

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Saad Dahleb - Blida 1



Faculté des Sciences

Département d'Informatique

Mémoire de fin d'études pour l'obtention  
Du diplôme de Master en informatique

Option : Ingénierie des logiciels

Thème

Conception et implémentation d'un Système de Recherche  
D'information à base d'une ontologie de domaine relatif à l'E-  
Learning

Réalisé par

Mr KHAILI Souhaib

Mr MOULOUD Ibrahim

Devant le membre Respecté

Mme MESKALDJI Khouloud (Présidente)

Mme CHERFA Imene (Examinatrice)

Mme ZAHRA Fatma Zohra (Promotrice)

Mr BOUDINA Abdelmadjid (Encadreur)

Promotion 2020-2021



# Remerciements

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

Tous nos infinis remerciements à notre recteur, vice-recteur et chef du département, notre promotrice Mme Zahra et Mr Boudina pour leurs aides, leurs conseils et leurs remarques qui nous ont permis de présenter notre travail dans sa meilleure forme.

Un remerciement particulier et sincère à nos chers parents et familles pour le soutien et les encouragements qu'ils nous ont toujours apportés, ainsi qu'à tous nos amis qui ont contribué à notre réussite.

Nos gratitude s'expriment à tous les consultants et internautes rencontrés lors des recherches effectuées et qui ont accepté de continuer notre travail en toute gentillesse

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

Enfin nous remercions tous ceux qui ont contribué de près ou de loin à l'élaboration de ce modeste travail, trouvant ici l'expression de notre profonde gratitude et profonds respects. Merci à tous et à toutes.

# Dédicaces

Je dédie mon travail à mes chers parents pour tous les sacrifices, l'amour,  
le soutien, les conseils et les motivations durant mon parcours éducatif

Et à tous mes frères qui m'ont encouragé à continuer mon chemin vers le  
succès

Et à tous les membres de ma famille qui m'ont aidé à atteindre mes  
objectifs

Et à tous mes amis qui ont été toujours avec moi

Merci à tous ceux qui ont prié pour moi jusqu'au bout

*Mouloud Ibrahim*

# Dédicaces

Je dédie ce mémoire à :

Mes chers parents, que nulle dédicace ne puisse exprimer mes sincères sentiments, pour leur aide, leur encouragement et leur patience illimitée en témoignage de mon profond amour et respect pour leurs grands sacrifices.

Mes chères sœurs, pour leur grand amour et leurs soutiens qu'ils trouvent ici l'expression de ma haute gratitude.

Ma très chère grande mère pour le gout à l'effort qu'il a suscité en moi, de par sa rigueur.

Mes chers amis (e), qui sans leur encouragement ce travail n'aura jamais vu le jour.

Et à toute ma famille, et à tous ceux que j'aime.

*KHAILI Souhaib*

Dans ces derniers temps, le volume des ressources sur internet a connu une augmentation significative, Dans cet immense entrepôt de données, les systèmes de recherche d'informations ne permettent pas de retourner aux utilisateurs les documents pertinents pour répondre à leurs besoins exprimés par une requête sur une collection de documents, Afin d'améliorer la pertinence et la précision et diminuer le temps de réponse du système de recherche d'information, nous proposons dans ce travail une approche qui se base sur l'utilisation d'une ontologie de domaine pour l'indexation d'une base de documents et l'utilisation des liens sémantiques entre documents ou fragments de documents de la collection, pour permettre l'inférence de tous les documents pertinents et l'ajout d'un module de classification des utilisateurs afin d'éliminer le bruit et répondre exactement à leurs interrogations , Cette approche est appliquée sur le domaine de l'E-learning dans le contexte du web sémantique. Quelques résultats obtenus sont également présentés.

**Mot clés :** recherche d'information, ontologie, e-learning, web sémantique.

In recent times, the volume of resources on the internet has increased significantly, In this huge data warehouse, information retrieval systems do not allow users to get relevant documents to meet their needs expressed by a query on a collection of documents, In order to improve the relevance and precision and decrease the response time of the information retrieval system, we propose in this work an approach which is based on the use of a domain ontology for the indexing of a database and the use of semantic links between documents or fragments of documents of the collection, to allow the inference of all relevant documents and the addition of a user classification module to eliminate noise and answer their requests exactly, This approach is applied to the field of E-learning in the context of the semantic web. Some results obtained are also presented.

**Keywords :** information retrieval, ontology, e-learning, semantic web.

في الآونة الأخيرة ، زاد حجم الموارد على الإنترنت بشكل كبير ، على مستوى مستودع البيانات الضخم هذا ، أنظمة استرجاع المعلومات لا تتيح للمستخدمين إمكانية الحصول على المستندات ذات الصلة لتلبية احتياجاتهم التي يعبر عنها استعلام حول مجموعة من المستندات ، من أجل تحسين ملاءمة ودقة وتقليل وقت استجابة نظام استرجاع المعلومات ، نقترح في هذا العمل منهجًا يعتمد على استخدام أنطولوجيا المجال لفهرسة قاعدة البيانات واستخدام الروابط الدلالية بين المستندات أو أجزاء من المستندات من المجموعة ، للسماح بالاستدلال على جميع الوثائق ذات الصلة وإضافة وحدة تصنيف المستخدم للتخلص من الضوضاء والإجابة على طلباتهم بالضبط ، تم تطبيق هذا النهج في مجال التعلم الإلكتروني في سياق الويب الدلالي. كما تم عرض بعض النتائج التي تم الحصول عليها.

**الكلمات المفتاحية:** استرجاع المعلومات، الأنطولوجيا، التعلم الإلكتروني، الويب الدلالي.



---

## Tables des matières

---

Remerciements.....	I
Dédicaces .....	II
Dédicaces .....	III
Résumé .....	IV
Tables des matières.....	VII
Liste des figures.....	XI
Liste des tableaux .....	XI
Liste des abréviations.....	XII
Introduction Générale .....	13
<i>Partie 1</i> .....	16
<i>CHAPITRE 1</i> .....	17
1. Définition de E-Learning.....	18
2. Les avantages et les inconvénients d'E-Learning.....	19
2.1. Les avantages d'E-Learning .....	19
2.2. Les inconvénients d'E-Learning.....	19
3. Les types d'apprentissage en ligne .....	20
3.1. Sychrone .....	20
3.2. Asynchrone.....	20
3.3. Mode autodirigé .....	20
4. Les caractéristiques d'E-Learning.....	20
5. L'évolution de l'E-Learning .....	20
7. Dispositif de formation à distance .....	22
7.1. Dispositif de formation.....	22
7.2. Composantes d'un dispositif de formation à distance .....	22
8. Principe de E-Learning.....	23
9. Conclusion.....	24
<i>CHAPITRE 2</i> .....	25
1. Système de recherche d'information .....	26
1.1. Définition de la recherche d'information .....	26
1.2. Définition d'un système de recherche d'information .....	26
1.3. Définition de la pertinence.....	27

1.3.1.	Calcul de pertinence .....	27
2.	Les modèles de recherche d'information .....	27
2.1.	Modèle booléen .....	28
2.2.	Modèle spatial vectoriel .....	28
2.3.	Modèle probabiliste .....	28
3.	Processus de recherche d'information .....	28
3.1.	Collection de documents .....	28
3.2.	L'indexation .....	29
3.2.1.	L'indexation manuelle .....	29
3.2.3.	L'indexation automatique .....	29
3.3.	La recherche .....	30
3.4.	Le classement .....	30
4.	Les objectifs d'un système de recherche d'information .....	31
5.	Evaluation d'un système de recherche d'information .....	31
5.1.	Mesures alternatives .....	32
5.1.2.	Mesure harmonique ( F-mesure) .....	32
5.1.3.	Mesure d'évaluation « E-Mesure » (ou F-Mesure paramétrique) .....	32
6.	Système de recherche d'information sémantique .....	33
6.1.	Définition d'ontologie .....	33
6.2.	Exemple d'ontologie .....	34
6.3.	L'ontologie en Recherche d'information .....	35
6.4.	L'indexation sémantique .....	36
7.	L'apprentissage Automatique .....	36
7.1.	Apprentissage supervisé .....	36
7.2.	Apprentissage non supervisé .....	37
7.3.	Apprentissage semi-supervisé .....	37
7.4.	Apprentissage partiellement supervisé .....	37
7.5.	Apprentissage par renforcement .....	38
7.6.	Apprentissage par transfert .....	38
8.	Conclusion .....	38
	<i>CHAPITRE 3</i> .....	39
1.	Approche classique .....	40
1.1.	Méthode de l'indexation .....	40
2.	L'approche sémantique .....	41

3.	La méthode d'indexation sémantique .....	42
4.	Travaux existants dans le domaine E-Learning .....	43
4.1.	Comparaison des approches.....	44
4.2.	Classement des approches relative à l'E-Learning .....	46
5.	Conclusion.....	46
<b>Partie 2 .....</b>		<b>48</b>
<b>CHAPITRE 4.....</b>		<b>49</b>
1.	L'architecture globale du système .....	50
1.1.	Le moteur de recherche .....	51
1.2.	Le module de recherche .....	51
1.3.	Base de ressources indexées .....	51
1.4.	Moteur d'inférence .....	51
1.5.	Module de classification des requêtes.....	51
1.6.	Les acteurs intervenants .....	54
1.6.1.	Personnel .....	54
1.6.2.	Apprenants.....	55
1.7.	L'ontologie globale.....	55
1.7.1.	L'ontologie des ressources pédagogiques.....	55
1.7.2.	L'ontologie des objets pédagogiques .....	57
2.	Définition des liens sémantiques.....	59
3.	Définition des règles d'inférence.....	59
4.	Méthode de recherche.....	60
4.1.	Le cas général .....	60
4.2.	Les cas de spécification du niveau d'apprenant .....	60
4.3.	Les cas de spécification de la catégorie d'apprenant .....	61
5.	Conclusion.....	61
<b>CHAPITRE 5.....</b>		<b>62</b>
1.	Matériels et outils.....	63
1.1.	Matériels utilisés.....	63
1.2.	Outils et langage de développement.....	63
1.3.	Langage Python .....	64
1.4.	Flask (Framework).....	64
1.5.	NLTK (Natural language Toolkit) .....	64

1.6.	Scikit-Learn.....	64
1.7.	OWLready 2.....	65
1.8.	Rdflib.....	65
1.9.	RDF .....	65
1.10.	SPARQL.....	66
1.11.	HTML .....	66
1.12.	CSS .....	67
1.13.	JavaScript .....	67
1.14.	Bootstrap.....	68
1.15.	UML .....	68
1.16.	Microsoft visuel studio.....	68
1.17.	Protégé.....	69
1.18.	Bootstrap studio.....	69
2.	Implémentation du système.....	70
2.1.	Préparations des données .....	70
2.2.	Traitement des données (indexation).....	70
3.	Déploiement du Module de classification .....	70
4.	La recherche d'information .....	71
5.	Déploiement du système.....	71
6.	Page de recherche .....	72
7.	La page des résultats .....	73
8.	Evaluation .....	73
8.1.	Tests et Résultats .....	74
8.1.1.	Jeu de test.....	76
9.	Conclusion.....	77
	Conclusion Générale.....	78
	Références.....	80

---

## Liste des figures

---

<b>Figure 1</b> : Le positionnement de E-Learning	18
<b>Figure 2</b> : Les composants d'un dispositif E-Learning. [9]	22
<b>Figure 3</b> : Principe de E-Learning. [11]	24
<b>Figure 4</b> : Processus de recherche d'information [25]	31
<b>Figure 5</b> : une partie d'ontologie de thésaurus MeSH. [30]	35
<b>Figure 6</b> : Croissance de nombre des publications pour "ontologie", "E-Learning". [42]	43
<b>Figure 7</b> : Classification des approches existantes (basé sur nos études).	46
<b>Figure 8</b> : Architecture générale du système.	50
<b>Figure 9</b> : Exemple de données d'entraînement	54
<b>Figure 10</b> : Diagramme de classe des acteurs intervenant.	55
<b>Figure 11</b> : Diagramme de classe d'ontologie des ressources pédagogique . [45]	56
<b>Figure 12</b> : Diagramme d'ontologie des objets pédagogique. [46]	57
<b>Figure 13</b> : Partie abstraite du code source.	71
<b>Figure 14</b> : Architecture externe du système.	72
<b>Figure 15</b> : Capture de l'interface de recherche du système.	72
<b>Figure 16</b> : Capture de l'interface des résultats du système.	73

---

## Liste des tableaux

---

<b>Tableau 1</b> : Tableau de comparaison entre E-Learning et l'apprentissage traditionnel .....	21
<b>Tableau 2</b> : Comparaison entre les approches existantes. ....	45
<b>Tableau 3</b> : Tableau des liens sémantiques .....	59
<b>Tableau 4</b> : Tableau des règles d'inférence .....	59

---

## Liste des abréviations

---

**ACM:** Association for Computing Machinery

**API:** Application Programming Interface

**BSD:** Berkeley Software Distribution

**DAWG:** Data access working group

**DOM:** Document Object Model

**D-learning:** Distance learning (Apprentissage a distance)

**E-Learning:** Electronic learning (apprentissage en ligne)

**GUI:** Graphical user interface

**IDE:** Integrated development environment

**IDF:** Inverse document frequency

**IEEE:** Institute of Electrical and Electronics Engineers

**LMS:** learning manager system

**MEV:** Modèle d'espace vectoriel (VSM : Eng)

**MIT:** Massachusetts Institute de Technologie

**MeSH:** Medcin subject headings

**NLTK:** Natural language toolkit

**OWL:** Web Ontology Language

**OWLREaDY:** Ontology-oriented programming module for Python

**RDF:** Resource Description Framework

**RDFLIB:** Resource Description Framework library

**RI:** Recherche d'information

**SGA:** Système de gestion d'apprentissage

**SPARQL:** Standard query language and protocol for Linked Open Data and RDF databases

**SRI:** Système de recherche d'information

**TF:** Term frequency

**TICE:** Technologie de l'information et de la communication de l'enseignement

**UML:** Unified Modeling Language

**W3C:** World Wide Web Consortium

**WWW:** World Wide Web

**XML:** Extensible markup language

# Introduction Générale

## Contexte et problématique

Les systèmes de recherche d'information sont des outils qui ont permis, jusqu'à aujourd'hui, d'améliorer sans cesse la qualité des services d'accès à l'information surtout dans le domaine d'E-Learning, grâce à la capitalisation des théories issues de nombreux travaux de recherche, cependant, en raison de la surabondance de l'information d'une part et de sa large accessibilité à travers le Web, d'autre part, leur mise en œuvre est confrontée à de nouveaux problèmes.

En effet la situation est actuellement paradoxale, la masse d'informations est telle que l'accès à une information pertinente, adaptée aux besoins d'un utilisateur donné devient à la fois difficile et nécessaire. En clair, le problème n'est pas tant la disponibilité de l'information, mais sa pertinence relativement à un contexte éducatif d'utilisation spécifique.

C'est pourquoi, on est orienté vers la conception et la mise en œuvre d'un système de recherche d'information à base d'ontologie qui permettent de retourner aux utilisateurs les documents répondant exactement à leurs besoins exprimés par une requête sur une collection de documents, autrement pour améliorer la pertinence de recherche du besoin en information au bout d'une base des ressources massives.

Dans le cadre de la conception de ce système on s'est basé sur les critères importants : la structure des ressources, le type des ressources, le type des apprenants ainsi que leurs niveaux.

Cela revient d'abord à indexer une base de données par une ontologie de domaine puis définir et exploiter les liens sémantiques inter-documents.

## Objectifs

Notre travail consiste à effectuer les tâches suivantes :

1. **Indexation d'une base de données** : consiste à la représentation (index) synthétique d'une collection de données. Cela revient à définir des concepts et établir des relations entre eux de type « formé de », « dépend de » et « précédé par » ;
2. **Définition de liens sémantiques entre documents** (règles d'inférence) : consiste à définir des liens entre documents, soit la relation qui lie un document à un autre en établissant un pointeur-typé reliant deux documents (prédécesseur) et (successeur) de type « cause-résultat », « référence » et « implication » ;
3. **Développement d'un moteur de recherche** : application dédiée à l'exécution des règles d'inférences.
4. **Implémentation** : le système sera implémenté sous forme d'une application Web.  
Il est demandé pour cela, après le choix du cours à implémenter, de :
  - a. Assurer une meilleure flexibilité et portabilité de l'application ;
  - b. Développer l'ontologie moyennant un langage de représentation des ontologies ;
  - c. Permettre l'exploitation et la manipulation de l'ontologie.

### **Organisation du mémoire**

Afin de réaliser un bon travail nous avons organisé notre mémoire comme suit :

Le travail est divisé en deux parties, la première était sur la recherche bibliographique et la deuxième concerne la conception et l'implémentation du système.

