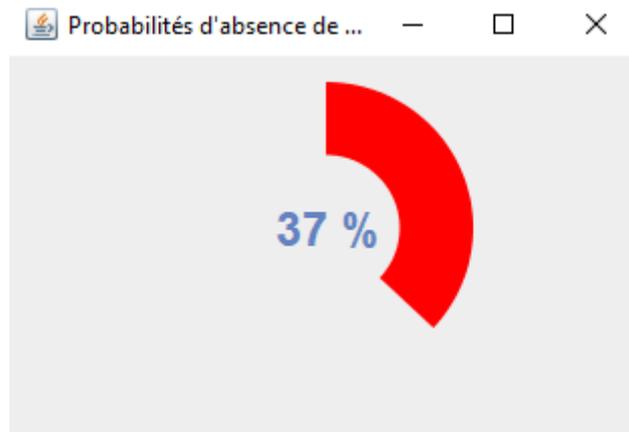


FIGURE 5.7 – Interface "Probabilités d'absence de la fuite".

## 5.6 Validation du système

Validation : La validation vise à établir que le "bon système a été construit", c'est-à-dire qu'il réalise son utilisation prévue dans son environnement opérationnel (Carvalho et al., 2016).

La validation consiste à tester l'efficacité du système, la précision de ses résultats par rapport aux données précédemment introduites à partir du dataset, afin de connaître le niveau de performance du système et les taux des résultats correctes et incorrectes partagés sur quatre (04) variables :

True Positive (TP) : Le système donne un résultat positif similaire au résultat du dataset.

True Negative (TN) : Le système donne un résultat négatif similaire au résultat du dataset.

False Positive (FP) : Le système donne un résultat positif alors que le résultat dans le dataset est négatif.

False Negative (FN) : Le système donne un résultat négatif alors que le résultat dans le dataset est positif.

Dans notre cas, nous avons effectué ce test en utilisant 20% du dataset, ou nous avons testé cas par cas cette sélection.

**La précision :** Permet de répondre à la question suivante : Quelle proportion d'identifications positives était effectivement correcte ?

Mathématiquement, la précision est définie comme suit :  $\text{Précision} = \text{TP} / (\text{TP} + \text{FP})$  (Prec, 2020).

**Le rappel :** Le rappel permet de répondre à la question suivante : Quelle proportion de résultats positifs réels a été identifiée correctement ?

Mathématiquement, le rappel est défini comme suit :  $Rappel = TP / (TP+FN)$  (Prec, 2020).

**F-mesure :** Moyenne harmonique de la précision et du rappel.

Mesure la capacité du système à donner toutes les solutions pertinentes et à refuser les autres.

Mathématiquement, le F-mesure est défini comme suit :  $2(Précision \times Rappel) / (Rappel+Précision)$  (Mesure, 2020).

**Fallout :** Le Fallout est le pourcentage de l'erreur du système.

Mathématiquement, le Fallout est défini comme suit :  $Fallout = FP / (FP+TP)$ .

Les résultats du système étaient satisfaisantes avec une valeur de précision de 93%, Rappel de 95%, F-mesure de 94% et Fallout de 7%.

L'utilisateur peut voir le taux de validité du système à partir de l'interface "Informations sur le système" ou il pourra afficher le taux (Fig. 5.8).

