



N° d'ordre :

Centre des Etudes Doctorales : Sciences et Techniques et Sciences Médicales
Formation Doctorale : Sciences de l'Ingénieur, Sciences Physiques, Mathématiques et Informatique
Discipline : Informatique
Spécialité : Informatique
Laboratoire : Laboratoire d'Intelligence Artificielle, de Sciences de données et Systèmes Emergents (LIASSE, ENSA de Fès)

Thèse pour l'obtention de DOCTORAT

Personnalisation de l'apprentissage en ligne par système de recommandation basé sur l'ontologie, l'analyse de données massives et le web mining

Préparée par : Mme Meryem AMANE

Soutenue le samedi 17 juin 2023 à 10h à l'Ecole Nationale des Sciences Appliquées de Fès,
devant le jury composé de :

Nom et Prénom	Grade	Etablissement	Qualité
Mohamed LAMRINI	PES	Faculté des Sciences Dhar El Mehraz de Fès	Président
Mohamed SABBANE	PES	Faculté des Sciences de Meknès	Rapporteur
Aissam BERRAHOU	PH	Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes de Rabat	Rapporteur
Khalid FARDOUSSE	PH	Faculté de Charia de Fès	Rapporteur
Nour El Houda CHAOUI	PES	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Hakim EL FADILI	PES	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Karima AISSAOUI	PH	Faculté des Sciences d'Oujda	Examineur
Mohammed Berrada	PES	Ecole Nationale des Sciences Appliquées de Fès	Directeur de thèse

Dédicaces

À mes très chers parents, vous êtes les premiers à qui je dédie cette thèse. Vous m’avez appris l’importance de la persévérance, de la discipline et du travail acharné. Votre amour, votre soutien inconditionnel et vos précieux conseils ont été mes piliers tout au long de ce parcours académique. Vous êtes les meilleurs parents du monde.

À mon mari, tu es mon soutien indéfectible et mon partenaire de vie. Tu as été là pour moi dans les moments difficiles, tu m’as encouragé à continuer et tu m’as aidé à garder le cap. Cette thèse est également dédiée à notre fils Fares, notre petit rayon de soleil et notre plus grand accomplissement.

À mes chers frères, vous êtes ma famille et mes plus grands supporters. Vous m’avez toujours poussé à donner le meilleur de moi-même, vous m’avez soutenu dans toutes mes décisions et vous avez partagé avec moi les moments de joie et de peine. Vous êtes des personnes formidables et je suis fière de vous avoir comme frères.

Enfin, à tous mes professeurs, je vous remercie de tout cœur pour votre enseignement, votre mentorat et votre passion pour la recherche. Vous m’avez appris à être rigoureuse, curieuse et persévérante, et pour cela, je vous en suis éternellement reconnaissante.

...

Remerciements

Ce travail de recherche a été effectué au sein du laboratoire LIASSE (l'Intelligence Artificielle, Sciences des données et Systèmes Émergents), à l'Ecole Nationale des Sciences Appliquées de Fès, de l'Université Sidi Mohammed Ben Abdellah.

Je suis emplie d'une profonde gratitude envers mes encadrants au sein du laboratoire "l'Intelligence Artificielle, Sciences des données et Systèmes Émergents", Professeurs M Mohamed Berrada et Mme Karima Aissaoui, qui ont su m'inspirer et me guider tout au long de cette aventure intellectuelle passionnante. Leur savoir-faire, leur expérience et leur expertise incontestable ont constitué un soutien indéfectible, leur engagement et leur disponibilité ont été des sources de motivation inépuisables pour moi.

Mon remerciement est également adressé aux rapporteurs, notamment M. Mohamed SABBANE, M. Aissam BERRAHOU et M. Khalid FARDOUSSE.

Je tiens également à exprimer ma gratitude aux membres du jury pour leur évaluation rigoureuse et constructive de ce travail, notamment M. Mohamed LAMRINI (Président du jury), M. Mohammed BERRADA, Mme Karima AISAOUI, M. Mohamed SABBANE, M. Aissam BERRAHOU, M. Khalid FARDOUSSE, Mme. Nour El Houda CHAOUI et M. Hakim EL FADILI.

Je tiens à adresser mes sincères remerciements à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail de recherche au sein du laboratoire "l'Intelligence Artificielle, Sciences des données et Systèmes Émergents" de l'Ecole Nationale des Sciences Appliquées de Fès.

Résumé

Dans le cadre de cette thèse, nous avons travaillé sur les défis et les opportunités liés à l'utilisation des données massives dans l'e-learning et les technologies de formation en ligne. Au fil de nos recherches, nous avons constaté que la croissance exponentielle des données liées à l'apprentissage en ligne ouvrait de nouvelles perspectives pour améliorer l'efficacité de ce processus. Toutefois, cette évolution rapide soulignait également les défis cruciaux auxquels il fallait faire face pour exploiter au mieux ces données massives.

Pour répondre aux défis de l'apprentissage en ligne, nous avons proposé deux approches novatrices. La première approche, l'ERSDO (système de recommandation E-learning avec des ontologies dynamiques), exploite l'ontologie dynamique et le web sémantique pour recommander des contenus d'apprentissage personnalisés en fonction des préférences et des besoins des apprenants. Cette approche permet une personnalisation accrue de l'apprentissage et une meilleure adéquation entre les besoins individuels des apprenants et les contenus proposés. La deuxième approche est une méthode de classification des objets d'apprentissage, qui combine une approche de classification multi-labels avec la logique floue et des techniques de web-mining pour identifier les objets les plus pertinents pour répondre aux besoins d'apprentissage des utilisateurs. Cette approche offre une sélection plus précise des objets d'apprentissage en fonction des besoins spécifiques des utilisateurs. En combinant ces deux approches, nous avons ainsi proposé des solutions pour améliorer l'efficacité et la pertinence de l'apprentissage en ligne.

Mots-clés : Données massives, Système de recommandation, Ontologie dynamique, Web sémantique, Classification des objets d'apprentissage, Classification multi-labels, Logique floue, Web-mining.

Abstract

In this thesis, we worked on the challenges and opportunities related to the use of big data in e-learning and online training technologies. We found that the exponential growth of data related to online learning opened up new perspectives for improving the effectiveness of this process. However, this rapid evolution also highlighted the crucial challenges that needed to be addressed to exploit this massive data best.