#### **Partie 1 :**

Chapitre 1 : ce chapitre vise à définir les notions de base de l'E-Learning.

Chapitre 2 : Défini les systèmes de recherche d'information.

Chapitre 3 : Contiens une recherche sur les travaux connexes du domaine E-Learning.

#### **Partie 2 :**

Chapitre 4 : la proposition d'une approche pour la résolution de la problématique mentionnée.

Chapitre 5 : L'implémentation de l'approche proposée.



## *Introduction Générale*

---

Ce travail est finalisé par une conclusion générale reprenant les principaux points abordés proposé, et en exposant certaine prospective d'amélioration de notre travail.

---

*Partie 1*

*Recherche bibliographique*

---

---

*CHAPITRE 1*  
*Apprentissage en ligne*  
*« E-Learning »*

---

L'évolution du domaine informatique a permis l'apparition de nombreuses méthodes d'apprentissage. Leur intégration dans le milieu scolaire permet d'adapter l'enseignement aux besoins des élèves pour passer de la méthode dite traditionnelle à des activités d'apprentissage centrés sur eux. [1]

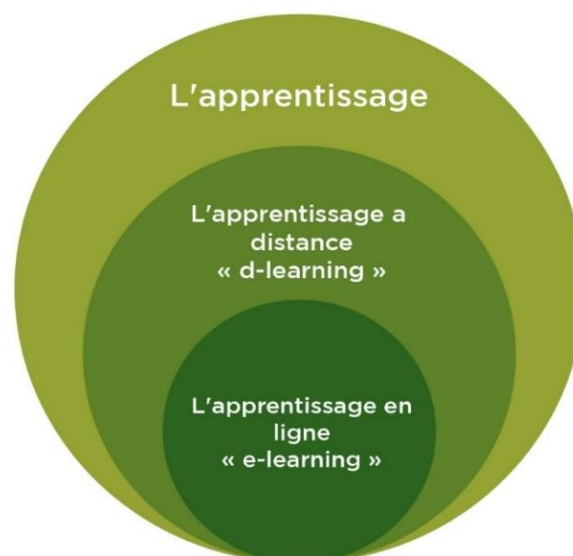
## 1. Définition du E-Learning

Le E-Learning ou l'apprentissage en ligne, signifie littéralement apprendre sur internet, est un processus d'apprentissage à distance s'appuyant sur des ressources multimédias, qui permet à une ou plusieurs personnes de se former à partir de leur ordinateur. Les supports multimédias utilisés peuvent combiner du texte, des graphismes en deux ou trois dimensions, du son, de l'image, de l'animation et même de la vidéo. [2]

Historiquement, l'apprentissage à distance (d-learning) a plus de cent années d'expérience et de traditions. Son caractéristique principale est la séparation, dans le temps et dans l'espace, de l'apprenant et de l'enseignant. L'E-Learning offre de nouvelles méthodes d'enseignement à distance basée sur les technologies informatiques et d'internet.

Aujourd'hui, l'E-Learning occupe une partie considérable de l'apprentissage à distance particulièrement et d'apprentissage en généralement, et cela dû à ses avantages par rapport aux autres méthodes d'apprentissage.

La Figure 1 illustre le positionnement de l'apprentissage en ligne en comparant avec le d-learning et l'apprentissage en général.



**Figure 1** : Le positionnement du E-Learning

## **2. Les avantages et les inconvénients d'E-Learning**

### **2.1. Les avantages d'E-Learning**

L'avantage le plus important d'apprentissage à distance est la flexibilité qu'il fournit par rapport à la méthode d'apprentissage classique, l'apprenant peut apprendre n'importe où et n'importe quand et peut consulter différents sujets plus rapidement, on peut résumer quelques autres avantages dans les points suivants :

- L'économie du temps.
- Minimiser les coûts.
- L'E-Learning peut toucher un public plus large.
- Une occasion pour l'apprenant de prendre confiance en soi, de se libérer d'une certaine pression, face à son écran, il dispose de tout son temps (autonomie de l'apprenant)
- Égaliser l'accès à l'éducation à tous les apprenants.
  - Rend les informations et les notes de cours accessibles à tout le monde.
- Faciliter une plus grande interaction entre l'enseignant et ses apprenants.

### **2.2. Les inconvénients d'E-Learning**

Bien que l'apprentissage en ligne présente de nombreux avantages, il présente certains inconvénients qui sont présentés comme suit :

- Une interruption d'internet peut affecter le processus d'apprentissage.
- Une panne dans la plate-forme peut causer l'arrêt du cours ou l'incapacité de l'apprenant d'entrer et télécharger les ressources dont il avait besoin.
- Il y a un manque de contact visuel entre les étudiants et les enseignants.
- L'enseignant n'est pas présent physiquement.
- L'E-Learning nécessite une maîtrise suffisante des outils informatiques et l'internet pour pouvoir suivre la formation.
- L'apprentissage à distance ne convient pas aux apprenants indisciplinés ou aux enseignants inflexibles.

- Manque de contrôle d'enseignant sur les apprenants qui cause parfois la négligence d'apprenant à effectuer ses devoirs.

### **3. Les types d'apprentissage en ligne**

#### **3.1. Synchrone**

L'apprentissage en ligne synchrone exige que tous les apprenants participent aux formations en même temps. Ce type d'E-Learning est généralement interactif et comprend des sessions de discussion sur Internet, des téléconférences, des télés cours et des conférences Web [3].

#### **3.2. Asynchrone**

L'apprentissage en ligne asynchrone est plus flexible par rapport à l'enseignement en ligne synchrone, les instructions asynchrones ne nécessitent pas la participation simultanée de tous les apprenants d'une formation précis, elles sont donc plus flexibles, [3] cela permet une plus grande autonomie pour les apprenants.

#### **3.3. Mode autodirigé**

Utilisation d'outils d'apprentissage informatisés et d'autoévaluation afin d'ajuster au mieux la formation.

### **4. Les caractéristiques d'E-Learning**

L'apprentissage en ligne doit présenter les caractéristiques suivantes [4] :

- L'identification et la gestion des utilisateurs.
- Préparation des contenus de cours en ligne.
- Gestion des cours.
- Suivi et analyse des comportements des élèves.
- Évaluation de l'état d'avancement des élèves.
- La création et la gestion de supports de communication interactifs.

### **5. L'évolution de l'E-Learning**

Comme l'origine du mot E-Learning n'est pas certaine ; il est proposé que le terme soit probablement apparu en 1980 [5]. Dans cette ère numérique et dans plusieurs cas, le E-Learning est devenu une alternative complète à la méthode d'apprentissage traditionnel. Le E-Learning peut être considéré comme une évolution naturelle de la formation à distance.

Il a toujours profité de la technologie moderne pour développer et adapter le cadre de l'outil pédagogique pour façonner l'éducation [6].

Quand l'internet et les ordinateurs personnels sont devenus phénoménaux et ont commencé à prospérer à la fin du 20e siècle, c'est vraiment à ce moment-là que le concept d'apprentissage en ligne a commencé à prendre forme. La technologie, le concept et l'appareil se complétaient bien, offrant une nouvelle tendance d'apprentissage. Le premier système de gestion de l'apprentissage (LMS-SGA) basé sur le Web nommé Cecil a été lancé en 1996 [7]. LMS est une application logicielle qui organise, documente, enregistre et dispense des cours en ligne. Les LMS modernes sont principalement basés sur le Web et permettent l'hébergement et/ou la diffusion de différents types de contenu d'apprentissage (supports de lecture, des vidéos et audio, des wikis, des conférences Web, des discussions, des forums, des blogs, des jeux d'apprentissage, des tests, des évaluations. etc.)

Il est aujourd'hui réalisable d'obtenir des diplômes de célèbres universités mondiales sans quitter la zone géographique ou n'importe quel endroit où l'on dispose d'une connexion haute débit. On peut citer Harvard Berkeley et le MIT qui se sont regroupées pour lancer une plateforme mondiale du E-Learning « plateforme EDX » [8].

## **6. Les différences entre l'apprentissage traditionnel et le E-Learning**

Il existe de nombreuses différences entre le E-Learning et l'apprentissage traditionnel, le tableau suivant représente les différences les plus évidentes :

**Tableau 1 :** Tableau de comparaison entre E-Learning et l'apprentissage traditionnel

<b>Apprentissage en ligne</b>	<b>Apprentissage traditionnel</b>
Se déroule en ligne	Se déroule hors ligne
En tout lieu et à tout moment	Dans un lieu défini, à un horaire défini
Rythme flexible	Rythme imposé
Tout seul	Avec des collègues
Favorise un style d'apprentissage indépendant	Favorise un style d'apprentissage où l'on apprend les uns des autres et les uns avec les autres

La principale source d'informations est en ligne	La principale source d'informations est le formateur
Interaction limitée	Beaucoup d'interaction avec les formateurs et les collègues

## 7. Dispositif de formation à distance

### 7.1. Dispositif de formation

Un dispositif de formation est l'ensemble des moyens techniques, logistiques et humains organisés dans le temps et dans l'espace pour répondre à la demande du commanditaire pour la formation d'une population précise.

### 7.2. Composantes d'un dispositif de formation à distance

Dans la mise en place d'un dispositif de formation, de nombreux services de l'entreprise sont impactés et de nombreux éléments entrent en ligne de compte. La figure ci-dessus représente un schéma qui regroupe les composantes d'un dispositif de formation en ligne.

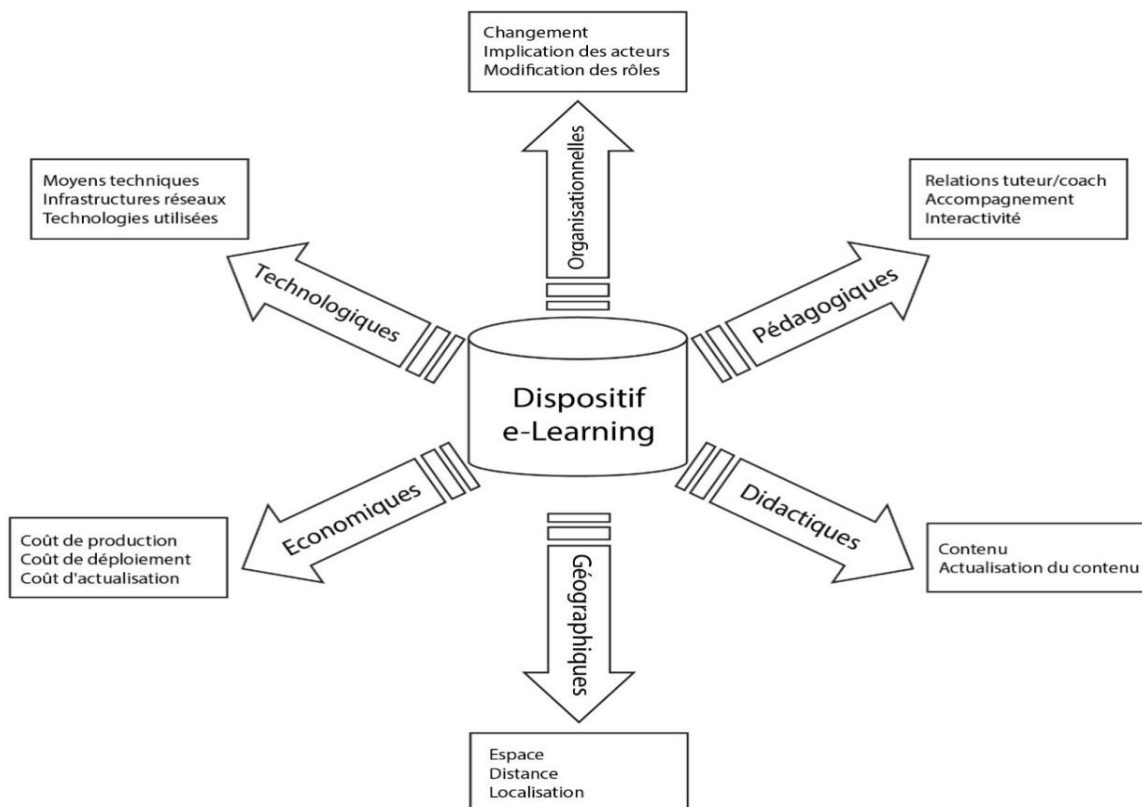


Figure 2 : Les composantes d'un dispositif E-Learning. [9]

M Faisa et al [10] , ont défini les composants comme suit :



- **Composantes organisationnelles** : concerne l'organisation des différents acteurs ainsi que leurs rôles en assurant le changement.
- **Composantes pédagogiques** : modes d'apprentissage centrés sur l'apprenant.
- **Composantes didactiques** : le contenu des modules.
- **Composantes géographiques** : les TICE réduisent à néant les frontières temporelles et géographiques en matière de formation.
- **Composantes économiques** : le coût diffère de la formation présentielle. Il est très important d'évaluer le coût de production, déploiement et d'actualisation des modules E-Learning.
- **Composantes technologiques** : les contraintes techniques sont prépondérantes dans la construction d'un dispositif E-Learning. Une bonne évaluation est conseillée afin que les apprenants ne soient pas découragés par ces dernières.

## **8. Principe du E-Learning**

Le principe du E-Learning s'agit de permettre l'accès au cours via le poste de travail de l'apprenant, sans quitter son endroit, (établissements, classes.) n'existent plus physiquement, l'opération d'apprentissage se déroule dans un système de formation en ligne qui est considérée comme un lieu de rencontre entre formateurs et cours et apprenants.

L'apprenant peut à travers le système consulte des ressources qu'il lui est attribué à différents types, assister à des séances de cours en ligne et consulter les formateurs en cas d'ambiguïté. Le formateur peut effectuer des séances de cours en présence d'un groupe d'apprenants sur le système d'apprentissage, les-envoyé des ressources au cours de séance et répond à leurs questions.

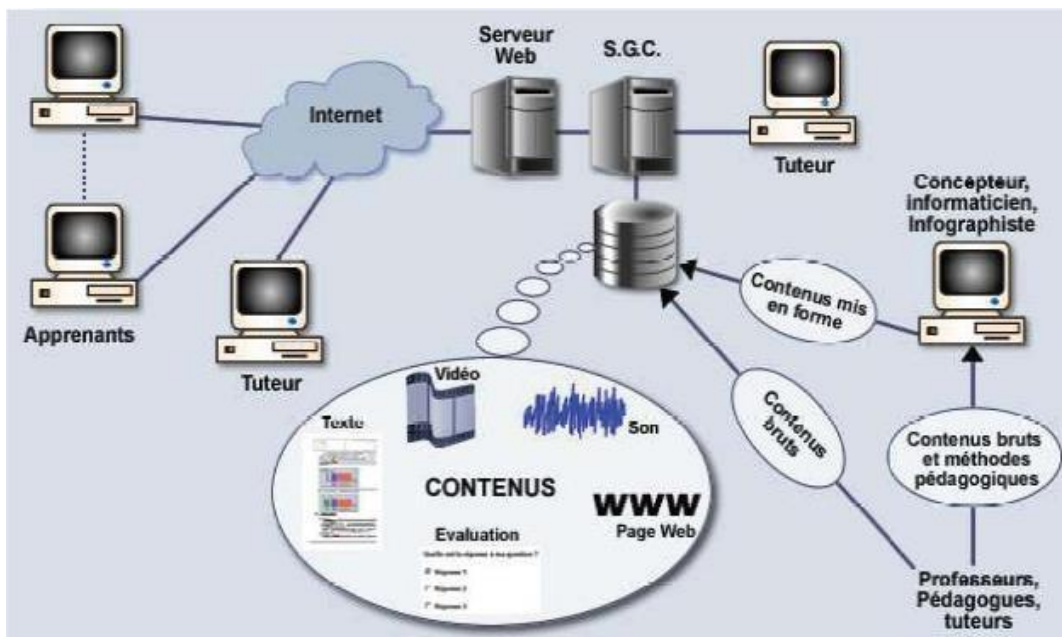


Figure 3 : Principe de E-Learning. [11]

## 9. Conclusion

Dans ce chapitre nous avons présenté l'enseignement en ligne le E-Learning en général, on a commencé par définir ce domaine, ensuite on a présenté ses avantages et ses inconvénients. Après ça, on a donné les caractéristiques du E-Learning passons par ses types.

Nous avons conclu notre chapitre par la présentation des composants essentiels d'un dispositif de formation en ligne et les principes de l'E-Learning. Il apparait clairement que le E-Learning a un impact considérable sur le système éducatif dans la dernière période.