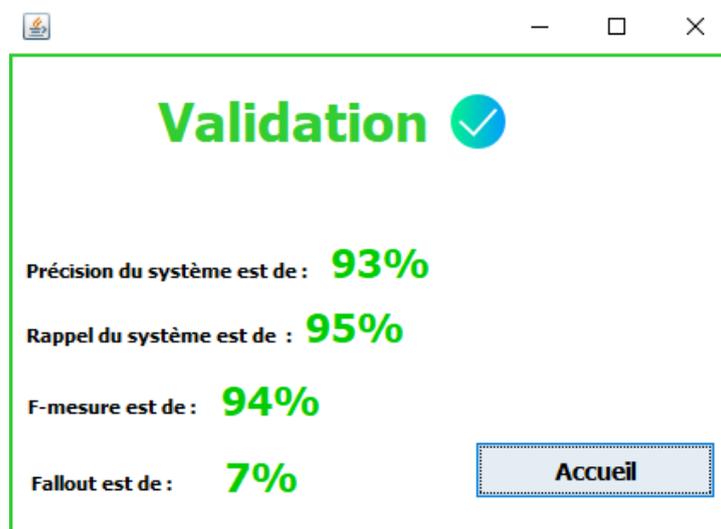


FIGURE 5.8 – Interface "Informations sur le système".

## 5.7 Conclusion

Pour ce dernier chapitre, nous avons montré la finalité de notre travail qui se conclut par l'implémentation de l'ontologie probabiliste avec la base de connaissances, ou nous avons montré les outils utilisés dans la programmation avec leurs étapes d'utilisation suivi par les captures comme des preuves visuelles pour faciliter la compréhension de notre système.



## Chapitre 6

---

# Conclusion générale

---

Les ontologies font partie des méthodes de modélisation utilisées pour définir des entités du monde réel d'une manière explicite et formelle, afin de pouvoir les introduire ensuite dans des systèmes informatiques, ce qui rapporte un apport considérable dans le domaine de développement des solutions informatisées dans la vie réelle.

Cependant le secteur de la gestion des eaux et plus précisément le domaine de la gestion des composants de l'infrastructure de base, qui est un domaine exige une grande précision et un suivi ponctuel vu qu'une simple erreur engendrera de grosses pertes, expose un besoin, ressenti chez les acteurs du domaine notamment les sociétés algériennes chargées de la gestion des eaux, ce besoin est la prédiction des fuites avant leurs apparitions dans les réseaux d'AEP, la connaissance des probabilités de présence d'une fuite dans une conduite précise, une activité clé qui aura un impact positif sur ces acteurs sur les plan écologique et économique.

Afin de répondre a cette demande, ce travail a été consacré pour trouver une solution informatique idoine a la prédiction des fuites, pour ce fait, la notion des probabilités doit être introduite, donc la première évolution durant le processus de développement du système était le passage de l'ontologie classique vers l'ontologie probabiliste, afin d'introduire la notion des probabilités vu que cette dernière ne peut pas être prise par les ontologies classiques.

Ce passage obligatoire ne peut être fait qu'en suivant une méthodologie précise, UMP-ST, résumée dans quatre (04) principales étapes, les exigences, l'analyse et la conception, l'implémentation et au final le test, le résultat de cette démarche était une ontologie probabiliste basée sur le contenu de l'ontologie classique avec l'ajout des notions liées aux probabilités.

L'ontologie probabiliste est sous format PR-OWL est établie selon un réseau bayésien multi-entités, un modèle que nous avons réalisé grâce à l'aide de notre promotrice, permettant d'effectuer des opérations de prédiction ce qui représente le centre de notre travail.

La prédiction consiste à calculer les probabilités de la présence et de l'absence d'une potentielle fuite dans une conduite du réseau d'AEP, en utilisant les valeurs des facteurs influents sur ce canal, ces facteurs sont définis sur le système sous forme de nœuds incluent dans des MFragments et liés entre eux selon la relation de causalité, ces nœuds définissent les distributions de probabilité locales (LPD).

Les démarches suivies ont mené à la création de trois (03) fichiers principaux, une ontologie probabiliste, un réseau bayésien multi-entités et une base de connaissances, cet ensemble représente le noyau que nous avons utilisé pour le développement de notre système que nous avons exposé par le billet des captures.

Au final, nous tenons à exprimer notre forte volonté de progression pour atteindre des perspectives en améliorant notre travail d'une manière qu'il pourra gérer plus de facteurs ce qui aidera à accroître la qualité de ses prédictions et aussi d'ajouter d'autres fonctionnalités selon les besoins des acteurs du domaine ce qui mettra en valeur notre travail.