To address the challenges of online learning, we proposed two innovative approaches. The first approach, the ERSO (E-learning Recommendation System with Dynamic Ontologies), leverages dynamic ontology and the semantic web to recommend personalized learning content based on learners' preferences and needs. This approach allows for greater customization of learning and better alignment between individual learners' needs and the content offered. The second approach is a method for classifying learning objects, which combines a multi-label classification approach with fuzzy logic and web-mining techniques to identify the most relevant objects to meet users' learning needs. This approach offers a more precise selection of learning objects based on users' specific needs. By combining these two approaches, we proposed solutions to improve the efficiency and relevance of online learning.

Keywords : Big data, Recommendation system, Dynamic ontology, Semantic web, Learning object classification, Multi-label classification, Fuzzy logic, Web-mining.

Table des matières

Dédicaces	i
Remerciements	ii
Résumé	iv
Abstract	v
Liste des abréviations	xii
Introduction générale	1
1 Introduction	1
2 Problématique.	3
3 Approches proposées	5
4 Organisation du manuscrit	7
Chapitre I E-learning : Théorie et évolution	9
1 Introduction	9
2 Histoire de l'e-learning	11
2.1 Première phase (de 1960 au début des années 1980).	12
2.1.1 PLATO - Logique programmée pour l'opération d'enseignement automatique	12
2.1.2 TICCIT – Temps partagé, interactif, Contrôlé par l'ordinateur, Télévision d'information	13
2.2 Deuxième phase (de 1980 au début des années 1990)	14
2.3 Troisième phase (de la fin des années 1990 à aujourd'hui)	14
3 Définition de l'E-learning	15
3.1 Formation assistée par ordinateur (CBT)	16
3.2 Formation basée sur le web (WBT).	17

3.3	Plateforme d'apprentissage	17
3.4	Séminaire virtuel	17
4	Outils et Systèmes e-learning	18
4.1	Learning management system (LMS)	19
4.2	Course Management System (CMS)	19
4.3	Learning Content Management System (LCMS)	21
4.4	MOOC (Massive Open Online Courses)	22
5	Normes et standard	23
5.1	Spécifications SCORM	23
5.2	Le standard LOM	25
5.2.1	Structure de LOM	26
5.3	IMS-LD	28
5.4	Plateformes e-learning	30
5.5	Plateforme Moodle.	32
5.6	Plateforme Blackboard	33
5.7	Plateforme Chamilo	33
6	M-learning : apprentissage mobile	34
6.1	Quelques définitions	35
6.2	Classement de l'apprentissage mobile	35
6.3	Avantages et défis du m-learning	37
6.3.1	Avantages du m-learning	37
6.3.2	Défis et limites du m-learning	38
6.4	Caractéristiques de l'apprentissage mobile	40
6.5	Principes de l'apprentissage mobile	41
7	Conclusion	41

Chapitre II	Big data et systèmes de recommandation au service de l'apprentissage	43
1	Introduction	43
2	Big Data	45
2.1	Définition.	45
2.2	Brève histoire	47
2.3	Caractéristiques de Big Data	48
2.3.1	Volume.	48
2.3.2	Vélocité	48

2.3.3	Variété	48
2.3.4	Véracité	48
2.3.5	Value	49
2.4	Technologies pour l'analyse Big Data	49
2.4.1	Hadoop.	49
2.4.2	HDFS – Hadoop Distributed File System	52
2.4.3	MapReduce	53
2.4.4	Apache spark	54
2.4.5	Apache YARN.	56
2.4.6	Apache Flume.	58
2.4.7	Apache Scoop	59
2.5	Travaux connexes	60
2.6	Intégration du big data dans les systèmes d'apprentissage en lignes	65
3	Les systèmes de recommandation.	67
3.1	Définition.	67
3.2	Brève histoire	68
3.3	Les domaines d'application	70
3.4	Classification des systèmes de recommandation	70
3.4.1	Filtrage basé sur le contenu	71
3.4.2	Filtrage collaboratif	75
3.4.3	Systèmes hybrides	80
3.4.4	Recommandation communautaire	82
3.4.5	Recommandation basée sur le filtrage démographique	83
3.4.6	Recommandation à base de connaissance	84
3.5	Critères d'évaluation des systèmes de recommandation	85
3.5.1	Qualité de la prédiction.	85
3.5.2	Qualité de l'ensemble des recommandations	86
3.6	Utilisation des systèmes multi-agents dans les systèmes de recom- mandation	88
3.7	Notre Approche de recommandation basé sur les systèmes multi- agents	90
3.7.1	Méthodologie	90
3.7.2	Résultats expérimentaux	91
3.7.3	Discussion	93

4	Conclusion	93
---	----------------------	----

Chapitre III Mise en œuvre d'un modèle de recommandation

	E-learning à base d'ontologie et du web sémantique	96
1	Introduction	96
2	Le web sémantique et l'apprentissage en ligne	98
2.1	Le Web sémantique	98
2.2	Le Web sémantique social	101
2.3	Le web sémantique social et L'e-learning	102
3	Les ontologies	103
3.1	Structure d'une ontologie	105
3.1.1	Les concepts	105
3.1.2	Les relations	106
3.1.3	Les axiomes	106
3.1.4	Les instances	107
3.2	Langages de description d'ontologie	108
3.3	Travaux connexes sur les systèmes de recommandation basé sur l'ontologie dans l'E-learning	109
4	Approche proposée	111
4.1	Perspective modulaire du système proposé	112
4.2	Perspective en couche du système proposé	114
4.3	L'ontologie du profil de l'apprenant (Learner Profile Ontology : LPO)	116
4.3.1	Attributs de learner Profil (l'apprenant)	117
4.3.2	Jeu de données du profil des apprenants	120
4.3.3	Prétraitement des attributs de l'apprenant	120
4.4	L'architecture de l'ontologie des ressources matérielles (Material Resource Ontology : MRO)	121
4.4.1	L'ontologie de ressource matérielle (MRO)	122
4.4.2	Méthodes pour formaliser des ressources de formation	122
4.4.3	Ontologie dynamique de relation apprenant-cours (Learner- Course Relation Ontology : LCRO))	123
4.4.4	Similarité de LCRO	124
4.4.5	Algorithme de recommandation utilisé par LCRO	124

5	Résultats expérimentaux	125
5.1	Analyse du jeu de données utilisée par LCRO	127
5.2	Processus expérimental	129
5.3	Précision du classificateur	130
5.4	Discussion	131
6	Conclusion	132

Chapitre IV Implémentation d'un système de recommandation basé sur la Classification multi-labels des objets d'apprentissage 135