---

***CHAPITRE 2***  
***Systeme de recherche***  
***d'information &***  
***apprentissage automatique***

---

Depuis le début du XXI<sup>e</sup> siècle la quantité de ressources et des articles pédagogiques augmentent d'une manière accélérée, Cependant l'accès à l'information pertinente, adaptée aux besoins d'un utilisateur donné devient à la fois difficile et nécessaire. Pour cela, de nombreux systèmes de Recherche d'Informations (RI) ont été développés pour traduire une requête (besoin d'utilisateur) vers un ensemble de document, ces systèmes se basent sur diverses approches dont le but est d'automatiser le processus de la recherche d'information (RI).

## **1. Système de recherche d'information**

### **1.1. Définition de la recherche d'information**

La recherche d'information est le domaine qui étudie la manière de retrouver des informations dans un corpus. Celui-ci est composé de documents d'une ou plusieurs bases de données, qui sont décrits par un contenu ou les métadonnées associées. Les bases de données peuvent être relationnelles ou non structurées, telles celles mises en réseau par des liens hypertexte comme dans le World Wide Web, l'internet et les intranets. Le contenu des documents peut être du texte, des sons, des images ou des données.

La recherche d'information est historiquement liée aux sciences de l'information et à la bibliothéconomie qui visent à représenter des documents dans le but d'en récupérer des informations, au moyen de la construction d'index. L'informatique a permis le développement d'outils pour traiter l'information et établir la représentation des documents au moment de leur indexation, ainsi que pour rechercher l'information. La recherche d'information est aujourd'hui un champ pluridisciplinaire, intéressant même les sciences cognitives. [12]

### **1.2. Définition d'un système de recherche d'information**

D'après [13] [14] [15], un Système de Recherche d'Information (SRI) est un système capable de stocker, de récupérer et de maintenir des informations, ce

système s'appuie sur le mécanisme de recherche pour effectuer la tâche de récupération des documents pertinents qui peuvent répondre au besoin en information de l'utilisateur formulé par une requête de recherche sous forme d'un langage naturel. Ou peut être sous une composition de requêtes basiques avec un ensemble d'opérateurs logiques (ET, OU, SAUF, etc.) appelée requête logique.

### **1.3. Définition de la pertinence**

La RI est définie comme la discipline qui trouve des documents pertinents par opposition à de simples correspondances avec des modèles lexicaux dans une requête. Cela souligne un aspect fondamental de la RI, c'est-à-dire que la pertinence des résultats est évaluée par rapport au besoin d'information et non à la requête. [16]

Les concepteurs de SRI distinguaient deux types de pertinence : la pertinence système, est définie comme étant une correspondance entre un document et une requête et la pertinence utilisateur qui se traduit par des jugements de pertinence sur les documents fournis en réponse à une requête.

#### **1.3.1. Calcul de pertinence**

Le TF-IDF est une méthode de pondération souvent utilisée en recherche d'information et en particulier dans la fouille de textes. Cette mesure statistique permet d'évaluer l'importance d'un terme contenu dans un document, relativement à une collection ou un corpus.

$TF-IDF = TF * IDF$  (Term Frequency \* Inverse Document Frequency) *Equation 1*

$TF(t) = (\text{Nombre de fois où le terme } t \text{ apparaît dans un document}) / (\text{Nombre total de termes dans le document})$  *Equation 2*

$IDF = \text{Log}(\text{le nombre total de documents dans la collection}) / (\text{le nombre de documents contenant le terme } t)$  *Equation 3*

## **2. Les modèles de recherche d'information**

Les modèles de RI sont classés comme suit :

## **2.1. Modèle booléen**

Le modèle booléen permet l'utilisation d'opérateurs d'algèbre booléenne, ET, OU et NON, pour la formulation de requêtes. Dans le modèle booléen, un document est associé à un ensemble de mots-clés. Les requêtes sont également des expressions de mots-clés séparés par AND, OR ou NOT/BUT. La fonction de récupération dans ce modèle traite un document comme pertinent ou non pertinent. [17]

## **2.2. Modèle spatial vectoriel**

Le modèle d'espace vectoriel peut être mieux caractérisé par sa tentative de classer les documents par la similitude entre la requête et chaque document [18]. Dans le modèle d'espace vectoriel (MEV), les documents et la requête sont représentés comme un vecteur et l'angle entre les deux les vecteurs sont calculés en utilisant la fonction cosinus de similarité.

## **2.3. Modèle probabiliste**

La caractéristique la plus importante du modèle probabiliste est sa tentative de classer les documents selon leur probabilité de pertinence pour une requête [19]. Les documents et les requêtes sont représentés par des vecteurs binaires  $\tilde{d}$  et  $\tilde{q}$ , chaque élément vectoriel indiquant si un attribut ou un terme de document apparaît ou non dans le document ou la requête. Au lieu de probabilités, le modèle probabiliste utilise les cotes  $O(R)$ , où  $O(R) = P(R)/1 - P(R)$ ,  $R$  signifie « le document est pertinent » et «  $\bar{R}$  signifie que le document n'est pas pertinent ». [20]

## **3. Processus de recherche d'information**

Les systèmes de recherche d'informations ont connu une grande évolution et amélioration depuis leur première apparition dans les années 1950, les processus de recherche d'information sont les suivants :

### **3.1. Collection de documents**

Avant la mise en fonction du système, la présence d'un ensemble des documents numériques manipulable de divers domaines nécessaires.

### **3.2. L'indexation**

L'indexation est la représentation des documents et des requêtes à travers un ensemble de règles et de notions permettant la traduction d'une requête ou d'un document, d'une description brute vers une description structurée. L'indexation a pour rôle d'extraire à partir d'un document ou d'une requête, une représentation paramétrée qui couvre au mieux son contenu sémantique.

On distingue trois types d'indexation sont les suivants :

#### **3.2.1. L'indexation manuelle**

Chaque document est analysé par un documentaliste ou par un spécialiste du domaine, qui extrait les mots basant sur un vocabulaire contrôlé (liste hiérarchique, thésaurus, lexicque, ...) [21], elle permet d'assurer un meilleur rapport entre les documents et les termes choisis par les spécialistes [22] ,mais l'inconvénient, elle est couteuse en termes de temps (nécessite un temps important) et plus d'effort intellectuel (nombre de personnes).

#### **3.2.2. L'indexation semi-automatique**

Appelée aussi indexation supervisée. Ce type d'indexation fait une combinaison des deux modes précédant [23], les termes du document sont extraits en premier par un processus automatique. Mais le choix final des mots clés est fait par l'indexeur ou le spécialiste du domaine, généralement les indexeurs utilisent un vocabulaire contrôlé sous forme de thésaurus ou de base terminologique [22].

#### **3.2.3. L'indexation automatique**

L'indexation dans ce cas-là est faite par un SRI basé sur des algorithmes et des méthodes, l'expert du domaine n'intervient pas [23]. Elle détecte d'une façon automatique les concepts significatifs d'un document en analysant le document mot par mot aussi

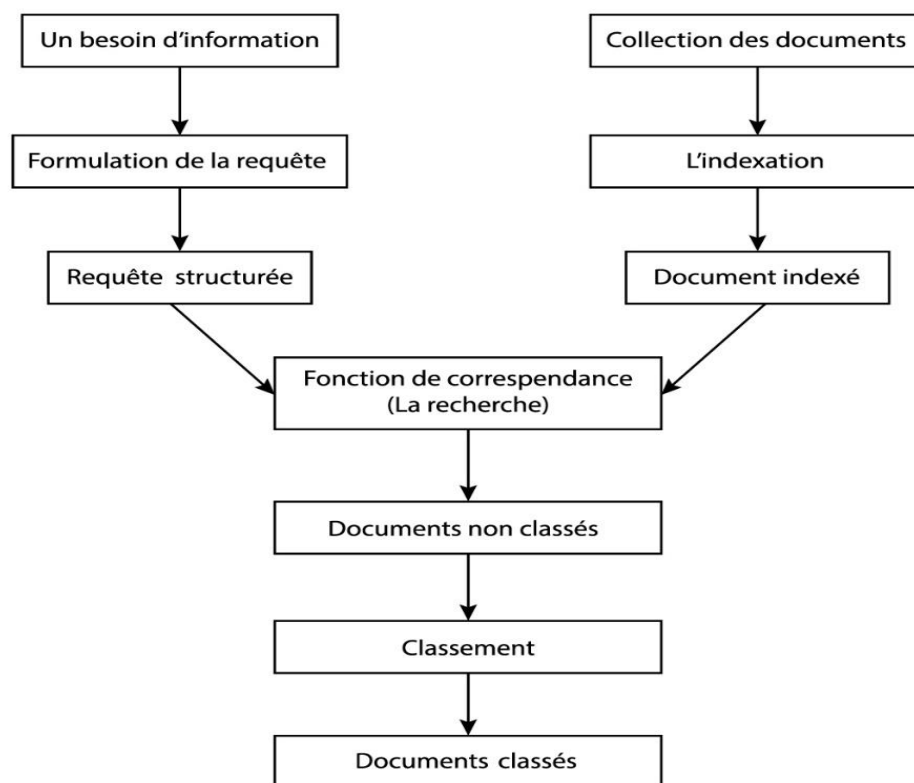
l'élimination des mots vides, la lemmatisation, la pondération des termes, à la fin la création de l'index. Ce type d'indexation est souvent le plus utilisé [24].

### 3.3. La recherche

La recherche est effectuée en faisant la correspondance entre les documents indexés et les mots clés de la requête, tous les documents qui contiennent les occurrences des mots clés seront récupérées.

### 3.4. Le classement

Les documents sont triés en fonction de la qualité de correspondance de telle sorte que les documents les plus pertinents seront affichés en premier, le classement dépend du modèle de RI. La figure 4 ci-dessous présente les processus de recherche.





**Figure 4** : Processus de recherche d'information [25]

#### **4. Les objectifs d'un système de recherche d'information**

- Les objectifs principaux d'un système de recherche d'informations d'après [14] sont :
- Minimiser la surcharge d'un utilisateur pour localiser les informations nécessaires.
- Fournir un support de génération de recherche d'utilisateurs.
- Assister l'utilisateur automatiquement et via l'interaction du système à développer une spécification de recherche qui représente le besoin de l'utilisateur et le style d'écriture de divers auteurs.
- Récupérer un ensemble de ressources pertinentes.

#### **5. Evaluation d'un système de recherche d'information**

La qualité d'un système doit être mesurée en comparant les réponses du système avec les réponses idéales que l'utilisateur espère recevoir. La comparaison des réponses d'un système pour une requête avec les réponses idéales nous permet d'évaluer les deux métriques suivantes :

- **Précision**

La précision mesure la proportion de documents pertinents retrouvés parmi tous les documents retrouvés par le système.

$$\text{Précision} = \frac{\text{Nombre de documents pertinents retrouvés}}{\text{Nombre de documents retrouvés}} \quad \text{Equation 4}$$

- **Rappel**

Le rappel mesure la proportion de documents pertinents retrouvés parmi tous les documents pertinents dans la base.

$$\text{Rappel} = \frac{\text{Nombre de documents pertinents retrouvés}}{\text{Nombre de documents pertinents}} \quad \text{Equation 5}$$

### 5.1. Mesures alternatives

D'autres mesures essaient de combiner les mesures de rappel et de précision afin de sortir avec une seule valeur. Parmi les mesures proposées nous pouvons citer : la mesure harmonique et la mesure d'évaluation appelé *F - mesure*.

#### 5.1.2. Mesure harmonique ( F-mesure)

Introduite par van Rijbergen, 1979 , la mesure harmonique *H* (appelée aussi *F - mesure*) ou *F - score* est une fonction qui combine les valeurs de rappel et de précision en une seule valeur (moyenne) incluse dans l'intervalle [0,1]. En effet, rappel et précision ayant tendance à varier en sens inverse.

La mesure harmonique mesure la capacité du système à retrouver tous les documents pertinents et à rejeter les autres.

$$H(j) = F - \text{mesure} = \frac{2}{\frac{1}{R(j)} + \frac{1}{P(j)}} = \frac{2 * R(j) * P(j)}{R(j) * P(j)} \quad \text{Equation 6}$$

Où : *R(j)* et *P(j)* représentent respectivement le rappel et la précision du *j* ème document renvoyé par le système. Cette mesure tend vers 0 lorsqu'aucun document pertinent n'est restitué et égale à 1 lorsque tous les documents restitués sont pertinents.

La valeur de *F - mesure* diminue quand l'un de ses paramètres est petit et elle augmente quand les deux paramètres sont proches tout en étant élevés.

#### 5.1.3. Mesure d'évaluation « E-Mesure » (ou F-Mesure paramétrique) :

Le but de cette mesure est de permettre à l'utilisateur de spécifier laquelle des valeurs de rappel et de précision est plus intéressante. La mesure est ainsi définie par :

$$E(j) = H_\beta = F_\beta - \text{Mesure} = \frac{1+\beta^2}{\frac{\beta^2}{R(j)} + \frac{1}{P(j)}} = \frac{(1+\beta^2)*P(j)*R(j)}{\beta^2 * P(j)+R(j)} \quad \text{Equation 7}$$

Le paramètre  $\beta > 0$  est un paramètre de l'utilisateur qui permet de spécifier l'importance du rappel et de précision. Il permet de contrôler le compromis  $R, P$ .

Si  $\beta = 1$ , même poids précision et rappel  $E(j) = H(j)$ .

Si  $\beta < 1$ , l'utilisateur favorise la précision.

Si  $\beta > 1$ , l'utilisateur favorise le rappel.

## 6. Système de recherche d'information sémantique

Un nombre croissant de systèmes de recherche d'informations récents utilisent des ontologies pour aider les utilisateurs à clarifier leurs besoins en informations et à proposer des représentations sémantiques des documents.

### 6.1. Définition d'ontologie

Neches et al [26] ont défini l'ontologie : « 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 la plus fréquente est celle de Gruber [27] : « une ontologie est une spécification explicite d'une conceptualisation ».

Elle a été modifiée par Borst [28] : « une spécification explicite et formelle d'une conceptualisation partagée ».

Studer et al [29], ont expliqué en détail la définition de l'ontologie :

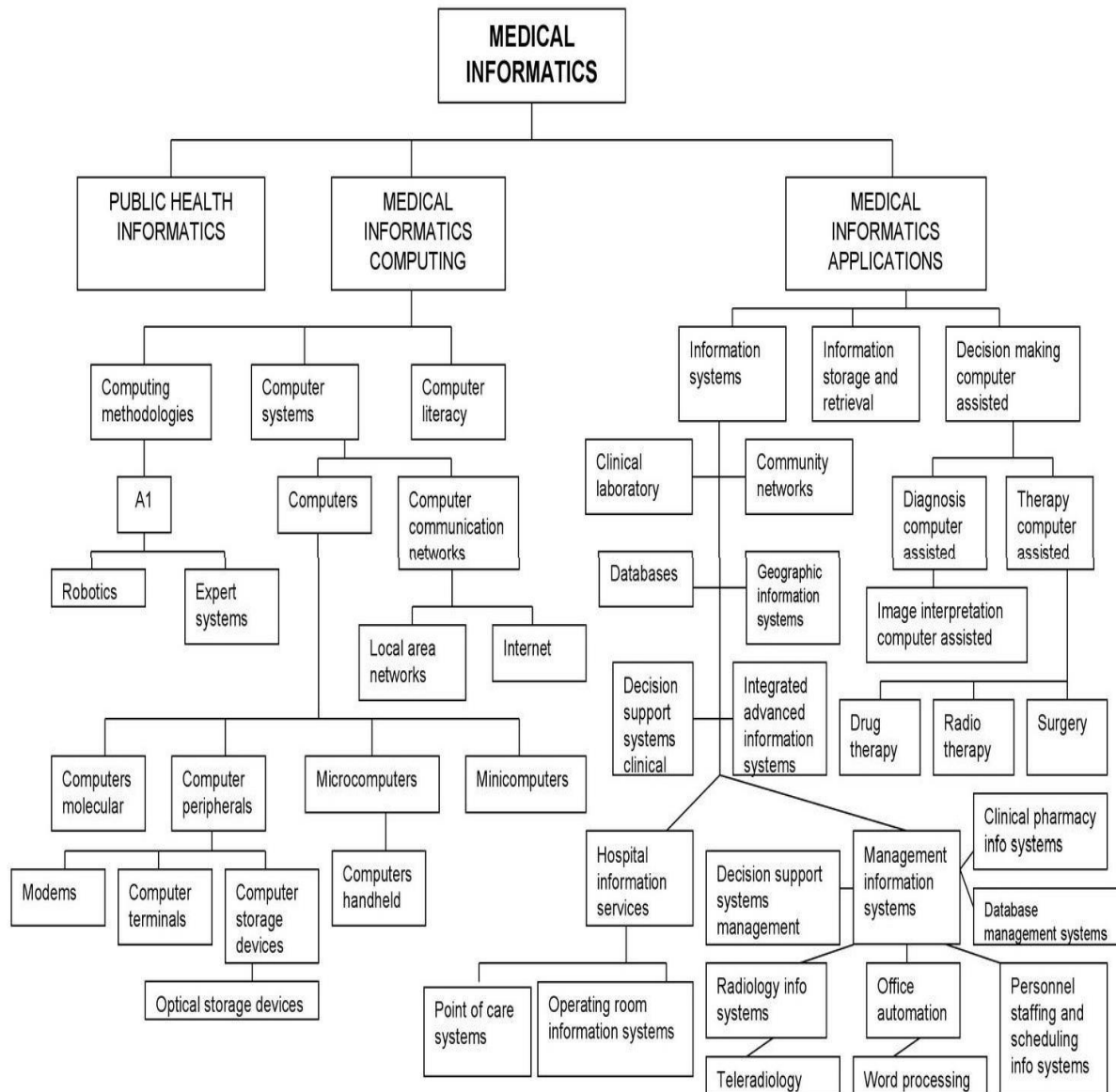
- **Explicite** signifie que le type des concepts et les contraintes sur leurs utilisations sont explicitement définis,
- **Formelle** se réfère au fait que la spécification doit être lisible par une machine.