---

# Bibliographie

---

- M. S. A.Ejah Umraeni Salam<sup>1\*</sup>, Muh. Tola<sup>1</sup> and F. Maricar<sup>2</sup>. Application of svm and elm methods to predict location and magnitude leakage of pipelines on water distribution network. 2015.
- A. R. . Alberto Martini, Marco Troncosi. Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements. 2015.
- A. Baneyx. *Construire une ontologie de la Pneumologie Aspects théoriques, modèles et expérimentations*. PhD thesis, 2007.
- A. Ben Mrad. *Observations probabilistes dans les réseaux bayésiens*. PhD thesis, Valenciennes, 2015.
- A. Bouzaïeni. Réseaux bayésiens et quelques applications en traitement d'images. In *CORIA-CIFED*, pages 377–380, 2014.
- R. N. Carvalho, K. B. Laskey, and P. C. Da Costa. Uncertainty modeling process for semantic technology. *PeerJ Computer Science*, 2 :e77, 2016.
- P. C. Costa. *Bayesian semantics for the Semantic Web*. Citeseer, 2005.
- P. C. Costa, K. B. Laskey, and K. J. Laskey. Pr-owl : a bayesian framework for the semantic web. 2005.
- 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*, pages 88–107. Springer, 2006.
- A. Dupont. *Hydraulique urbaine*, volume 1. édition Eyrolles, 1979.
- Eclipse-IDE, 2020. URL <https://www.techno-science.net/definition/517.html#:~:text=Eclipse>.

- en mailles, 2020. URL <https://www.wikitp.fr/distribution/les-types-de-reseaux>.
- R. Fung and B. Del Favero. Backward simulation in bayesian networks. In *Uncertainty Proceedings 1994*, pages 227–234. Elsevier, 1994.
- B. Fusion, 2020. URL <https://www.bayesfusion.com/smile/>.
- H. Ghorbel, A. Bahri, and R. Bouaziz. Les langages de description des ontologies : Rdf & owl. *Acte des huitièmes journées scientifique des jeunes chercheurs en Génie Electrique et Informatique (GEI), Sousse-Tunisia.*, 2008.
- T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2) :199–220, 1993.
- A. Hammache and R. Ahmed-Ouamer. Système d’inférence pour une indexation de documents basée sur une ontologie de domaine. In *INFORSID*, 2006.
- M. Henrion. Propagating uncertainty in bayesian networks by probabilistic logic sampling. In *Machine Intelligence and Pattern Recognition*, volume 5, pages 149–163. Elsevier, 1988.
- E. Hlel, S. Jamoussi, and A. B. Hamadou. Intégration d’un réseau bayésien dans une ontologie. 2014.
- L. Houichi. *APPROCHE METHODOLOGIQUE POUR LE DIAGNOSTIC DES RESEAUX D’EAU*. PhD thesis, Université de Batna 2, 2012.
- iStock(2016). Stockfotos royalty free beelden en videos, 2016. URL <https://www.istockphoto.com/fr/photo/fonte-tuyau>.
- Java, 2020. URL [https://www.java.com/fr/download/faq/whatis\\_java.xml](https://www.java.com/fr/download/faq/whatis_java.xml).
- D. M. John Mashford, Dhammika De Silva and S. Burn. An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. 2009.
- Kolli(2011). *Intégration d’ontologies dans le cadre du web sémantique : une détection des relations sémantiques basée sur le RÀPC*. PhD thesis, Thèse présentée pour obtenir le grade de doctorat en Science en Informatique .Université Mentouri de Constantine.
- K. B. Laskey, P. C. Costa, and T. Janssen. Probabilistic ontologies for multi-int fusion. *Ontologies and Semantic Technologies for Intelligence*, 213(147) :2, 2010.
- K. B. Laskey, R. Haberlin, R. N. Carvalho, and P. C. G. da Costa. Pr-owl 2 case study : A maritime domain probabilistic ontology. In *STIDS*, pages 76–83, 2011.