1	Introduction	135
2	Preliminaires	137
2.1	Objets d'apprentissage	138
2.2	La classification multi-Labels (MLC)	143
2.3	Types de Classification Multi-label	144
2.4	Fuzzy C-Means (FCM)	146
2.4.1	Les types de FCM	148
2.4.2	Fuzzy C-Means (FCM) et les objets d'apprentissages	150
2.5	Data Mining	151
2.5.1	Définition.	151
2.5.2	Origine du concept Data mining	151
2.5.3	Méthodes d'exploration de données	152
2.5.4	Data mining - Processus	152
3	L'apprentissage automatique	154
3.1	L'apprentissage supervisé	155
3.2	Les algorithmes d'apprentissage non supervisé	156
3.3	L'apprentissage par renforcement.	157
4	Travaux relatifs aux objets d'apprentissage dans les systèmes d'apprentissage en ligne	161
5	Notre approche	164
5.1	La classification des objets d'apprentissages basée sur les techniques de Clustering et Feature Selection	164
5.2	Évaluation et discussion	167
5.2.1	Corpus utilisé	167
5.2.2	Procédure expérimentale et mesure de la performance	168

5.2.3 Réduction de la dimensionnalité	170
6 Conclusion	172
Conclusion générale et perspectives	174
1 Synthèse	174
2 Bilan des travaux	176
3 Perspectives.	178
Annexes	180
A Exemple d'implémentation de l'approche proposée ERSDO	180
1 création d'une ontologie pour un domaine lié aux ressources d'appren- tissage en ligne des mathématiques.	184
2 Création des profils d'utilisateurs en fonction de leur interaction avec les ressources d'apprentissage en ligne :	188
3 Génération des recommandations pour un utilisateur en fonction de son profil et de l'ontologie :	189
4 La liste des recommandations	192
B Productions scientifiques	194
1 Publications	194
2 Communications.	194
Bibliographie	196

Liste des abréviations

API	Application Programming Interface
CBT	Computer-Based Training
CMS	Course Management System
CNN	Convolutional Neural Network
DAO	Data Access Object
DTO	Data Transfer Object
EAD	Enseignement à distance
E-learning	Electronic Learning
ERSDO	E-learning Recommender System based on Dynamic Ontology
FAD	Formation à Distance
FCM	Fuzzy C-means
FOAD	Formation Ouverte A Distance
HDFS	Hadoop Distributed File System
HTML	Hypertext Markup Language
IMS	Instructional Management Systems
IEEE	Institute of Electrical and Electronics Engineers
JSON	JavaScript Object Notation
KIF	Knowledge Interchange Format
KNN	K-Nearest Neighbors
LCMS	Learning Content Management System
LCRO	Learner-Course Relation Ontology
LMS	Learning Management System

LOM	Learning Object Metadata
LPO	Learner Profile Ontology
LTSC	Learning Technology Standards Committee
M-learning	Mobile Learning
MOOC	Massive Online Open Courses
Moodle	Modular Object-Oriented Dynamic Learning Environment
MRO	Material Resource Ontology
OWL	Web Ontology Language
PLATO	Programmed Logic for Automatic Teaching Operation
RNN	Recurrent Neural Network
SCO	Sharable Content Object
SCORM	Sharable Content Object Reference Model
SGBD	Système de Gestion de Base de Données
SKOS	Simple Knowledge Organization System
SQL	Structured Query Language
SVM	Support Vector Machine
SWS	Semantic Web Services
TF-IDF	Term Frequency-Inverse Document Frequency
TI	Technologie d'Information
TIC	Technologie d'Information et de Communication
TICCIT	Time-shared Interactive Computer Controlled Information Television
VLE	Virtual Learning Environment
WBT	Web-Based Training
XML	Extensible Markup Language
YARN	Yet Another Resource Negotiator

Liste des figures

Figure 1 : Structure d'un LMS	20
Figure 2 : Principes de base de SCORM	25
Figure 3 : représentation des éléments de la hiérarchie du modèle de LOM	27
Figure 4 : Modèle conceptuel de la conception globale de l'apprentis- sage	31
Figure 5 : Positionnement du m-learning	36
Figure 6 : les principes de M-learning	41
Figure 7 : le volume des données massive d'après Statista 2022	46
Figure 8 : Architecture du framework Hadoop	51
Figure 9 : Schéma de traitement dans l'architecture HDFS	52
Figure 10 :Modèle MapReduce	53
Figure 11 :Composants d'apache Spark	56
Figure 12 :composants d'apache Hadoop YARN	57
Figure 13 :architecture d'apache Flume	58
Figure 14 :Architecture de Apache Sqoop	60
Figure 15 :Framework Big Data pour les systèmes d'apprentissage en ligne	67
Figure 16 :Classification des systèmes de recommandation	71
Figure 17 :Filtrage basé sur le contenu	72
Figure 18 :Filtrage collaboratif	76
Figure 19 :Système de recommandation hybride	81
Figure 20 :Recommandation à base de connaissance	84
Figure 21 :Précision et Rappel concernant tous les Items	87
Figure 22 :Proposition d'architecture de système de recommandation pour l'apprentissage en ligne avec un système multi-agents .	91
Figure 23 :L'en-tête des cours "DataFrame"	92

Figure 24 :Le modèle en couches original selon Tim Berners-Lee	99
Figure 25 :Une version révisée du modèle en couches de Steve Bratt .	100
Figure 26 :Un exemple à partir d'un monde en blocs, avec les objets a,b,c,d,e dans deux couches d'objets différentes	104
Figure 27 :les ontologies utilisées du système proposé	112
Figure 28 :Perspective modulaire du système proposé	114
Figure 29 :Perspective en couche du système proposé	114
Figure 30 :Architecture du système de recommandation e-learning basé sur l'ontologie dynamique (ERSDO)	116
Figure 31 :Ontologie de l'apprenant (LPO)	118
Figure 32 :L'architecture de l'ontologie des ressources matérielles . . .	122
Figure 33 :Architecture de l'ontologie dynamique de l'apprentissage en ligne	124
Figure 34 :le flux de travail de notre système de recommandation d'on- tologie proposé	125
Figure 35 :Algorithme 1 : Top N Recommandations en utilisant LCRO	126
Figure 36 :Precision, Recall et F1-measure de ERESDO et CONTENT BASED	132
Figure 37 : : Représentation d'un modèle de contenu pour un LCMS avec des objets d'apprentissage réutilisables.	140
Figure 38 :Représentation d'un objet d'apprentissage composé d'un sous- objet didactique et d'un sous-objet technique	142
Figure 39 :Illustration de notre conception avec l'algorithme de classi- fication multi-label combiné aux techniques Fuzzy C-Means et Web-Mining.	166
Figure 40 :le nombre d'objets d'apprentissage pour les ensembles de données Moodle, Blackboard et Schoology.	168
Figure 41 :Création de la base de donnée	181
Figure 42 :Définition l'espace de noms de l'ontologie et importation de toutes les ontologies externes nécessaires	184
Figure 43 :Définition des classes de l'ontologie	185
Figure 44 :Définition des propriétés de l'ontologie	186
Figure 45 :Définition des individus dans l'ontologie	187