- **Partagée** se rapporte à la notion, selon laquelle une ontologie capture la connaissance consensuelle, 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.

## **6.2. Exemple d'ontologie**

L'ontologie présentée dans la figure 6 est un thésaurus médical MeSH « Medical Subject Headings », une liste de termes ayant entre eux des relations hiérarchiques, synonymiques et de proximité. Les relations hiérarchiques sont celles du descripteur au sein de son arborescence.

Figure 5 : une partie d'ontologie de thésaurus MeSH. [30]



### 6.3. L'ontologie en Recherche d'information

La recherche d'information se tourne vers les ontologies et les modèles assimilés pour caractériser le contenu (la sémantique) des documents et les besoins des utilisateurs afin d'améliorer les approches classiques.

Un traitement plus efficace des documents et la restitution de résultats plus près à leur centre d'intérêt (apporter un meilleur ciblage des besoins) par une sélection plus pertinente au sein des métadonnées (grande source de données). [31]

Une représentation riche, précise et efficace des documents, une indexation sémantique qui s'appuie sur les relations et les concepts entretiennent dans un modèle pour donner du sens au contenu de ces documents.

L'amélioration d'indexation en distinguant les occurrences d'un terme qui correspondent à des sens différents et en regroupant les mots de même sens et d'une même idée pour retrouver des documents n'utilisant pas exactement les termes de l'utilisateur, mais répondre par des documents similaires à son besoin.

#### **6.4. L'indexation sémantique**

L'utilisation d'une ontologie lors de la phase d'indexation permettrait de lever les ambiguïtés des sens des termes utilisés et de mieux représenter les connaissances inhérentes dans les documents. En termes d'indexation sémantique, des concepts de l'ontologie sont associés à chaque document selon les sémantiques qui y sont véhiculées.

### **7. L'apprentissage Automatique**

L'apprentissage automatique, apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmées pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes. [32]

Selon [32] les types d'apprentissage automatique sont les suites :

#### **7.1. Apprentissage supervisé**

L'apprentissage supervisé a pour but d'établir des règles de comportement à partir d'une base de données contenant des exemples de cas déjà étiquetés. La base de données est en principe un ensemble de couples entrées / sorties  $\{(X, Y)\}$ . Le but est d'apprendre à prédire pour toute nouvelle entrée  $X$ , la sortie  $Y$  [33], l'apprentissage supervisé utilise plusieurs méthodes telles que :

- **L'arbre de décision**

L'apprentissage par arbre de décision est une méthode classique en apprentissage automatique [34]. Son but est de créer un modèle qui prédit la valeur d'une variable-cible depuis la valeur de plusieurs variables d'entrée.

- **Régression linéaire**

Un modèle de régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, dite expliquée et une ou plusieurs variables, dites explicatives. [32]

## **7.2. Apprentissage non supervisé**

Contrairement à l'apprentissage supervisé, le non-supervisé traite le cas où on dispose seulement des entrées  $\{X\}$  sans avoir au préalable les sorties. L'apprentissage non supervisé ou le « clustering » vise à construire des groupes (clusters) d'objets similaires à partir d'un ensemble hétérogène d'objets [35].

On distingue plusieurs algorithmes de « clustering », comme [35]:

- **K-moyennes**

KMeans est un algorithme de partitionnement des données en K nombre de groupes ou clusters. Chaque objet sera associé à un seul cluster. Le nombre K est fixé par l'utilisateur.

- **Fuzzy KMeans**

Il s'agit d'une variante du précédent algorithme proposant qu'un objet ne soit pas associé qu'à un seul groupe.

## **7.3. Apprentissage semi-supervisé**

L'apprentissage semi-supervisé est une approche de l'apprentissage automatique qui combine une petite quantité de données étiquetées avec une grande quantité de données non étiquetées pendant l'entraînement.

## **7.4. Apprentissage partiellement supervisé**

Probabiliste ou non, quand l'étiquetage des données est partielle. C'est le cas quand un modèle énonce qu'une donnée n'appartient pas à une classe A, mais peut-être à une classe B ou C (A, B et C étant trois maladies par exemple évoqués dans le cadre d'un diagnostic différentiel).

### **7.5. Apprentissage par renforcement**

L'algorithme apprend un comportement étant donné une observation. L'action de l'algorithme sur l'environnement produit une valeur de retour qui guide l'algorithme d'apprentissage.

### **7.6. Apprentissage par transfert**

L'apprentissage par transfert peut être vu comme la capacité d'un système à reconnaître et appliquer des connaissances et des compétences, apprises à partir des tâches antérieures, sur des nouvelles tâches ou domaines partageant des similitudes.

## **8. Conclusion**

Dans ce chapitre on a défini les concepts de base d'un système de recherche d'information plus particulièrement les processus de la RI, les modèles, les objectives de RI, et les méthodes d'évaluation. Ensuite, nous avons parlé sur les apports d'ontologie dans les SRI.



---

# *CHAPITRE 3*

## *Travaux connexes*

---

Dans ce chapitre, on va étudier les approches proposées dans la littérature consacrée au système de recherche pour le E-Learning. Cette étude permet de recenser les avantages et les inconvénients des approches proposées toute en commençant par la conception de chaque système passant par la structure vers le noyau.

### **1. Approche classique**

L'approche classique consiste à considérer une requête comme une chaîne de caractère et un document pertinent comme celui qui contient cette chaîne de caractère. À partir de cette vision simpliste, on peut imaginer l'approche qui consiste à balayer les documents indexés séquentiellement, en les comparent avec la chaîne de caractère qui est la requête (requête indexé). Si on trouve la même chaîne de caractère dans un document, alors il est sélectionné comme réponse. Cette approche est évidemment très simple à réaliser.

L'opération de recherche est très lente. Pour chaque requête, on doit parcourir toute la base de documents. Bien que cette base soit volumineuse. En effet cette approche n'est utilisée que sur des collections très limitées (des dizaines jusqu'aux centaines de documents).

Recherche d'information dans des sources distribuées [36] est l'une des approches classiques.

#### **1.1. Méthode de l'indexation**

Cette méthode consiste à effectuer des pré-traitements sur les documents et les requêtes, ce qu'on appelle l'indexation. Cette opération vise à construire une structure d'index qui permet de retrouver très rapidement les documents incluant des mots demandés.

La structure d'index est de la forme suivante (la structure de fichier inversé) :

$$\text{Mot} \rightarrow \{\dots, \text{Doc}, \dots\}$$

C'est-à-dire, chaque mot est mis en correspondance avec les documents qui le contiennent.

Une requête peut être maintenant une expression plus complexe, incluant des opérateurs logiques (ET, OU, ...) ou d'autres types d'opérateurs. On commence par évaluer les éléments de base (par exemple : les mots) dans la requête, obtenant ainsi des listes de documents, qui se combine, selon l'opérateur qui relie ces éléments pour obtenir finalement une seule liste de documents.

### **2. L'approche sémantique**

La recherche sémantique a pour objectif d'améliorer la précision de recherche par la compréhension approfondie des requêtes des utilisateurs, par la machine et la signification exacte des termes tels qu'ils apparaissent dans le corpus, afin de générer des résultats plus pertinents.

La sémantique a été introduite à différents niveaux et à différentes étapes du processus de RI. Cependant, il convient de souligner que toute tentative d'apporter de la sémantique doit permettre d'équilibrer la quantité de traitement complexe du langage naturel nécessaire à l'augmentation des performances de RI.

En présente quelques travaux qui sont basés sur l'approche sémantique dans la recherche.

SPARK [37] est un système de récupération qui peut interroger des données sémantiques avec des requêtes SPARQL générées à partir de mots-clés. Il se compose de trois étapes principales : la cartographie des termes, la construction du graphique de requête et le classement des requêtes. L'étape de mappage de termes essaie de faire correspondre les termes de la requête aux ressources ontologiques.

Q2Semantic [38] essaie de combler le fossé entre les requêtes par mot-clé et les requêtes formelles. Les auteurs traitent trois problèmes pour atteindre leur objectif, à savoir l'appariement des termes, le classement et l'évolutivité.

Le premier problème est abordé en enrichissant les requêtes des utilisateurs avec les termes extraits de Wikipédia, afin que les termes de la requête soient facilement mis en correspondance avec les entités de l'ontologie. Pour résoudre le deuxième problème, ils implémentent un mécanisme de classement qui prend en compte de nombreux facteurs tels que la longueur de la requête, la pertinence et l'importance des éléments de l'ontologie, etc.

Enfin, pour le problème d'évolutivité, ils proposent une structure de graphe groupé pour représenter des résumés de graphes RDF. Mais ils ont toujours besoin des procédures coûteuses de construction de graphes et de parcours au moment de l'exécution.

SemSearch [39] est un système similaire qui génère des requêtes formelles à partir des mots-clés. Il diffère des systèmes ci-dessus de la manière dont il construit les requêtes. Il utilise un certain nombre de modèles prédéfinis qui sont remplis de diverses combinaisons de termes de requête. Étant donné que chaque mot-clé de la requête peut être une classe, une instance ou une propriété, le nombre de combinaisons est assez important, ce qui affecte négativement les temps de réponse de la requête. Lorsque les requêtes sont générées, elles sont utilisées directement pour rechercher les données sémantiques. Le classement est effectué après la récupération des résultats de toutes les requêtes.

### **3. La méthode d'indexation sémantique**

Une alternative évolutive à la construction de requêtes à partir de mots-clés, est l'indexation sémantique. Dans cette approche, les données sémantiques des bases de connaissances RDF sont indexées de manière structurée et rendues directement disponibles pour être utilisées avec des requêtes par mot-clé.

OWLIR [40] est un système de recherche sémantique qui adapte une approche similaire. Il se compose d'un module d'extraction d'informations, d'un module d'inférence et d'un module de récupération. Ils utilisent le système AeroText1 pour extraire des phrases clés du texte libre et les représenter sous forme de triplets RDF. Ces données sont enrichies de raisonnements et de règles sémantiques pour obtenir des triplets RDF inférés. Enfin, ces triplets sont indexés avec le texte libre correspondant. Les détails du mécanisme d'indexation et du classement sont omis. Dans la partie évaluation, ils montrent les améliorations en comparant les trois indices : texte libre, texte libre avec triplets RDF et texte libre avec triplets RDF inférés.

Squiggle [41] est un Framework de récupération qui utilise l'indexation sémantique pour un accès efficace aux connaissances sémantiques. Au lieu d'utiliser toute la base de connaissances pour la recherche, il indexe et recherche uniquement les entités reconnues

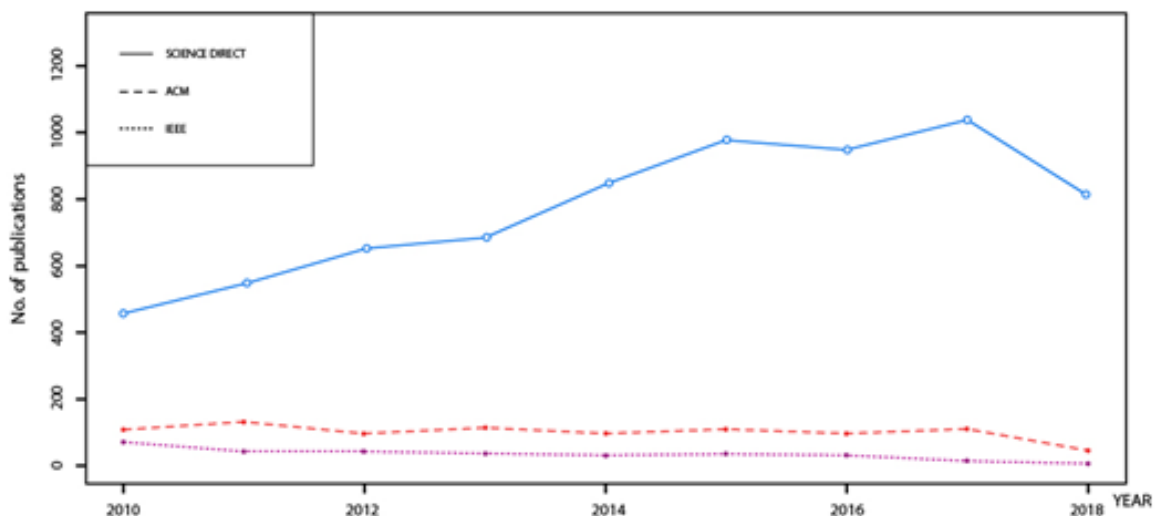
lors de la phase d'annotation sémantique. L'unité d'indexation est à nouveau le triple. Contrairement aux systèmes décrits ci-dessus, Squiggle conserve le texte libre dans un index séparé et la requête soumise par l'utilisateur est recherchée dans les deux index. La récupération se fait principalement à l'aide de l'index de texte libre et l'index sémantique est utilisé pour faire des suggestions à l'utilisateur en fonction des connaissances ontologiques.

#### 4. Travaux existants dans le domaine du E-Learning

Selon [42], à partir de l'an 2000, il y a eu une augmentation continue de l'intérêt des chercheurs concernant le développement et l'utilisation des ontologies en général, comme le montre l'augmentation du volume de publications ayant le mot « ontologie » dans son titre ou la liste des mots clés. Plus précisément, comme le montre la figure 7, l'utilisation des ontologies dans les systèmes d'apprentissage en ligne a également suscité un intérêt accru, tel qu'indiqué dans les recherches menées sur quatre (4) bibliothèques numériques : IEEE, ACM, ScienceDirect.

Figure 6 : Croissance de nombre des publications pour "ontologie", "E-Learning". [42]

Tous les travaux connexes attribuent une sémantique au contexte ou aux ressources de l'apprentissage et l'objectif principal visé est d'améliorer les résultats de recherches par des systèmes de recherche ou des systèmes de recommandation basée sur une ontologie, certaines approches ont focalisé sur le profil de l'apprenant et d'autre aux ressources



pédagogiques.

Les systèmes de recommandation sont une forme spécifique de filtrage de l'information visant à présenter les éléments d'information qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence et cherche à prédire l'« avis » que donnerait un utilisateur [32].

Gina et al [43] ont déclaré que les systèmes de recommandation utilisent l'ontologie et des méthodes de l'intelligence artificielle pour fournir des recommandations personnalisées. L'ontologie est utilisée pour modéliser les apprenants et les ressources d'apprentissage et permet de récupérer des détails et prennent en charge les mécanismes d'inférence, ce qui aide à fournir des recommandations améliorées.

Dans le but de proposer une solution au problème de la découverte de Ressources en ligne par différentes catégories d'utilisateurs, Olaperi yeside sowunmi & al [44] présente un cadre pour la recherche d'informations personnalisée de contenu éducatif. Le Framework exploite les technologies du web sémantique. Ces technologies comprennent : la pile Web sémantique, l'annotation sémantique, les vocabulaires (ontologies), les graphiques de connaissances et de données ouvertes liées et la recherche d'informations sémantiques.

Finalement, Rachid Ahmed-Ouamer & al [45] ont proposé une approche qui se base sur l'utilisation d'une ontologie de domaine pour l'indexation d'une base de documents et l'utilisation des liens sémantiques entre documents ou fragments de documents de la collection, pour permettre l'inférence de tous les documents pertinents. Cette approche est testée sur le domaine de l'E-Learning de l'informatique dans le contexte du web sémantique.