- leaking2020. Image d'un cuivre pipe a eau cassée.  
 URL <https://www.gettyimages.fr/detail/photo/broken-leaking-copper-water-pipe-image-libre-de-droits/157436620?adppopup=true>.
- Mesure, 2020. URL <http://blog.onyme.com/apprentissage-artificiel-evaluation-precision-rappel-f-mesure/>.
- R. Mizoguchi, K. Sinita, and M. Ikeda. Task ontology design for intelligent educational/training systems. Position Paper for ITS'96 Workshop on Architectures and Methods for . . . , 1996.
- A. B. Mrad. *Observations probabilistes dans les réseaux bayésiens*. PhD thesis, 2015.
- P. Naïm, P.-H. Wuillemin, P. Leray, O. Pourret, and A. Becker. Réseaux bayésiens. *Eyrolles, Paris*, 3, 1999.
- P. Naïm, P.-H. Wuillemin, P. Leray, O. Pourret, and A. Becker. *Réseaux bayésiens*. Editions Eyrolles, 2011.
- 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.
- H.-T. Nguyen. *Réseaux bayésiens et apprentissage ensembliste pour l'étude différentielle de réseaux de régulation génétique*. PhD thesis, Université de Nantes, 2012.
- OWL-API, 2020. URL <https://www.contentside.com/owl-api-une-api-java-pour-manipuler-des-ontologies/>.
- J. Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Elsevier, 2014.
- Prec, 2020. URL <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall?hl=fr>.
- Protege-wiki, 2020. URL [https://protegewiki.stanford.edu/wiki/ProtegeOWL\\_API\\_Programmers\\_Guide](https://protegewiki.stanford.edu/wiki/ProtegeOWL_API_Programmers_Guide).
- Protégé, 2020. URL <https://protege.stanford.edu/>.
- R. ramifié, 2020. URL <https://www.wikitp.fr/distribution/les-types-de-reseaux>.
- I. Riali, M. Fareh, and H. Bouarfa. Fzmebn : Toward a general formalism of fuzzy multi-entity bayesian networks for representing and reasoning with uncertain knowledge. In *International Conference on Enterprise Information Systems*, volume 2, pages 520–528. SCITEPRESS, 2017.

- L. H. Royce A. Francis, Seth D. Guikema. Bayesian belief networks for predicting drinking water distribution system pipe breaks. 2015.
- F. Santos. L'algorithme em : une courte présentation. *CNRS, UMR, 5199*, 2015.
- A. H. Schinkel, U. Mayer, E. Wagenaar, C. A. Mol, L. Van Deemter, J. J. Smit, M. A. Van Der Valk, A. C. Voordouw, H. Spits, O. Van Tellingen, et al. Normal viability and altered pharmacokinetics in mice lacking mdr1-type (drug-transporting) p-glycoproteins. *Proceedings of the National Academy of Sciences*, 94(8) :4028–4033, 1997.
- SEAAL. Société des eaux et de l'assainissement d'alger, 2020. URL <http://www.seaal.dz/qui-sommes-nous/>.
- S. Staab and A. Maedche. *Axioms are objects, too : Ontology engineering beyond the modeling of concepts and relations*. Citeseer, 2000.
- R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering : principles and methods. *Data & knowledge engineering*, 25(1-2) :161–197, 1998.
- UnBBayes, 2020. URL <https://sourceforge.net/projects/unbbayes/>.
- M. Uschold, M. Gruninger, et al. *Ontologies : Principles, methods and applications. TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR*, 1996.
- Z. Zaabot. *Les Réseaux bayésiens. application en reconnaissance de formes à partir d'informations complètes ou incomplètes*. PhD thesis, Université Mouloud Mammeri, 2012.