Figure 46 :la hiérarchie des classes dans l'ontologie	187
Figure 47 :Création des profils d'utilisateurs en fonction de leur inter- action avec les ressources d'apprentissage en ligne	188
Figure 48 :Génération des recommandations pour un utilisateur en fonc- tion de son profil et de l'ontologie	190
Figure 49 :La liste des recommandations	193

Introduction générale

Sommaire

1	Introduction	1
2	Problématique	3
3	Approches proposées	5
4	Organisation du manuscrit	7

1 Introduction

L'utilisation de l'apprentissage en ligne a véritablement transformé le domaine de l'éducation en tirant profit des avancées technologiques de l'information et de la communication (TIC). Ainsi, avec l'essor de l'internet et des technologies numériques, les apprenants peuvent désormais accéder à des ressources pédagogiques et à des supports de cours dans n'importe quel emplacement et à tout moment, ce qui représente un changement majeur par rapport à l'apprentissage traditionnel en classe limité par des contraintes d'espace physique et de temps. En outre, l'apprentissage en ligne offre une flexibilité accrue et permet aux apprenants de personnaliser leur expérience d'apprentissage en fonction de leurs besoins et préférences individuels. Cette personnalisation peut inclure la possibilité d'apprendre à leur propre rythme, d'accéder à une variété de ressources multimédia et de s'engager dans des activités d'apprentissage interactives et col-

laboratives avec leurs collègues et leurs formateurs. En analysant les données des apprenants, telles que leurs styles d'apprentissage, leurs préférences et leurs performances, les systèmes d'apprentissage en ligne peuvent recommander des contenus d'apprentissage adaptés, identifier les besoins spécifiques dans lesquels un soutien supplémentaire peut être nécessaire et fournir un retour en temps réel pour améliorer le cycle d'apprentissage.

De plus, la croissance des plateformes d'apprentissage en ligne au cours des dernières années a eu un effet profond sur l'éducation, tant en termes d'accès aux multiples choix de ressources éducatives que de quantité de données générées par ces plateformes. Ainsi, les plateformes d'apprentissage sont devenues une partie intégrante de l'enseignement en ligne, fournissant une structure importante de communication entre les apprenants et les formateurs et offrant aux apprenants une expérience d'apprentissage plus flexible. Les données de ces plateformes peuvent être utilisées pour identifier les tendances et les modèles de performance des apprenants, ainsi que pour évaluer l'efficacité des méthodes d'enseignement et des ressources pédagogiques.

Afin d'optimiser les performances en matière d'enseignement et d'apprentissage, les spécialistes de l'éducation se tournent de plus en plus vers le traitement de données pour recueillir des informations précieuses telles que les interactions entre les apprenants, les évaluations, les progrès de l'apprentissage et les retours d'information. Toutefois, il est primordial d'utiliser des technologies innovantes comme le "Big Data", l'exploration des objets d'apprentissage, et les théories de classification pour analyser efficacement ces données et en tirer des enseignements utiles pour les pratiques d'enseignement et d'apprentissage.

Dans le domaine de l'apprentissage en ligne, les systèmes e-learning collectent une grande quantité d'informations sur les apprenants, les objets d'apprentissage et les activités d'apprentissage. Pour cette raison, un système de recommandation performant est crucial pour tout système d'apprentissage en ligne. Cependant, la plupart des systèmes actuels sont conçus pour répondre aux demandes des apprenants à partir de bases de données statiques, qui ne prennent pas en compte les changements dynamiques de l'environnement d'apprentissage. C'est dans ce niveau-là que nous avons pensé à introduire l'ontologie dynamique. Cette méthode structurée et formelle permet de décrire les connaissances d'un domaine

spécifique. Pour améliorer la performance de la recommandation dans l'apprentissage en ligne, nous avons créé une ontologie de l'apprentissage, cette ontologie décrit les concepts, les relations et les dépendances entre les différents éléments de l'environnement d'apprentissage en ligne. Elle nous permet de mieux comprendre les interactions complexes qui se produisent entre ces différents éléments et donc d'optimiser la précision des recommandations.

Les technologies de classification représentent un domaine en pleine expansion dans le domaine de l'éducation, offrant un potentiel considérable pour l'analyse de données éducatives. En utilisant des algorithmes sophistiqués, ces technologies peuvent être utilisées pour classer et grouper des données en fonction de critères spécifiques, tels que les caractéristiques démographiques des apprenants, les styles d'apprentissage, ou les mesures de performance. Cette approche permet aux pédagogues de détecter des tendances et des modèles en fonction des facteurs quantifiables, ce qui peut aider à optimiser les pratiques d'enseignement et d'apprentissage. Par exemple, les algorithmes de classification peuvent être utilisés pour identifier des apprenants en difficulté, ce qui permet aux enseignants de fournir des interventions ciblées.

L'objectif central de cette thèse est double. Tout d'abord, on cherche à mettre en évidence comment l'analyse des données provenant des plateformes d'apprentissage en ligne peut apporter des informations précieuses sur les pratiques et les stratégies éducatives, en exploitant les avancées technologiques de l'ère numérique. De plus, elle vise à construire une nouvelle approche innovante, basée sur les nouvelles technologies, pour améliorer de manière dynamique la recommandation dans le domaine de l'apprentissage à distance. À travers une analyse nuancée de ces thèmes, cette recherche a pour ambition de susciter une nouvelle vague d'innovation dans le domaine de l'éducation, et de favoriser un environnement d'apprentissage plus adapté et plus efficace pour tous les apprenants.

2 Problématique

Le développement de l'e-learning et des technologies de formation en ligne a conduit à une augmentation spectaculaire de la quantité de données générées par les apprenants. Cette croissance rapide des données offre de nouvelles opportuni-

tés pour améliorer l'efficacité de l'apprentissage en ligne, mais soulève également des défis importants qui portent sur la question de comment exploiter efficacement ces données massives pour améliorer l'efficacité de l'apprentissage en ligne. En effet, la gestion de ces données volumineuses représente un défi majeur pour les systèmes e-learning. Les plateformes d'apprentissage en ligne génèrent des quantités massives de données, notamment des données sur les performances des apprenants, des données d'interaction et des données de journal du système. Ces données doivent être stockées et gérées efficacement pour être utilisées de manière constructive dans l'amélioration de l'efficacité de l'apprentissage en ligne.