Dans cette approche le profil d'utilisateur ou l'apprenant est divisé en deux classes (apprenant, facilitateur) dans ce cadre ils ont utilisé une ontologie simple pour définir le profil de l'apprenant.

#### **4.1. Comparaison des approches**

Le but des ontologies utilisées dans toutes les approches sont d'ajouter un sens aux Informations sur le Web (contenu éducatif dans notre cas), ils représentent les connaissances de manière à ce que des informations puissent être déduites.

On peut baser sur plusieurs critères pour faire une comparaison de ces approches étudiées.

Parlons du noyau de tous les travaux : la base de connaissance ou l'ontologie, certaines approches se basent sur le profil des apprenants plus que les ressources d'apprentissage, l'ontologie est construite pour faire la différence entre les profils d'utilisateurs finaux, dans les systèmes de recommandation, cette ontologie est fréquemment utilisée, car le but de ce système est de donner un résultat de recherche recommandé pour un utilisateur spécifique ici on remarque que la recherche dépend du personnel qui utilise le système.

L'ontologie des ressources (documents ou fragments) est toujours utilisée pour donner du sens au contenu.

D'autres approches ont considéré que ces ressources pédagogiques sont les plus essentiels pour donner des résultats de recherche plus pertinents dans le domaine E-Learning là on parle sur les systèmes de recherche.

Une ontologie plus détailler des ressources est utilisée avec des ontologies de l'E-Learning sont fusionnées pour donner une base de connaissance vaste ou large, pour donner plus de sémantique ils ont compté sur des moteurs d'inférences, il existe aussi d'autres qui ont utilisé l'ontologie du domaine d'enseignement pour améliorer la précision du système.

Finalement, dans cette comparaison on retient que les différences fondamentales entre les approches de recherche ou recommandation sont les ontologies utilisées (l'ontologie des ressources pédagogiques, ontologie de profile d'apprenant, l'ontologie de l'E-Learning et l'ontologie du domaine...etc.) et les règles d'inférences.

On peut résumer cette comparaison par le Tableau 2 qui montre la différence entre les approches :

**Tableau 2:** Comparaison entre les approches existantes.

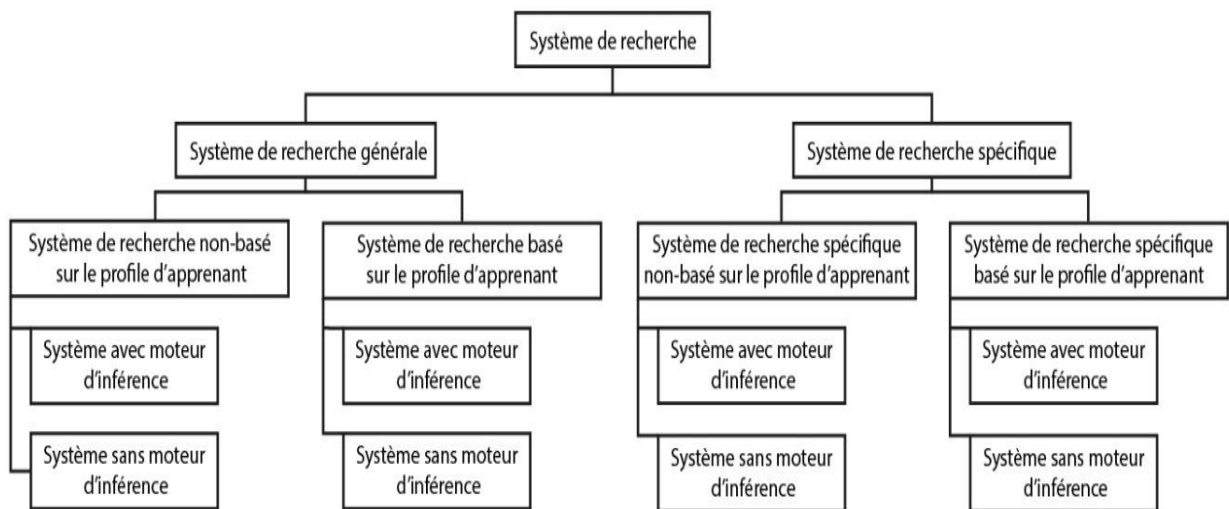
Approches	Type du système	Profile de l'apprenant	Ontologies utilisé	Moteur d'inférence
Ontology-based recommender systems in E-Learning [43]	Système de recommandation	Très intéressant	Dépend du modelé de système recommandation	Non utilisé

A semantic web-based framework for RI in E-Learning system [44]	Système d'apprentissage qui intègre un moteur de recherche	Intéressent	-Profil d'apprenant -Objet d'apprentissage	Utilisé
Un système de recherche d'information pour l'E-Learning [45]	système de recherche	S'intéresse au niveau	-Profil d'apprenant -Ressources pédagogique -Domaine d'enseignement (instance du domaine informatique)	Utilisé

### 4.2. Classement des approches relative à l'E-Learning

On s'intéresse au système de recherche d'information relative à l'E-Learning et on les classes selon notre étude de comparaison.

Le schéma ci-dessous présente une classification des SRI relative à l'E-Learning :



*Figure 7: Classification des approches existantes (basé sur nos études).*

## 5. Conclusion



Après cette étude on va exploiter les faiblesses de chaque système en utilisant la comparaison précédente et fusionner la qualité pour donner un résultat plus pertinent et plus évolué.

---

***Partie 2***  
***Contribution***

---

---

*CHAPITRE 4*

*Conception*

---

Aujourd'hui, la satisfaction des besoins en information d'un apprenant demeure un objectif très important à atteindre pour les systèmes de recherche d'information (SRI) dans le domaine E-Learning. C'est dans ce contexte, que nous avons choisi d'adopter l'intégration d'une ontologie des objets pédagogiques dans le processus de recherche d'information.

L'ontologie des objets pédagogiques constitue un facteur clé pour organiser les ressources dans la meilleure structure possible. À cet effet, nous allons ce qui suit détailler la solution proposée dans notre projet. La solution est basée sur cette ontologie qui améliore grandement les résultats de recherche.

### 1. L'architecture globale du système

Dans ce paragraphe, nous proposons notre Conception de l'architecture générale du système, mettant en évidence le détail de fonctionnement tel qu'illustrée par la figure 9 ci-dessous.

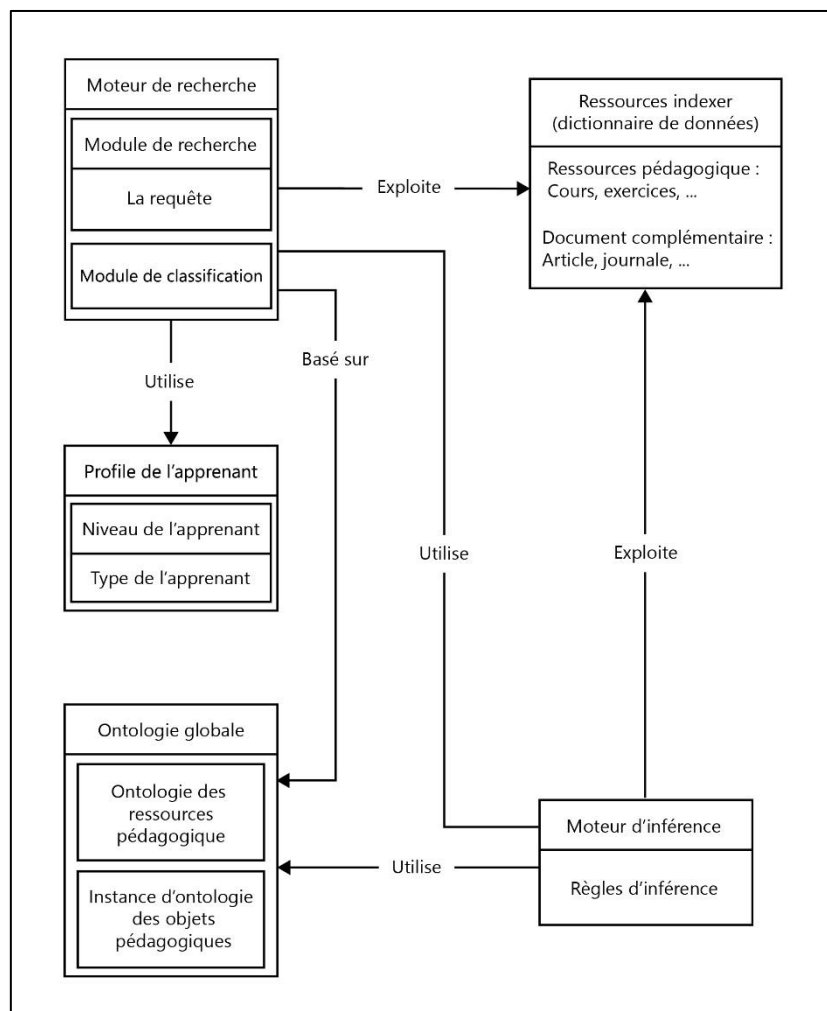


Figure 8 : Architecture générale du système.

### **1.1. Le moteur de recherche**

Le fonctionnement du moteur de recherche est basé essentiellement sur le module de recherche ainsi que sur d'autres éléments du système tels que : l'ontologie globale, le niveau de l'apprenant, la base de ressource indexée et le moteur d'inférence.

### **1.2. Le module de recherche**

A travers le module de recherche, l'apprenant peut exprimer son besoin en information par une requête formulée en langage naturel, cette requête va être indexée et exploitée pour récupérer les documents pertinents par rapport au centre d'intérêt exprimé par la requête.

### **1.3. Base de ressources indexées**

La base des ressources indexées stocke une énorme quantité de documents indexés de manière organisée à même de faciliter leur exploitation par le moteur d'inférence pendant le processus de la recherche.

### **1.4. Moteur d'inférence**

C'est un mécanisme qui permet de conduire à d'autres liens sémantiques à partir d'une base de fait initiale pour récupérer plus de ressources pertinentes. Il est utilisable par le moteur de recherche durant l'opération de recherche pour récupérer plus de ressource possible.

### **1.5. Module de classification des requêtes**

Le module qu'on a ajouté dans ce système est basé sur l'apprentissage supervisé qui utilise l'algorithme de classification, la raison principale d'utilisation de ce module est que l'évaluation du niveau est difficile pour l'apprenant. Alors, le principe consiste à entraîner la machine avec un ensemble des données.

Ces données représentent un ensemble des requêtes attachées avec le niveau de l'utilisateur,

Ce Module renforce la recherche et le rend plus précis avec une assistance de l'apprenant à partir de sa requête d'interrogation.

Ce module est construit à l'aide de l'algorithme SVM de classification, le détail ci-dessous montre le déploiement de l'algorithme

### **1.5.1. Définition du SVM**

« Support Vector Machine » est un algorithme d'apprentissage automatique supervisé qui peut être utilisé à la fois pour des défis de classification ou de régression. Cependant, il est principalement utilisé dans les problèmes de classification. Dans l'algorithme SVM, nous traçons chaque élément de données comme un point dans un espace à  $n$  dimensions (où  $n$  est le nombre de caractéristiques que vous avez) avec la valeur de chaque caractéristique étant la valeur d'une coordonnée particulière. Ensuite, nous effectuons la classification en trouvant l'hyper-plan qui différencie très bien les deux classes.

### **1.5.2. L'application de l'algorithme SVM :**

Notre cas consiste à classifier les requêtes de l'apprenant pour connaître son niveau cette méthode et considérer comme une assistance de l'apprenant, l'utilité du SVM est pour :

- **La catégorisation de texte**

Pour la catégorisation de texte le SVM utilise les données de formation (dataset contenant l'ensemble des requêtes avec le niveau) pour classer les requêtes dans différentes catégories telles qu'une requête d'un débutant, requête d'expert ou intermédiaire.

Pour chaque requête, calculez un score et comparez-le à une valeur seuil prédéfini. Lorsque le score d'une requête dépasse la valeur seuil, alors elle est classée dans une catégorie définie.

Classez les nouvelles instances en calculant le score de chaque requête et en le comparant au seuil appris.

### 1.5.3. Fonction du noyau SVM (SVM KERNEL)

Les algorithmes SVM utilisent un ensemble de fonctions mathématiques définies comme le noyau. La fonction du noyau est de prendre des données en entrée et de les transformer dans la forme requise.

Les fonctions utilisées sont donc RBF et Linear pour renvoyer le modèle SVC correspondant

### 1.5.4. Paramètre utilisé

Le paramètre C compromet la classification correcte des exemples d'apprentissage contre la maximisation de la marge de la fonction de décision. Pour des valeurs plus élevées de C, une marge plus petite sera acceptée si la fonction de décision est meilleure pour classer correctement tous les points d'entraînement. Un C plus faible favorisera une marge plus importante, donc une fonction de décision plus simple, au détriment de la précision de l'apprentissage. En d'autres termes, C se comporte comme un paramètre de régularisation dans le SVM.

---

#### **Pseudo-algorithme**

---

**Inputs :** Querie

**Output :** Niveau d'apprenant

$C \leftarrow [1,10]$

Kernel  $\leftarrow$  ('Linear' , 'RBF')

**Repeat**

**For all**  $\{x_i, y_i\}, \{x_j, y_j\}$  **do**

    Optimise  $\alpha_i$  and  $\alpha_j$

**end for**

**until** no changes in alpha or other ressource constraint

**ensure :** retain only the support vectors ( $\alpha_i > 0$ )

---

### 1.5.5. Préparation des données

Dans cette étape nous cherchons à préparer et optimiser nos données pour que notre modèle puisse tirer les meilleurs résultats possibles.

On va illustrer ce qu'on a dit avec un exemple de données pour montrer les données d'entrée de l'algorithme de classification SVM.

<p><i>what team did baseball 's st. louis browns become ? ... , debutant</i> <i>what is the oldest profession ? ... , debutant</i> <i>what are liver enzymes ? ... , debutant</i> <i>do i need to tear off the black outside before i put the new filter into the machine ? ... , intermediaire</i> <i>name the scar-faced bounty hunter of the old west . ... , expert</i> <i>why do heavier objects travel downhill faster ? ... , debutant</i> <i>who was the pride of the yankees ? ... , debutant</i> <i>who killed gandhi ? ... , debutant</i> <i>what is considered the costliest disaster the insurance industry has ever faced ? ... , debutant</i> <i>what sprawling u.s. state boasts the most airports ? ... , debutant</i></p>
---

Figure 9: Exemple de données d'entraînement

- **Explication**

Les débutants généralement utilisent des questions ouvertes pour chercher une réponse à leur besoin ex : what is SQL injection ?

Alors que les intermédiaires utilisent des questions fermées ex : do we use SQL in SQL injection, pour un expert est différent, ils utilisent des phrases impératives ex : SQL injection.

## 1.6. Les acteurs intervenants

Afin de pouvoir effectuer une formation en ligne, un groupe d'acteurs doit intervenir dans le système, chacun ayant des rôles précis à accomplir au sein du système. Nous pouvons les découper en deux sous-groupes.

### 1.6.1. Personnel

Il s'agit d'un groupe d'intervenant dont l'objectif principal est de veiller au bon fonctionnement du système à même de permettre aux apprenants de suivre leurs formations dans les meilleures conditions possibles. Il s'agit des formateurs, développeurs et administrateurs.



### 1.6.2. Apprenants

C'est une personne qui suit un enseignement et/ou formation quelconque. Nous distinguons deux catégories d'apprenant.

- **Les apprenants holistiques** : ils utilisent une approche globale quand ils ont une recherche à réaliser. Ils préfèrent les stratégies générales et aiment les exemples permettant une généralisation.
- **Les apprenants sérialistes** : ce type d'apprenant préfère progresser pas à pas en faisant attention aux détails et aux aspects particuliers.

Le diagramme représenté par la Figure 11 ci-dessous présente les acteurs intervenants dans le système de recherche.

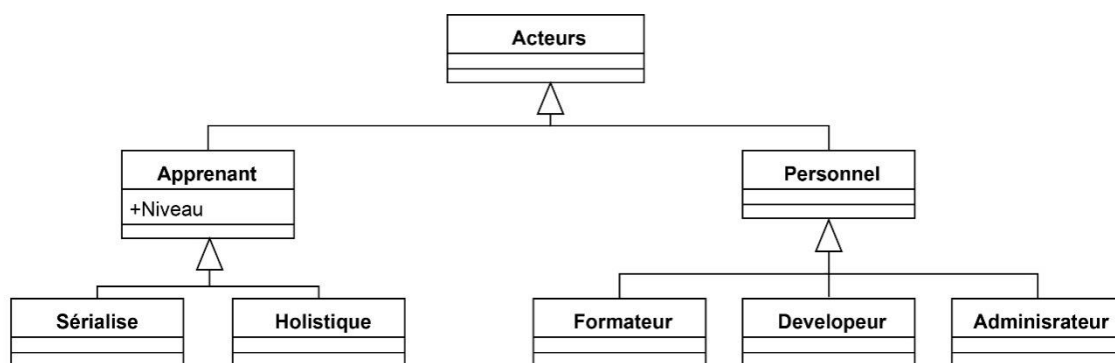


Figure 10: Diagramme de classe des acteurs intervenant.