En outre, la problématique des systèmes de recommandation inefficaces est un autre défi majeur pour les systèmes d'apprentissage en ligne. Les systèmes de recommandation actuels ne parviennent souvent pas à recommander des contenus d'apprentissage pertinents en raison de la complexité de l'environnement d'apprentissage et de l'évolution constante des préférences et des besoins des apprenants. Le développement d'un système de recommandation dynamique qui permet d'extraire des informations pertinentes des données massives collectées par les systèmes d'apprentissage en ligne est notre proposition pour résoudre ce problème. Cette méthode permet d'analyser en temps réel les données collectées et de recommander des contenus d'apprentissage pertinents pour chaque apprenant en fonction de ses besoins et de ses préférences spécifiques. Ainsi, en combinant l'ontologie dynamique et les approches de recommandation, les systèmes d'apprentissage en ligne peuvent fournir des recommandations plus précises et personnalisées pour chaque apprenant, améliorant ainsi l'efficacité de l'apprentissage et les résultats globaux.

Par ailleurs, la problématique de classification précise des objets d'apprentissage est un autre défi qui confronte les systèmes d'apprentissage en ligne. Les objets d'apprentissage peuvent être de nature et de complexité différentes, ce qui rend leur classification difficile. Néanmoins, une classification précise est essentielle pour améliorer la recommandation et personnaliser l'expérience d'apprentissage pour chaque apprenant.

Ainsi, les chercheurs ont exploré différentes méthodes, telles que l'apprentissage automatique, pour améliorer la précision de classification. Cependant, l'utilisation de telles méthodes nécessite des ressources importantes et une qualité

de données élevée. Il est donc crucial de trouver un équilibre entre la précision de la classification et les ressources disponibles. En somme, la gestion, l'analyse et la recommandation des contenus des systèmes d'apprentissage en ligne sont confrontées à plusieurs défis majeurs, tels que l'exploitation des données massives, la recommandation inappropriée et la classification des objets d'apprentissage.

3 Approches proposées

Pour mieux comprendre le contexte et prendre conscience des problématiques sous-jacentes, nous avons mené dans un premier temps une revue de la littérature sur l'utilisation des technologies Big Data dans les systèmes d'apprentissage en ligne. Cette démarche a abouti à la publication d'un premier article de conférence intitulé « Big Data in E-learning : Literature Review and Challenges », qui a été inclus dans les actes de la conférence internationale AI2SD'2020 (Advanced Intelligent Systems for Sustainable Development) indexé par Springer.

Pour mettre en lumière les connaissances acquises lors de cette revue de la littérature, nous avons forgé deux pistes de recherche dans le domaine des systèmes de recommandation pour les plateformes d'apprentissage :

Dans notre premier axe, nous avons présenté une approche innovante de recommandation, qui repose sur l'utilisation de l'ontologie dynamique et du web sémantique. Cependant, la mise en place d'un système de recommandation E-learning avec des ontologies dynamiques (ERSDO) a été un véritable défi pour nous, puisque d'une part les plateformes e-learning en libre accès posaient des problèmes en termes de personnalisation des contenus dynamiques, et d'autre part, la création de profils utilisateurs dynamiques s'est avérée complexe. L'ontologie de l'e-learning que nous avons proposé pour achever cet objectif est conçue comme un ensemble de ressources ontologiques homogènes qui ont été identifiées pour être utilisées dans multiples environnements d'apprentissage. Cette ontologie pourrait être expliquée à partir de deux perspectives, une perspective modulaire, et une perspective en couches qui approfondit la mise en œuvre de la vue modulaire. Cela a fait l'objet d'un article intitulé «ERSDO : E-learning Recommender System based on Dynamic Ontology » dans le journal «Education

and Information Technologies».

Dans le deuxième axe de recherche, nous avons établi un modèle théorique pour la classification des objets d'apprentissage en combinant une approche de classification multi-labels avec la logique floue. L'approche de classification multi-label considère la présence de plusieurs étiquettes ou catégories pour un seul objet d'apprentissage, tandis que la méthode de la logique floue tient compte de l'imprécision et de l'incertitude dans le processus de classification. Cette contribution, vise la proposition d'une nouvelle méthode de partage d'objets basée sur la classification combinée avec les techniques de web-mining. Cela permet d'améliorer la précision et l'efficacité globales du processus de classification. Dans cette contribution, l'accent est mis sur la catégorisation des objets d'apprentissage à l'aide de la classification multi-labels (MLC) et des algorithmes Fuzzy C-Means (FCM). L'algorithme FCM est une technique de regroupement qui utilise la mesure de la distance euclidienne pour améliorer la précision du processus de classification. L'utilisation du MLC garantit que chaque objet d'apprentissage sera identifié par plusieurs étiquettes ou catégories, ce qui permet une catégorisation plus nuancée et plus précise. Pour valider l'efficacité de l'approche proposée, une série d'expériences a été menée à l'aide d'ensembles de données du monde réel. Cette contribution a fait l'objet d'un article intitulé « Multi-Label Classification of Learning Objects Using Clustering Algorithms Based on Feature Selection » dans le journal « International Journal of Emerging Technologies in Learning ».

Pour parvenir à une contribution solide, nous avons entrepris dans un troisième article intitulé « Fuzzy C-Means, and web Mining Methods for Learning Object Analysis » une analyse approfondie des avantages offerts par les objets d'apprentissage. Nous avons ainsi étudié de manière approfondie cette catégorie d'outils afin de mieux comprendre leur fonctionnement et leurs bénéfices. Cette démarche nous a permis de proposer une méthode d'analyse novatrice, basée sur la classification des objets d'apprentissage avec les méthodes de web Mining .

En effet, ces objets sont au cœur des systèmes d'apprentissage en ligne, car ils permettent de fournir aux apprenants des contenus interactifs et structurés, qui peuvent être utilisés pour atteindre des objectifs pédagogiques spécifiques. Les objets d'apprentissage offrent ainsi de nombreux avantages, tels que la flexibilité, la

personnalisation, l’adaptabilité, l’interactivité, la modularité et la réutilisabilité.

Notre méthode d’analyse se concentre sur la classification des objets d’apprentissage en fonction de leurs caractéristiques et de leur contenu pédagogique. Cette classification nous permet d’identifier les objets les plus pertinents pour répondre aux besoins d’apprentissage des utilisateurs. Nous pouvons ainsi proposer des contenus adaptés à leurs préférences et à leurs niveaux de connaissances.