## 1.7. L'ontologie globale

Il s'agit de l'instanciation de l'ontologie des objets pédagogiques, ainsi que celle de l'ontologie des ressources pédagogiques, il constitue l'élément fondamental dans le fonctionnement du moteur d'inférence.

### 1.7.1. L'ontologie des ressources pédagogiques

L'ontologie des ressources pédagogiques est utilisée pour attribuer une sémantique aux différents types de documents afin de pouvoir les classer. Cette ontologie est utilisée

particulièrement dans les systèmes d'apprentissage et les systèmes de recherche/recommandation relatives à l'E-Learning (tel que relevé dans la littérature).

La figure 12 ci-dessous représente le diagramme UML des ressources pédagogiques :

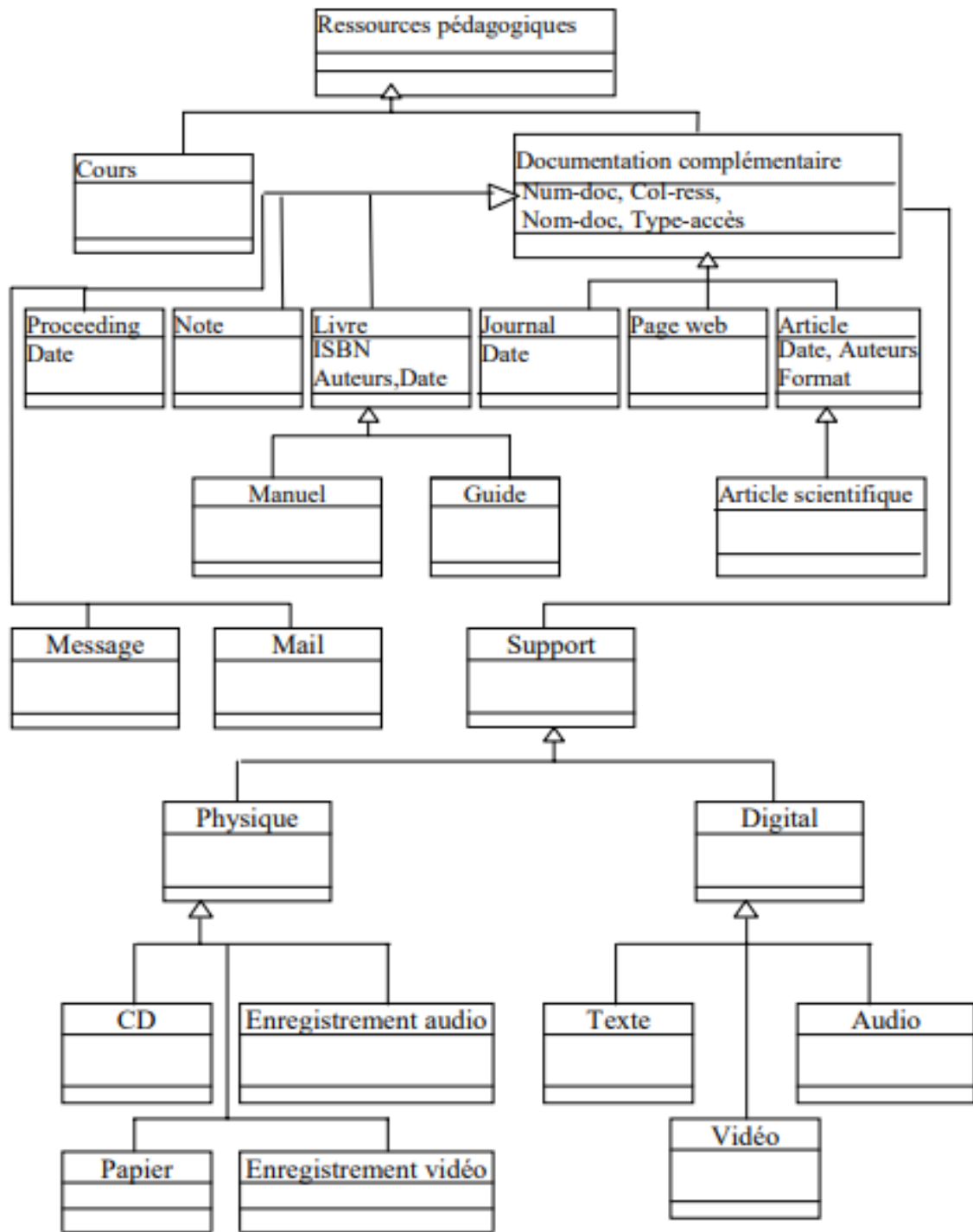


Figure 11: Diagramme de classe d'ontologie des ressources pédagogique . [45]

### 1.7.2. L'ontologie des objets pédagogiques

On s'est inspiré du travail de Jovana & al [46] dans le but de découper une ressource pédagogique en plusieurs éléments (objets pédagogique) d'une manière hiérarchique. Chaque élément est classé dans son niveau approprié, en commençant par l'élément le plus général qui soit tout en haut de la hiérarchie vers l'élément le plus spécifique qui soit tout en bas de la hiérarchie. Il existe quatre types de relation/liens entre les objets de l'ontologie.

Le modèle suivant représenté par la Figure 13 illustre la conception de cette ontologie.

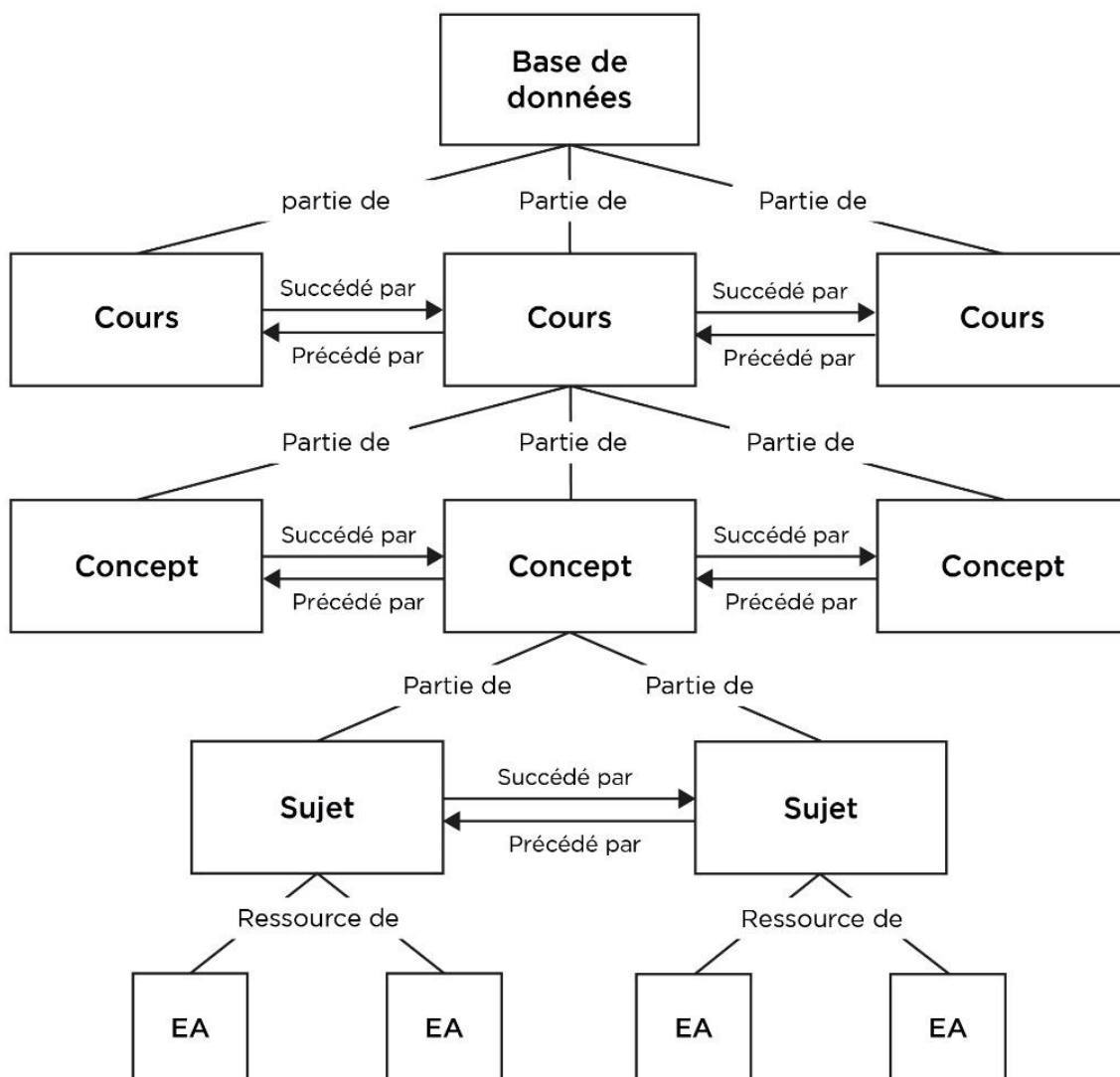


Figure 12: Diagramme d'ontologie des objets pédagogiques. [46]

### 1.7.2.1. Les propriétés entre les classes (object property)

- **Partie de** : signifie qu'un objet d'un niveau inférieur est un sous-objet d'un autre objet de niveau supérieur.
- **Ressource de** : cette propriété relie un sujet avec un élément d'apprentissage, il signifie que le contenu d'un élément d'apprentissage explique le sujet auquel il correspond.
- **Précédé par** : signifie qu'un objet pédagogique est précédé par un autre objet (par exemple : cours est précédé par un autre cours).
- **Succédé par** : ce lien relie entre les objets qui se trouvent au même niveau, il signifie qu'un objet (par exemple : concept) est succédé par un autre concept (c'est l'inverse du lien « précédé par »).

### 1.7.2.2. Les éléments de l'ontologie

Une ontologie est composée des éléments suivants :

- **Programme d'étude** : Cet objet constitue la tête/sommet de la hiérarchie de l'ontologie, il est classé dans le premier niveau de la hiérarchie. Un programme d'étude est constitué de plusieurs cours qu'un apprenant peut parcourir.
- **Cours** : un cours est un ensemble de concepts, il est classé dans le deuxième niveau de la hiérarchie de l'ontologie. Chaque cours représente une partie d'un programme d'étude.
- **Concept** : un Concept est un ensemble de sujets, il est classé dans le troisième niveau dans la hiérarchie de l'ontologie. Chaque concept représente une partie d'un cours.
- **Sujet** : un sujet est un ensemble des objets d'apprentissage, il est classé dans le quatrième niveau dans la hiérarchie de l'ontologie. Chaque sujet représente une partie d'un concept.
- **Objet d'apprentissage** : Cet objet est la base de la hiérarchie de l'ontologie, il explique le contenu de son sujet.

## 2. Définition des liens sémantiques

Un lien sémantique est une représentation relationnelle qui relie deux documents, le tableau ci-dessous représente la définition des liens sémantiques qu'on va utiliser :

Tableau 3 : Tableau des liens sémantiques

Lien sémantique	Dénotation	Signification
Similaire	$\text{Sim}(d1) = [d2]$	Le document d2 est similaire a d1
Référence	$\text{Ref}(d1) = [d3]$	Le document d3 est une référence de d1 (il est plus explicatif)
Implique	$\text{Imp}(d1) = [d2]$	La sémantique de document (fragment) prédécesseur d1 implique celle de son successeur d2.
Séquentiel	$\text{Seq}(d1) = d2$	Document prédécesseur d1 doit être parcouru par l'apprenant avant de parcourir le contenu de son successeur d2.
Cause-résultat	$\text{Cr}(d1) = [d2]$	Le document d1 est la raison d'existence de d2

## 3. Définition des règles d'inférence

Les règles d'inférence sont les règles qui fondent le processus de déduction, de dérivation ou de démonstration [32] , les règles d'inférence sont définies en exploitant les liens sémantiques :

Tableau 4 : Tableau des règles d'inférence

Règles d'inférence	Description
Règle 1	$\text{imp}(d)=d' ; \text{seq}(d')=d'' \rightarrow \text{seq}(d)=d''$
Règle 2	$\text{imp}(d)=d' ; \text{ref}(d')=d'' \rightarrow \text{ref}(d)=d''$
Règle 3	$\text{sim}(d)=d' ; \text{ref}(d')=d'' \rightarrow \text{ref}(d)=d''$
Règle 4	$\text{Seq}(d)=d' ; \text{sim}(d')=d'' \rightarrow \text{seq}(d)=d''$
Règle 5	$\text{Cr}(d)=d' ; \text{Sim}(d')=d'' \rightarrow \text{Cr}(d)=d''$

Règle 6	$\text{sim}(d)=d' ; \text{imp}(d')=d'' \rightarrow \text{imp}(d)=d''$
Règle 7	$\text{Cr}(d)=d' ; \text{ref}(d')=d'' \rightarrow \text{Cr}(d)=d''$

## 4. Méthode de recherche

L'apprenant peut rechercher les documents dans le domaine qui l'intéresse en exprimant ses centres d'intérêts à travers ses requêtes. À cet effet, nous devons définir explicitement les détails de l'exploitation des liens d'ontologie et des liens sémantiques. Nous devons également prendre en considération le type de l'apprenant ainsi que son niveau par rapport au centre d'intérêt exprimé à travers sa requête. À noter que la prise en charge du niveau de l'apprenant, lors du processus de recherche de l'information dans notre cas, reste optionnel.

### 4.1. Le cas général

Dans ce cas, ni le niveau ni la catégorie de l'apprenant n'est pas spécifiée, les résultats de recherche seront tous les documents qui ont une relation avec le document pertinent recherché quelle que soit sa classe en exploitant tous les liens sémantiques sans aucune exception. Nous exploiterons également les liens « partie de » et « ressource de ».

### 4.2. Les cas de spécification du niveau d'apprenant

- Apprenant « expert »

Dans le cas où l'apprenant est expert du domaine recherché, les résultats seront mieux filtrés. À ce titre, le moteur de recherche récupère tous les documents ayant une relation de type « succédé par » avec le document recherché.

- Apprenant « intermédiaire »

Dans le cas où l'apprenant à un niveau intermédiaire par rapport au domaine recherché, le moteur de recherche récupère tous les documents ayant une relation de type « succédé par » et « précédé par » avec le document recherché.

- **Apprenant « débutant »**

Dans le cas où l'apprenant à un niveau débutant par rapport au domaine recherché, le moteur de recherche récupère tous les documents ayant une relation de type « précédé par » avec le document recherché.

### 4.3. Les cas de spécification de la catégorie d'apprenant

- **Apprenant « sérialiste » :**

Pour ce type d'apprenant, la notion de la progression pas à pas, est prise en considération, il est accordé une importance capitale au détail. Ici, le moteur de recherche récupère tous les documents liés aux documents recherchés par des liens de types : « partie de », « séquentiel », « cause-résultat », « ressource de » et « implique ».

- **Apprenant « holistique »**

Pour ce type d'apprenant, nous sommes dans le cas inverse, car la notion de généralisation est prise en considération. Ici, le moteur de recherche récupère tous les documents liés aux documents recherchés par l'inverse des liens « partie de », « cause-résultat », « implique », « ressource de » et « référence ».

## 5. Conclusion

Dans ce chapitre nous avons expliqué tout ce qui concerne la conception du système, les ontologies utilisées, les liens sémantiques et les règles d'inférence.

Nous avons présenté l'architecture de système qui va nous permettre de réaliser le système pour répondre à nos exigences.

---

## ***CHAPITRE 5***

### ***Implémentation et évaluation***

---



Après avoir achevé la conception de notre système. Il est temps de concrétiser notre contribution en implémentant notre système pour qu'il soit prêt à être utilisé par les acteurs concernés.

Le but de notre système est de faire une sélection des ressources pertinentes pour l'utilisateur qui lance l'interrogation.

Dans ce chapitre nous allons présenter les outils utilisés dans le développement, les fonctionnalités incluent dans le prétraitement des données et la modélisation de nos modèles ainsi que l'implémentation de ces derniers.

Nous allons aussi présenter le système final et leurs interfaces afin d'introduire la méthode d'utilisation du logiciel. Pour finir nous allons voir et discuter les résultats obtenus et mesurer leurs fiabilités grâce aux tests réalisés.

### **1. Matériels et outils**

Pour développer notre système nous avons utilisé le matériel et les logiciels suivants :

#### **1.1. Matériels utilisés**

Nous avons travaillé sur un ordinateur personnel. Les informations suivantes représentent la configuration du matériel utilisé.

##### **Ordinateur personnel**

Processeur : Intel(R) Core(TM) i5-7300U CPU @ 2.60GHz 2.70 GHz

RAM : 8,00 Go

Type du système : Système d'exploitation 64 bits, processeur x64

Édition Windows 10 Professionnel

Version 20H2

Build du système d'exploitation 19042.1237

#### **1.2. Outils et langage de développement**

La mise en œuvre des différentes étapes discutées se fait avec plusieurs outils et logiciels de développement dans cette section on va parler sur tous les langages utilisés ainsi les logiciels pour implémenter et réaliser notre projet, l'utilisation des technologies reconnues et des versions stables sont nécessaires.

### 1.3. Langage Python

Python est un langage de programmation interprété, multi-paradigme et multi-plateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse miettes et d'un système de gestion d'exceptions. Python est un langage qui peut s'utiliser dans de nombreux contextes et s'adapter à tout type d'utilisation grâce à des bibliothèques spécialisées. Pour le développement de notre système nous avons utilisé la version 3.8.2 du python.

### 1.4. Flask (Framework)

Flask est un micro Framework open-source de développement web en Python. Il est classé comme micro-Framework car il est très léger. Flask a pour objectif de garder un noyau simple mais extensible. Il n'intègre pas de système d'authentification, pas de couche d'abstraction de base de données, ni d'outil de validation de formulaires. Cependant, de nombreuses extensions permettent d'ajouter facilement des fonctionnalités. Il est distribué sous licence BSD.

### 1.5. NLTK (Natural language Toolkit)

Natural Language Toolkit (NLTK) est une bibliothèque logicielle en Python permettant un traitement automatique des langues, développée par Steven Bird et Edward Loper du département d'informatique de l'université de Pennsylvanie. En plus de la bibliothèque, NLTK fournit des démonstrations graphiques, des données-échantillon, des tutoriels, ainsi que la documentation de l'interface de programmation (API).

### 1.6. Scikit-Learn

Scikit-learn est une bibliothèque d'apprentissage automatique gratuite pour Python. Elle comporte divers algorithmes tels que la machine vectorielle de support, les forêts aléatoires et les k-voisins, et elle prend également en charge les bibliothèques numériques et scientifiques Python comme NumPy. Son utilisation ici

peut être résumée par le besoin de test grâce à la matrice de confusion. Nous avons utilisé la version '0.24.2'.

### **1.7. OWLready 2**

Owlready2 est un package pour la programmation orientée ontologie en Python. Il peut charger des ontologies OWL 2.0 en tant qu'objets Python, les modifier, les enregistrer et effectuer un raisonnement via HermiT (inclus). Owlready2 permet un accès transparent aux ontologies OWL (contrairement aux API Java habituelles).

Owlready version 2 inclut un triplestore/quadstore optimisé, basé sur SQLite3. Ce quadstore est optimisé à la fois pour les performances et la consommation de mémoire. Contrairement à la version 1, Owlready2 peut gérer de grosses ontologies. Owlready2 peut également accéder à l'UMLS et à la terminologie médicale (en utilisant le sous-module PyMedTermino2 intégré).

### **1.8. RdfLib**

RDFLib est une bibliothèque Python pour travailler avec RDF, un langage simple mais puissant pour représenter des informations. Cette bibliothèque contient des analyseurs/sérialiseurs pour presque toutes les sérialisations RDF connues, telles que RDF/XML, Turtle, N-Triples et JSON-LD, dont beaucoup sont maintenant pris en charge dans leur forme mise à jour (par exemple Turtle 1.1). La bibliothèque contient également des back-ends Graph en mémoire et persistants pour stocker les informations RDF et de nombreuses fonctions pratiques pour déclarer les espaces de noms de graphes, héberger des requêtes SPARQL, etc. Il est en développement continu avec la version stable la plus récente, rdfliB 5.0.0 étant sorti le 18 avril 2020. Il a été créé à l'origine par Daniel Krech avec la première version en novembre 2002.

### **1.9. RDF**

Le Resource Description Framework (RDF) est une famille de spécifications du World Wide Web Consortium (W3C) conçue à l'origine comme un modèle de

données pour les métadonnées. Il est devenu une méthode générale pour la description conceptuelle ou la modélisation des informations mises en œuvre dans les ressources Web, en utilisant une variété de notations syntaxiques et de formats de sérialisation de données. Il est également utilisé dans les applications de gestion des connaissances. RDF a été adopté en tant que recommandation du W3C en 1999. La spécification RDF 1.0 a été publiée en 2004, la spécification RDF 1.1 en 2014.

### **1.10. SPARQL**

SPARQL est un langage de requête RDF, c'est-à-dire un langage de requête sémantique pour les bases de données, capable de récupérer et de manipuler les données stockées au format Resource Description Framework (RDF). Il est devenu une norme par le RDF Data Access Working Group (DAWG) du World Wide Web Consortium et est reconnu comme l'une des technologies clés du Web sémantique. Le 15 janvier 2008, SPARQL 1.0 a été reconnu par le W3C comme une recommandation officielle, et SPARQL 1.1 en mars 2013.

SPARQL permet à une requête de se composer de modèles triples, de conjonctions, de disjonctions et de modèles facultatifs.

### **1.11. HTML**

Le langage de balisage hypertexte, ou HTML, est le langage de balisage standard pour les documents conçus pour être affichés dans un navigateur Web. Il peut être assisté par des technologies telles que les feuilles de style en cascade (CSS) et des langages de script tels que JavaScript.

Les navigateurs Web reçoivent des documents HTML d'un serveur Web ou d'un stockage local et restituent les documents dans des pages Web multimédias. HTML décrit la structure d'une page Web de manière sémantique et inclut à l'origine des indices pour l'apparence du document.

### **1.12. CSS**

Les feuilles de style en cascade (CSS) sont un langage de feuille de style utilisé pour décrire la présentation d'un document écrit dans un langage de balisage tel que HTML. CSS est une technologie de base du World Wide Web, aux côtés de HTML et JavaScript.

CSS est conçu pour permettre la séparation de la présentation et du contenu, y compris la mise en page, les couleurs et les polices. Cette séparation peut améliorer l'accessibilité du contenu, offrir plus de flexibilité et de contrôle dans la spécification des caractéristiques de présentation, permettre à plusieurs pages Web de partager le formatage en spécifiant le CSS pertinent dans un fichier .css séparé, ce qui réduit la complexité et la répétition du contenu structurel et permet le fichier .css à mettre en cache pour améliorer la vitesse de chargement des pages entre les pages qui partagent le fichier et sa mise en forme.

### **1.13. JavaScript**

JavaScript, souvent abrégé en JS, est un langage de programmation conforme à la spécification ECMAScript. JavaScript est de haut niveau, souvent compilé juste à temps et multi-paradigme. Il a une syntaxe entre accolades, un typage dynamique, une orientation objet basée sur des prototypes et des fonctions de première classe.

Avec HTML et CSS, JavaScript est l'une des technologies de base du World Wide Web. Plus de 97 % des sites Web l'utilisent côté client pour le comportement des pages Web, incorporant souvent des bibliothèques tierces. La plupart des navigateurs Web disposent d'un moteur JavaScript dédié pour exécuter le code sur l'appareil de l'utilisateur.

En tant que langage multi-paradigmes, JavaScript prend en charge les styles de programmation événementiels, fonctionnels et impératifs. Il dispose d'interfaces de programmation d'applications (API) pour travailler avec du texte, des dates, des

expressions régulières, des structures de données standard et le modèle objet de document (DOM).

### **1.14. Bootstrap**

Bootstrap est un Framework CSS gratuit et open source destiné au développement Web frontal réactif et mobile. Il contient des modèles de conception CSS et (éventuellement) JavaScript pour la typographie, les formulaires, les boutons, la navigation et d'autres composants d'interface.

### **1.15. UML**

Le langage de modélisation unifié (UML) est un langage de modélisation de développement à usage général dans le domaine du génie logiciel qui vise à fournir un moyen standard de visualiser la conception d'un système.

La création d'UML a été motivée à l'origine par le désir de standardiser les systèmes de notation disparates et les approches de conception de logiciels. Il a été développé chez Rational Software en 1994-1995, avec un développement ultérieur mené par eux jusqu'en 1996.

### **1.16. Microsoft visual studio**

Microsoft Visual Studio est un environnement de développement intégré (IDE) de Microsoft. Il est utilisé pour développer des programmes informatiques, ainsi que des sites Web, des applications Web, des services Web et des applications mobiles. Visual Studio utilise des plates-formes de développement de logiciels Microsoft telles que Windows API, Windows Forms, Windows Presentation Foundation, Windows Store et Microsoft Silverlight. Il peut produire à la fois du code natif et du code managé.

Visual Studio inclut un éditeur de code prenant en charge IntelliSense (le composant de complétion de code) ainsi que la ré-factorisation de code. Le débogueur intégré fonctionne à la fois comme débogueur au niveau de la source et comme débogueur au niveau de la machine. D'autres outils intégrés incluent un profileur de code, un

concepteur pour la création d'applications GUI, un concepteur Web, un concepteur de classe et un concepteur de schéma de base de données. Il accepte les plug-ins qui étendent les fonctionnalités à presque tous les niveaux, y compris l'ajout de la prise en charge des systèmes de contrôle de source (comme Subversion et Git) et l'ajout de nouveaux ensembles d'outils tels que des éditeurs et des concepteurs visuels pour des langages spécifiques à un domaine ou des ensembles d'outils pour d'autres aspects du développement logiciel. Cycle de vie (comme le client Azure DevOps : Team Explorer).

### **1.17. Protégé**

Protégé est un système auteur pour la création d'ontologies. Il a été créé à l'université Stanford et est très populaire dans le domaine du Web sémantique et au niveau de la recherche en informatique.

Protégé est développé en Java. Il est gratuit et son code source est publié sous une licence libre (la Mozilla Public License).

Protégé peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL, etc.

Il possède plusieurs concurrents tels que Hozo, OntoEdit et Swoop. Il est reconnu pour sa capacité à travailler sur des ontologies de grandes dimensions.

### **1.18. Bootstrap studio**

Bootstrap Studio est une application de conception et de développement Web propriétaire. Il offre un grand nombre de composants pour créer des pages réactives, notamment des en-têtes, des pieds de page, des galeries et des diaporamas, ainsi que des éléments de base tels que les étendues et les divs.

Le programme peut être utilisé pour créer des sites Web et des prototypes. Il est construit sur le Framework Electron populaire et est multiplateforme.

## **2. Implémentation du système**

### **2.1. Préparations des données**

Cette première étape consiste à créer une base de ressources qui contient plusieurs ressources pédagogiques et de créer une ontologie cette dernière représente la structure de ces ressources.

### **2.2. Traitement des données (indexation)**

Le résultat concret de l'étape d'indexation est premièrement la production d'index qui vont associer à chaque Ressource, des termes censés représenter les sujets dont ils traitent. En second lieu, l'indexation fixe des stratégies de recherche, qui vont utiliser les termes indexés pour repérer le document "pertinent" parmi la masse documentaire.

Indexer un document, c'est donc élaborer un jeu d'éléments censé le représenter au mieux, puis définir les moyens d'accès à ce document à l'aide de ces éléments représentatifs.

L'indexation a pour principal objectif de rendre accessibles des informations, que l'on repère pour cela au moyen d'index. ... L'indexation est un processus consistant à reformuler le contenu d'un document sous une forme plus adaptée à son exploitation dans une application donnée (par exemple la recherche).

## **3. Déploiement du Module de classification**

Après l'étape de préparation l'examen de la requête de l'utilisateur sera effectuée avant le lancement de la recherche.

Ce Module contient deux fonctions principales sont les suivants :

Fonction d'Entraînement (Training) : cette fonction contient deux parties, la première consiste à lire le fichier qui contient l'ensemble des données de test Et la deuxième partie concerne l'entraînement de la machine. Elle retourne en sortie le modèle.

Fonction du Test : cette fonction prend en entrant la requête d'utilisateur pour faire le traitement et la classification pour que la sortie soit le niveau de ce dernier et le modèle déduit de la fonction d'entraînement.



L'algorithme ci-dessous représente une partie abstraite du code source :

```
Algorithme query_classifier
CONST
TYPE
VAR
DEBUT
FONCTION Train()
  VAR List train_label,Train_data
  debut
  train_data = Loadfile(path p)
  train_label = normalise(train_data)
  parameters = {use parameter}
  model = searchestimator(classifier,parameter)
  model = training(train_vector,train_label)
  final_model = create_model(model)
  return final_model
  fin
FONCTION testing(final_model,query)
  debut
  test_vector = normalise(query)
  return predict(created_model,test_vector)
  fin
fin
```

Figure 13:  
abstraite du

Partie  
code source.

#### 4. La

recherche

#### d'information

L'étape de recherche est la plus essentielle dans ce système, dans cette phase le système exploite l'index de données et l'ontologie en utilisant le moteur d'inférence pour des résultats plus pertinents.

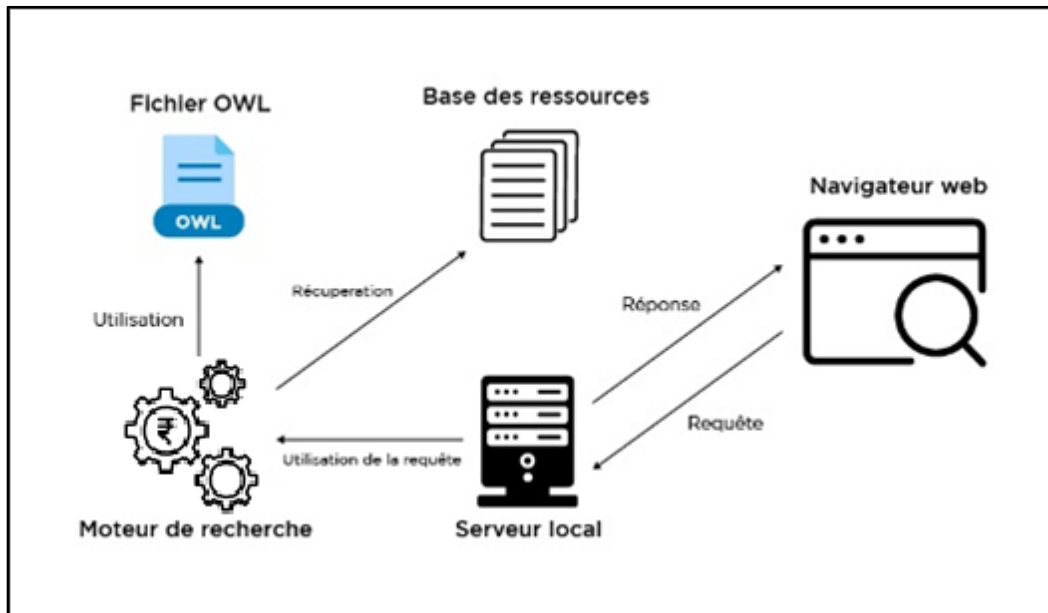
L'apprenant lance une requête d'interrogation pour faire une recherche, cette dernière sera indexée aussi, basant sur les liens sémantiques et les règles d'inférence le système effectue une sélection des documents pertinents (Recherche sémantique).

#### 5. Déploiement du système

Nous avons implémenté le système sous forme d'une application web. Pour cela nous avons utilisé le langage PYTHON et le Framework Flask qui donne une grande flexibilité et la portabilité de l'application. Celle-ci rentre dans le cadre de la nouvelle génération du web (le web sémantique). En effet nous avons utilisé le langage OWL pour représenter l'ontologie développée et les deux bibliothèques (OWLREADY2, RDFLIB) pour son

exploitation et sa manipulation. La figure ci-dessous présente l'architecture externe du système développé.

*Figure 14: Architecture externe du système.*



Dans cette section nous allons présenter l'interface graphique du système réalisé afin de guider les apprenant afin d'utiliser ce système, nous avons deux interfaces à expliquer : la page de recherche, la page des résultats  
Commençons de la première page essentielle :

## 6. Page de recherche



*Figure 15: Capture de l'interface de recherche du système.*

Cette interface contient un seul élément essentiel et deux éléments optionnels

1 : le premier élément contient trois Objets (champ de recherche, le type de ressources recherché, le bouton de recherche).

2 : le deuxième élément contient la première partie du filtrage, est le niveau de l'apprenant dans le domaine recherché.

3 : le troisième élément contient la deuxième partie du filtrage, l'apprenant doit choisir sont type.