En somme, notre analyse approfondie des objets d’apprentissage nous a permis de mettre en lumière leur importance dans les systèmes d’apprentissage en ligne, ainsi que les bénéfices qu’ils offrent.

4 Organisation du manuscrit

La thèse est organisée en quatre chapitres, chacun abordant une thématique spécifique relative à l’e-learning et à l’analyse de données.

- Le premier chapitre fournit une vue d’ensemble de l’e-learning en présentant les théories sous-jacentes et les avancées récentes dans ce domaine. Nous examinons également les défis auxquels les systèmes d’apprentissage en ligne sont confrontés et les solutions proposées pour les surmonter.
- Le deuxième chapitre est consacré à l’utilisation des données massives (big data) dans le domaine de l’e-learning. Nous explorons les différentes sources de données disponibles et les techniques de collecte, stockage, analyse et visualisation des données pour améliorer la qualité de l’apprentissage en ligne. Nous abordons également les systèmes de recommandation et leur approche, les critères d’évaluation, ainsi que le développement de notre premier système de recommandation qui utilise les données massives des apprenants pour améliorer la qualité des recommandations.
- Le troisième chapitre décrit la mise en œuvre d’un modèle de recommandation e-learning basé sur l’ontologie et le web sémantique. Nous décrivons les principales étapes de la mise en place du modèle, de la conception de l’ontologie à l’implémentation du système de recommandation.
- Le quatrième et dernier chapitre se concentre sur la classification multi-labels des objets d’apprentissage. Nous proposons une méthode d’analyse

basée sur la classification des objets d'apprentissage, en utilisant des techniques de clustering et de sélection de caractéristiques pour améliorer la précision de la classification.

Enfin, la conclusion récapitule les principales contributions de la thèse et suggère des pistes pour les travaux futurs dans le domaine de l'e-learning et de l'analyse de données.

Chapitre I

E-learning : Théorie et évolution

Sommaire

1	Introduction	9
2	Histoire de l'e-learning	11
3	Définition de l'E-learning	15
4	Outils et Systèmes e-learning	18
5	Normes et standard	23
6	M-learning : apprentissage mobile	34
7	Conclusion	41

1 Introduction

Au XXI^e siècle, avec l'essor de l'informatique, la communication et les technologies électroniques, la manière d'enseigner et d'apprendre s'est également modernisée. Aujourd'hui, les livres et les cours ont été transformés en livres électroniques, en PDF (fichiers de documents portables), en webinaires, en présentations, etc. Ce type de système éducatif, appelé système d'apprentissage en ligne, nous a permis d'acquérir des connaissances de manière plus efficace, améliorant ainsi la qualité de l'enseignement.

L'utilisation de l'e-learning permet d'exploiter les avantages des TIC modernes pour optimiser l'apprentissage. Le débat pédagogique autour de l'utilisation de l'apprentissage en ligne et l'utilisation d'Internet aux établissements coïncide avec

un changement de paradigme dans la théorie pédagogique de l'apprentissage. On passe des théories cognitives à des théories constructivistes [1]. Celles-ci perçoivent les apprenants comme des sujets actifs et constructifs et développent ainsi des exigences pour plus d'auto-direction, de coopération et d'authenticité.

Avec la popularité qu'ont connue les nouvelles technologies, et la grande familiarisation des étudiants avec les outils numériques et l'internet, l'intégration des nouvelles technologies d'information et de communication (TIC) dans le domaine de la formation/apprentissage est devenue primordiale. Ainsi, le développement et l'amélioration des systèmes d'apprentissage ont pris de la cadence[2].

L'e-learning a acquis une importance cruciale en tant qu'outil éducatif pendant et après la pandémie Covid 19. Alors que les fermetures d'écoles et les mesures de distanciation sociale ont entraîné une perturbation majeure du processus éducatif, l'e-learning a fourni une alternative sûre et efficace à l'apprentissage en présentiel, permettant aux étudiants de continuer leur éducation à distance. En outre, l'e-learning a également permis aux enseignants de dispenser leurs cours en ligne, contribuant ainsi à maintenir un certain niveau de normalité dans le processus éducatif.

L'e-learning représente une méthode pédagogique qui se concentre sur l'intégration des nouvelles technologies dans le processus d'apprentissage [3]. Cette approche est appliquée dans divers contextes pour offrir aux apprenants un accès à l'information en dehors des cours traditionnels, renforçant ainsi leur apprentissage tout en éliminant les contraintes liées à leur présence physique dans les lieux de formation. Dans la suite, nous commençons ce chapitre par les différentes phases marquantes dans l'histoire de l'e-learning, nous présentons après les différentes définitions de l'e-learning, ainsi que quelques termes qui appartiennent au même champ lexical et qui sont apparus pour accompagner les nouvelles méthodologies de formation et d'apprentissage en ligne. Nous abordons ensuite les normes et standards de l'e-learning, ainsi que les différentes plateformes d'apprentissages.

2 Histoire de l'e-learning

Le nouveau monde de l'apprentissage a été annoncé dès le début des années 1960, avec l'apparition des machines d'apprentissage et des ordinateurs dans l'enseignement. Le développement de la technologie informatique s'est déroulé en trois phases. La première phase (dès les années 1960 au début des années 1980) a été marquée par l'utilisation d'ordinateurs centraux, la deuxième phase (dans les années 1980) par l'émergence des ordinateurs personnels (PC) et la troisième phase a commencé dans les années 1990 avec la mise en réseau des ordinateurs [4].

De nos jours, les dispositifs numériques connaissent une très grande popularité, non seulement dans le milieu des jeunes étudiants des universités, mais aussi dans le milieu des élèves des établissements d'enseignement primaire et secondaire. Cette situation mène les décideurs dans le domaine de l'éducation à en tirer profit, en introduisant la formation en ligne, ce qui a donné par la suite naissance à plusieurs termes relatifs à l'apprentissage en utilisant les nouvelles technologies.

Dans les années 1960, l'utilisation de machines d'apprentissage visait d'une part à accélérer l'éducation, et d'autre part à remplacer le manque d'enseignants par des machines. Dans les années 1980, l'utilisation des ordinateurs personnels (PC) s'est développée et, avec elle, l'utilisation d'offres d'apprentissage spéciales telles que la formation et la pratique, les tutoriels, les simulations et les jeux. La technologie n'a cessé d'évoluer et a favorisé l'émergence de programmes d'apprentissage multimédia pour de larges groupes d'utilisateurs. L'e-learning a connu un essor important avec la diffusion d'Internet à partir du milieu des années 1990, époque à laquelle s'est également établi le terme e-learning pour désigner l'apprentissage assisté par ordinateur.

Dans les années 1960, les universités américaines ont mis en place les premiers grands systèmes d'apprentissage assisté par ordinateur [5]. Conformément à la théorie d'apprentissage dominante de l'époque, le behaviorisme, ces offres étaient strictement orientées vers le comportement. La recherche et le développement de ces systèmes ont été largement soutenus par les autorités américaines. On peut

donc distinguer les trois phases suivantes :

- du début des années 1960 au début des années 1980 (caractérisées par l'utilisation de gros ordinateurs)
- du début des années 1980 au début des années 1990 (caractérisées par l'utilisation croissante d'ordinateurs personnels fixes)
- du début des années 1990 à nos jours (caractérisées par les PC multimédia en réseau).

2.1 Première phase (de 1960 au début des années 1980)

Les premières approches visant à soutenir l'apprentissage humain par l'utilisation de machines électroniques remontent à la fin des années 1950. Les précurseurs étaient les États-Unis, qui ont poussé ce développement. En 1960, en raison de l'augmentation rapide du nombre d'élèves aux États-Unis, un déficit de 135.000 enseignants a été identifié. L'automatisation de l'enseignement promettait également de résoudre ce problème. On pensait pouvoir combler ce manque d'enseignants par l'utilisation à grande échelle de machines à enseigner[6].

Les deux grands projets (PLATO et TICCIT) ont été financés par la « National Science Foundation of America » à hauteur d'environ 60 millions de dollars[7]. Ils étaient considérés à l'époque comme les systèmes dominants et ont influencé de manière décisive le développement de systèmes ultérieurs.

2.1.1 PLATO - Logique programmée pour l'opération d'enseignement automatique

Deux décennies avant l'arrivée du « World Wide Web » sur la scène, le système PLATO a fait œuvre de pionnier en matière de forums et de tableaux d'affichage en ligne, de courrier électronique, de salles de chat, de messagerie instantanée, de partage d'écran à distance et de jeux multi-joueurs, conduisant à l'émergence de ce qui est peut-être la première communauté en ligne au monde[8].

Le projet PLATO (Programmed Logic for Automatic Teaching Operation)[9] est considéré comme l'un des premiers grands projets de recherche et d'application de systèmes d'apprentissage assistés par ordinateur. La première version

du système a été introduite dès le début des années 1960 au « Computer-Based-Education Research Laboratory » de l'université de l'Illinois[9]. L'intention initiale de développer le système était de résoudre les problèmes déjà mentionnés auxquels le système éducatif américain était confronté à la fin des années 1950. Outre le fait que le projet PLATO a en quelque sorte créé la première communauté en ligne, c'est surtout la flexibilité du système qui a fait son succès. Le langage de programmation TUTOR a été développé pour permettre aux enseignants de développer leurs propres programmes d'apprentissage[9].

2.1.2 TICCIT – Temps partagé, interactif, Contrôlé par l'ordinateur, Télévision d'information

Le coût élevé de la création de programmes d'apprentissage pour le système PLATO a été l'une des raisons du lancement d'un autre projet, le projet TICCIT (Time shared, Interactive, Computer-Controlled, Information Television)[10].

Afin de réduire les coûts élevés du matériel, le système TICCIT a combiné la technologie innovante des micro-ordinateurs avec des récepteurs de télévision conventionnels déjà disponibles. C'est ainsi que les premières offres multifenêtres ont été rendues possibles. On a renoncé à l'implémentation d'un langage de programmation flexible, de sorte que les offres du système TICCIT sont structurées de manière uniforme[7].

Le système TICCIT était également un système multi-utilisateur, mais il n'était pas aussi puissant que le système PLATO. Alors que ce dernier permettait à un millier d'apprenants de travailler simultanément, le système TICCIT était limité à environ 120 apprenants simultanés. Bien que des résultats positifs aient été obtenus en termes de réussite des apprenants, il s'est avéré que cette réussite dépendait de nombreux autres facteurs. Outre les connaissances préalables des apprenants, il s'agissait notamment du rôle des enseignants et de la conception didactique globale des programmes d'apprentissage[11]. Dès les années 1970 l'ordinateur était déjà considéré comme un outil alternatif pour transmettre des contenus d'apprentissage[11].

2.2 Deuxième phase (de 1980 au début des années 1990)

Dans les années 1970, les problèmes liés à la réalisation d'offres d'apprentissage au moyen de grands systèmes informatiques ont entraîné un désenchantement et une baisse d'intérêt pour l'utilisation des nouvelles technologies dans les scénarios d'apprentissage. L'apparition des premiers ordinateurs personnels (PC) et leur rapide pénétration du marché ont eu pour conséquence qu'ils ont également été utilisés pour concevoir des scénarios d'apprentissage. L'ordinateur personnel était considéré comme un outil d'apprentissage abordable, particulièrement adapté à l'apprentissage tout au long de la vie. Comme dans les années 1950, on a évoqué de nouveau une crise de l'éducation qui ne pourrait être maîtrisée que par l'utilisation massive d'ordinateurs personnels[12]. L'utilisation de l'ordinateur dans l'enseignement suscitait à nouveau de grands espoirs, mais des voix critiques s'élevaient également, craignant que l'existence des établissements d'enseignement classiques soit menacée.

Les PC des années 1980 fonctionnaient généralement comme des systèmes autonomes. Ce n'est que dans les grandes entreprises que des réseaux locaux ont été créés. Ce n'est qu'avec le développement du langage de description des pages Hypertext Markup Language (HTML) et l'introduction à grande échelle du protocole de transmission (TCP/IP) que la base d'un échange de données simple et de nouveaux canaux de communication dans des réseaux mondiaux a été rendue possible.

En conclusion de la deuxième phase, on peut dire que l'introduction de l'ordinateur personnel a permis de multiplier les expériences avec les possibilités offertes par les nouvelles technologies, mais qu'il n'a pas été possible de parvenir à une application durable des systèmes. Comme dans la première phase, cet échec peut à nouveau être attribué aux exigences parfois excessives envers les nouvelles technologies.

2.3 Troisième phase (de la fin des années 1990 à aujourd'hui)

La forte diffusion des ordinateurs dans les ménages privés, qui a débuté dans les années 1980, s'est poursuivie dans les années 1990. Deux évolutions en particu-

lier ont conduit à la poursuite de la diffusion des ordinateurs personnels également à une large utilisation des offres d'e-learning :

- Multimédia : les nouveaux médias, associés à la demande du nouveau marché, ont donné lieu à de nouvelles visions de ce que devait être l'apprentissage par ordinateur. Le multimédia et les systèmes hypertextes promettaient une individualisation du processus d'apprentissage, dont on espérait qu'elle augmenterait l'efficacité de l'apprentissage.
- Mise en réseau : le milieu des années 1990 a vu le début d'une vaste mise en réseau des PC utilisés jusqu'à présent localement. Celui-ci est ainsi devenu un instrument de communication permettant la communication interpersonnelle et la coopération via des réseaux publics[13].