### 7. La page des résultats



*Figure 16: Capture de l'interface des résultats du système.*

Concernant la page des résultats, elle contient deux éléments :

1 : le premier élément contient un champ de recherche pour que l'apprenant puisse lancer une autre requête d'interrogation.

2 : le deuxième élément contient les résultats de recherche.

### 8. Evaluation

Depuis la naissance du domaine de la RI, l'évaluation des modèles de RI et méthodes proposées a toujours été un centre d'intérêt important. En effet, elle permet de caractériser le modèle et de fournir des éléments de comparaison entre modèles.

Pour cela des mesures de qualité des systèmes de recherche ainsi que des ensembles de données ont été développés afin de tester ces systèmes sur une base commune.

La qualité d'un système doit être mesurée en comparant les réponses du système avec les réponses idéales que l'utilisateur espère recevoir. Plus les réponses du système correspondent à celles que l'utilisateur espère, meilleur est le système.

Pour réaliser une telle évaluation, une expérimentation qui utilise les éléments suivants doivent être établies :

- Une collection de documents de test.
- Un ensemble de requêtes de test.
- Une liste de documents pertinents pour chaque requête.
- Des mesures et des critères quantifiables.

Et même si le temps de réponse et l'espace utilisé pour le stockage d'informations sont plus ou moins importants dans l'évaluation des SRI, la qualité des résultats renvoyés par un système de recherche reste le critère le plus important dans l'évaluation et la comparaison des systèmes entre eux.

Le manque des éléments de test forme un problème d'évaluation, La mesure de performance et le calcul de précision du système seront plus difficiles.

Pour cela il existe deux façons d'évaluation, la première consiste à faire un jeu de test manuellement et la deuxième à l'aide d'un programme de test automatique.

Pour cela on a développé un programme de test automatique sert à générer des requêtes aléatoirement et (le type ou le niveau) d'apprenant, il les combine pour lancer le test.

### **8.1. Tests et Résultats**

Après l'exécution du programme de test automatique qu'on a créé nous pouvons classer les résultats obtenus de la recherche comme suite :

La comparaison des réponses d'un système de recherche pour une requête avec les réponses idéales nous permet d'évaluer les métriques d'évaluation de SRI les plus courantes :

Les taux de rappel et de précision sont les mesures les plus utilisées pour l'évaluation d'une recherche parce qu'ils permettent d'évaluer la capacité d'un SRI de répondre aux deux objectifs principaux qui sont :

- Retrouver tous les documents pertinents.

- Rejeter tous les documents non pertinents.

- **Taux de rappel**

Le rappel mesure la capacité du système à retrouver tous les documents pertinents répondant à une requête. Autrement dit, il mesure la proportion de documents pertinents restitués par le système relativement à l'ensemble des documents pertinents contenus dans la collection.

Si le rappel vaut 1 c'est que les documents pertinents disponibles ont été restitués par le système, inversement si le rappel vaut 0 c'est qu'aucun document pertinent n'a pas été restitué. Cette mesure permet aussi de déterminer le silence, c'est-à-dire la proportion de documents pertinents non trouvés. Voir l'équation (5)

C'est une estimation de la capacité du SRI de retourner tous les documents pertinents de la collection et à éliminer le silence.

D'un point de vue opposé, nous pouvons définir la notion du « silence » qui représente le pourcentage de documents pertinents n'ayant pas été extrait (faux négatifs).

$$\text{Silence} = 1 - \text{rappel} \quad \textit{Equation 6}$$

- **Taux de précision**

La précision mesure la capacité du système à rejeter tous les documents non pertinents à une requête.

Autrement dit, elle mesure la proportion de documents pertinents restitués relativement à l'ensemble des documents restitués par le système.

Elle mesure la capacité du système à trouver exclusivement des documents pertinents et donc à éliminer le bruit. Voir l'équation (4)

La précision vaut 1 quand tous les documents restitués sont pertinents. Elle vaut 0 si aucun des documents restitués n'est pas pertinent. Cette mesure détermine également le bruit, c'est-à-dire la proportion de documents non pertinents restitués par le système.

La précision est la capacité d'un système à ne sélectionner que des documents pertinents.

D'un point de vue opposé, nous pouvons définir la notion du « bruit » qui représente le pourcentage de documents non pertinents extraits par le système (faux positifs) :

$$\text{bruit} = 1 - \text{précision} \quad \textit{Equation 7}$$

### **8.1.1. Jeu de test**

Le jeu de test représenté dans cette section pour voir le fonctionnement du système.

**Requête :** Html table.

**Type et/ou niveau d'apprenant :** Intermédiaire.

**Résultat :** Padding & spacing.

**Résultat estimé :** Html Tables / Border.

**Commentaire :** Résultat non pertinent.

**Requête :** SQL injection.

**Type et/ou niveau d'apprenant :** holistique / sérialiste.

**Résultat :** Web Application Attacks /Mapping & Port Scanning/ WIFI Attacks/ Firewalls/Web application.

**Résultat estimé :** cyber attack/Networking/Cyber crime/wifi attacks.

**Commentaire :** Résultat pertinent.

**Requête :** what is electronic crime.

**Type et/ou niveau d'apprenant :** Expert.

**Résultat :** cyber crime/Money Making Threats /Networking/ dark web.

**Résultat estimé :** cyber crime/Money Making Threats/Networking/dark web.

**Commentaire :** Résultat pertinent.

**Requête :** Virus attack.

**Type et/ou niveau d'apprenant :** Sérialiste / debutant

**Résultat :** Code Execution/ Network Attacks/ Mapping & Port Scanning/ Vulnerability Scanners/ Buffer Overflow

**Résultat estimé :** Code Execution/ Vulnerability Scanners/ Buffer overflow

**Commentaire :** Résultat pertinent

**Requête :** Spoofing.

**Type et/ou niveau d'apprenant :** Sérialiste.

**Résultat :** -Networking transport/ Cyber Crime/ Networking/ WIFI Attacks/ Cyber Defence/ Networking layer/ Networking Basics/ Firewalls/ Web application.

**Résultat estimé :** Cyber attack/ Networking/ Cyber Crime/ WIFI Attacks.

**Commentaire :** Résultat pertinent.

D'après le jeu de test qu'on a effectué les résultats sont satisfaisants.

## **9. Conclusion**

Dans ce chapitre on a présenté les outils utilisés, qui ont permis d'implémenter le système ainsi le guide d'utilisation, on a éliminé le bruit et diminuer le silence des systèmes de recherche classique.

On n'a pas pu évaluer le système et calculer sa précision à cause de manque des benchmark (la collection de test), Malgré les résultats obtenus ont été satisfaisants, Ces résultats ont été obtenus après plusieurs tentatives non concluantes ainsi que plusieurs modifications dans l'implémentation.

## Conclusion Générale

Durant la dernière période vécue par la pandémie du corona virus, l'utilisation d'internet est devenue remarquable dans tous les domaines, L'enseignement en ligne est devenu une alternative partout dans le monde, ce qui a conduit à une augmentation accélérée du contenu académique sur Internet. À cet effet, la recherche du contenu pour chaque apprenant est devenue plus complexe, et vue à l'importance de temps car il sera obligé de consulter un nombre important des ressources pour satisfaire leurs attentes ainsi que leurs besoins en information.

Ce qui nous a conduit à étudier, la possibilité de la réduction de ce parcours afin d'améliorer le processus de recherche et récupérer le besoin en informations pertinentes qui a pour objectif à réduire le bruit et le temps.

Dans la première partie de notre mémoire nous avons réalisé une recherche sur le E-Learning, en première phase a permis de distinguer les aspects, les principes, les avantages, les inconvénients, les types d'apprentissage, les composant d'un dispositif d'E-Learning. Quant à la deuxième phase, elle a été réalisée sur les SRI, nous avons abordé dans ce cadre les modèles de RI, les processus de recherche, les objectifs d'un SRI, l'évaluation d'un SRI, les systèmes de recherche sémantique, les ontologies et l'apprentissage automatique de la machine.

Suite à ces recherches, nous avons classifié quelques travaux connexes relatifs à l'E-Learning après les avoir consultés.

La deuxième partie a été consacrée par la proposition d'un système de recherche d'information à base d'ontologie. L'ontologie globale mise en œuvre est formée de deux sous ontologie l'un pour le type de ressource et l'autre pour la structure des ressources, notre système a été renforcé par des liens sémantiques et d'un moteur d'inférence pour améliorer la précision. Sans oublier l'impact de l'apprenant sur le résultat de la recherche, tout en prenant en considération son niveau d'étude ainsi que son type, un module de classification



## *Conclusion Générale*

---

basée sur l'apprentissage automatique de la machine a été déployé afin de prédire le niveau de l'apprenant depuis la requête.

Le système réalisé est une application web simple à utiliser, qui contient deux interfaces pour lancer la recherche et sélectionner le niveau et/ou type ainsi qu'une autre pour afficher les résultats.

Les obstacles rencontrés durant la réalisation de notre travail sont les suivants :

- Le manque des données (data set) pour l'entraînement de la machine pour une classification bien précise des requêtes.
- Le manque des données de test (benchmark) pour l'évaluation du système.
- et surtout le manque de temps.

Arrivé à ce stade, nous pouvons dire que les objectifs fixés depuis le début de ce travail sont atteints, le système a permis de fournir une solution d'une part et que l'ontologie a permis l'organisation des ressources d'autre part.

De plus, il est possible de déployer le système pour qu'il soit opérationnel en temps réel.

Pour plus d'amélioration de notre système en propose les perspectives d'améliorations suivantes :

- Suivi d'un apprenant en basant sur l'historique de recherche et les centres d'intérêt.
- Les suggestions des sujets effectuer par le module de classification des sujets (sujet classifier).
- L'ajout de module de classification d'image pour effectuer la recherche par image.

## Références

- [1] B. M, Le E-Learning : la distance en question dans la fonction, Harmattan, 2005.
- [2] H. Fournier, «A Review Of The State Of The Field Of Adult Learning ELearning,,» National Research Council Canada, Institute for Information Technology Research Assistants: Cornelia Dragne and Daniel Romila, 2006.
- [3] H. J.C, «Different Types of Distance Education,» *ezonearticles*, 2011.
- [4] M. R. O. Al U., «Web-Based Distance Education Systems : Required Features and Standards,» *Bilgi Dünyası*, 2004.
- [5] J. L. D.-D. C. & G. K. Moore, «E-Learning, online learning, and distance learning environments : Are they the same?,,» *The Internet and Higher Education*, pp. 1-4, 2011.
- [6] A. V. D. & C. . Sangra, «Building an Inclusive Definition of E-Learning: An Approach to the Conceptual Framework,» 2012.
- [7] L. G. & D. W. Donald P. Sheridan, «CECIL: THE FIRST WEB-BASED LMS,» The University of Auckland, NEW ZEALAND, Auckland, 2002.
- [8] «lesechos,» 2012. [En ligne]. Available: [https://www.lesechos.fr/30/08/2012/lesechos.fr/0202236891247\\_harvard--berkeley-et-le-mit-lancet-une-plate-forme-mondiale-d-E-Learning-gratuite.htm](https://www.lesechos.fr/30/08/2012/lesechos.fr/0202236891247_harvard--berkeley-et-le-mit-lancet-une-plate-forme-mondiale-d-E-Learning-gratuite.htm).
- [9] M. Prat, « E-Learning, réussir un projet », 2008.
- [10] E. I. M faisa, «E-Learning,» 2015.
- [11] A. S, «E-Learning adaptatif,» Université Sophia-Antipolis, 2005.
- [12] cours d'initiation a la culture antique, charleroi, 2019.
- [13] «modélisation d'un système de recherche pour les systèmes hypertexte,» l'école nationale superieur des mines et de l'université jean monnet, saint-etienne, 2002.
- [14] information retrieval systeme, Norwel : Kluwer academic, 1997.
- [15] système de recherche d'information etendu basé sur une projection multi-espace, Qubec: université de quebec a chicoutimi, 2018.
- [16] Information retrieval : algorithms and heuristics, Norwel: kluwer academic, 2004.
- [17] «Web information retrieval and search,» *Al-satil journal*, pp. 55 - 92 .
- [18] Introduction to modern information retrieval, M.J.mcgill , 1983.
- [19] «Relevance weighting of search,» *Journal of the american society for information science*, 1976.

- [20] «Relating the new language,» 2000.
- [21] O.R.Bouidghaen, «Accès contextuel à l'information dans un environnement mobile : approche basée sur l'utilisation d'un profil situationnel de l'utilisateur et d'un profil de localisation des requêtes,» Université Paul Sabatier, 2011.
- [22] M. M. BAZIZ, «Indexation conceptuel guidée par ontologie pour la recherche d'information,» université paul sabatier, 2005.
- [23] M.Daoud, «Accès personnalisé à l'information : approche basée sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie de domaines à travers l'historique des sessions de recherche,» Université Paul Sabatier, 2009.
- [24] N. Nassr, «Croisement de langues en recherche d'information : traduction et désambiguïsation de requêtes,» Université Paul Sabatier, 2002.
- [25] S. C. e. F. Benaissa, «Prise en compte du profil utilisateur dans les systèmes d'information semi-structurée,» Ecole nationale supérieure d'informatique , Alger, 2008.
- [26] «Enabling technology for knowledge sharing,» *ALMag*, vol. ALMag, p. 36 , 1991.
- [27] T. Gruber, «Knoweldge acquisition,» *International journal of humain-computer studies*, vol. 5, n° %12, pp. 199 - 220, 1993.
- [28] B. WN, «Construction d'ontologie d'ingenieurie pour le partage et la réutilisation les connaissance,» Université twente, 1999.
- [29] B. R. F. D. Studer R, «Knowledge engeneering : Principles and methods,» *Data & knowledge engeneering* , vol. 25, n° %11-2, pp. 161 - 197, 1998.
- [30] F. B. ROGERS, «medical search hedaings,» U.S national library of medecin, [En ligne]. Available: <https://www.nlm.nih.gov/mesh/>.
- [31] D. -. K. R. F. B. B. M. CAO T-D, «Vers un système d'aide a la veille technologique guidé par une ontologie,» chez *Ingenieurie des connaissance* , 2005.
- [32] w. Foundation, «wikipedia,» wikimedia foundation, 15 janvier 2001. [En ligne]. Available: [https://fr.wikipedia.org/wiki/Syst%C3%A8me\\_de\\_recommandation?fbclid=IwAR1JwuelArdDExrazpi7-JqqK8-Z4CsSXk-7klw492ch\\_âiTUQZvZ9FWIVc](https://fr.wikipedia.org/wiki/Syst%C3%A8me_de_recommandation?fbclid=IwAR1JwuelArdDExrazpi7-JqqK8-Z4CsSXk-7klw492ch_âiTUQZvZ9FWIVc).
- [33] E. Fouz, «Clustering des News,» université de nice sophia antipolis, Nice, 2013.
- [34] O. Z. M. LIOR ROKACH, *Data Mining with Decision Trees : theory and applications*, Hackensack: World Scientific Pub Co Inc, 2008.
- [35] M. Bouguessa, «Forage de données,» Université du Québec à Montréal, Montréal, 2015.
- [36] Y. R. ,. F. A. Jacques savoy, «Recherche d'information dans des sources distribué,» chez *Acts du XIX ème congres inforsid* , 2001.
- [37] c. w. m. ,. h. w. Qi zhou, «Adapting key word query to semantic search,» 2017.

- [38] k. z. q. l. t. haofen wang, «Query to semantic,» chez *aliteweight keyword interface to semantic search*, shanghai, 2008.
- [39] v. s. ., e. m. yuangui lei, «semantic search,» chez *a search engine for the semantic web*, 2006.
- [40] t. f. a. j. r. urvy shah, «information retrieval on the semantic web,» chez *Proceeding of the eleventh international conference on information and knowledge management*, NEW YORK , 2002.
- [41] e. d. v. d. c. irene celino, «squiggle,» chez *An experience in model driving developement of real world semantic search engine* , 2007.
- [42] M. g. ., R. a. Al-yahia, «Ontology in E-Learning : review of litterature,» *International journal of software engeneering & application* , vol. 9, pp. 67 - 84, 2015.
- [43] A. M. Gina george, «Review of ontology based recommender system in E-Learning,» 2019.
- [44] s. m. o. ., r. d. Olaperi yeside sowunmi, «A semantic web based framework for RI in E-Learning system,» covenan university , Ogun State, 2018.
- [45] R. a.-o. Arezki hammache, «Un systeme de recherche d'information pour E-Learning,» *Cain info*, vol. 11, pp. 85 - 105 , 2008.
- [46] S. c. M. R. Jovana J, «System for learning object retrival in ontology based database course,» chez *the seventh international conference on E-Learning* , Belgrad, 2016.