Alors qu'au début des années 1990, ce sont surtout les CD-ROM qui étaient utilisés comme supports de stockage, l'utilisation d'Internet s'est généralisée. La mise en réseau croissante des ordinateurs a conduit à une plus grande offre d'apprentissage basées sur le réseau. Les plates-formes standardisées d'apprentissage ont favorisé cette évolution[14].

Le terme d'e-learning, apparu à cette époque, est devenu un « buzz word » désignant toutes les formes d'apprentissage assisté électroniquement[15]. Le concept le plus important de l'utilisation des offres d'e-learning est désormais la combinaison de phases en présentiel et en ligne, appelée *blended learning* (apprentissage mixte). Dans ce qui suit, nous présentons les différentes définitions du l'e-learning, ainsi que quelques termes appartenant au même champ lexical qui semblent accompagner les nouvelles méthodes de formation et de e-learning.

3 Définition de l'E-learning

Il existe une grande diversité de termes pour le domaine de l'e-learning, qui s'est encore élargie avec l'utilisation du Web 2.0. L'e-learning - abréviation de *electronic learning* - désigne les processus d'apprentissage qui sont guidés, dirigés ou soutenus électroniquement sous une forme ou une autre[16]. Le terme s'établit de plus en plus comme un terme générique pour tous les types d'apprentissage assistés par des médias et comprend aussi bien l'apprentissage avec un logiciel

installé localement (programmes d'apprentissage ou CD/DVD) et l'apprentissage via Internet. Ainsi, l'apprentissage est soutenu par des médias électroniques - surtout par l'utilisation d'ordinateurs - dans des cadres hors ligne et en ligne.

L'e-learning est considéré comme l'un des concepts pédagogiques pour l'utilisation des technologies numériques. Dans ce domaine, la plupart de ces activités d'enseignement et d'apprentissage se déroulent en ligne via Internet. Des programmes hors ligne sur CD-ROM sont également possibles, mais ne sont plus d'actualité de nos jours[17].

D'autres termes sont souvent utilisés comme synonymes d'e-learning. Des désignations telles que blended learning, apprentissage en ligne, e-teaching, apprentissage multimédia, apprentissage basé sur les médias ou encore apprentissage par Internet peuvent être utilisées[18].

Le terme e-learning n'est utilisé que depuis 1998. Depuis cette année, plusieurs générations d'e-learning ont déjà vu le jour en raison de divers développements numériques[19]. Dans la première génération, les données telles que les présentations ou les documents pouvaient être stockées sous forme électronique au sein d'une organisation, ce qui permettait à tous les collaborateurs d'accéder aux données. La génération suivante, la formation assistée par ordinateur (FAO), utilisait des programmes d'apprentissage sur CD-ROM. La troisième génération a utilisé pour la première fois des programmes d'apprentissage assistés par ordinateur via Internet ou Intranet[20]. Les quatre formes techniques de l'e-learning sont :

3.1 Formation assistée par ordinateur (CBT)

Ce terme désigne les programmes d'apprentissage multimédias classiques sur CD-ROM qui peuvent être lancés directement à partir du support ou qui doivent être installés au préalable sur l'ordinateur[21]. L'ordinateur peut ici assumer les trois fonctions didactiques de l'enseignement : transmettre des connaissances, proposer des exercices et évaluer ces derniers.

3.2 Formation basée sur le web (WBT)

Le programme d'apprentissage se trouve dans un environnement basé sur le web ou le réseau. L'apprentissage coopératif de plusieurs participants dans un programme d'apprentissage est possible[22]. Dans le cas des WBT, l'ordinateur ne joue qu'un rôle d'intermédiaire entre les participants.

3.3 Plateforme d'apprentissage

Les plateformes d'apprentissage (également appelées portails d'apprentissage en ligne ou systèmes de gestion de l'apprentissage, en abrégé LMS) sont des centres de formation virtuels pour la gestion des offres d'apprentissage en ligne et des utilisateurs[23]. Elles constituent une évolution logique des TIC et proposent, en plus des TIC et des TBI, du matériel d'apprentissage complémentaire (liens complémentaires, manuels d'étude, rapports pratiques, films pédagogiques, applications audiovisuelles, etc.) ainsi que des formes d'interaction et des possibilités de communication améliorées entre les participants, telles que des forums et des chats (thématiques). En outre, les apprenants bénéficient d'une assistance technique et professionnelle à laquelle ils peuvent s'adresser en cas de problèmes liés à la technique ou au contenu[24].

3.4 Séminaire virtuel

Un séminaire virtuel est une formation en ligne synchrone, c'est-à-dire simultanée pour tous les participants, mais indépendamment du lieu[25]. Elle est dirigée par un enseignant qui peut présenter et expliquer le contenu de la formation via un flux vidéo (avec son). Parallèlement, les participants peuvent poser des questions directement à l'enseignant. Les séminaires virtuels combinent donc les avantages des formations en classe traditionnelles (possibilité pour les participants de poser des questions) avec les avantages des formations en ligne[25].

L'apprentissage et l'enseignement dans l'espace virtuel ne nécessitent pas seulement une infrastructure fonctionnelle comme les ordinateurs et Internet, mais aussi et surtout des compétences spécifiques pour l'apprentissage et l'enseignement. Un rapport pédagogique entre l'autonomie et la discipline des apprenants et le rôle de médiateur des enseignants est nécessaire. Cela nécessite un

changement important de la conception traditionnelle de l'enseignement[26]. Les apprenants déterminent eux-mêmes les contenus qu'ils souhaitent apprendre, pendant combien de temps, à quel rythme, quand et où. Les tâches d'apprentissage peuvent être répétées à volonté dans les programmes. Cette variante s'applique à la plupart des programmes d'apprentissage assisté par ordinateur. D'autres variantes d'apprentissage en ligne, telles que les vidéoconférences, les cours en ligne ou les simulations sur ordinateur, suivent chacune leur propre voie de développement des compétences, mais supposent un minimum d'autonomie. En même temps, l'enseignement et l'apprentissage avec les médias numériques présentent l'avantage, par rapport aux formes traditionnelles, de permettre aux apprenants de construire et d'organiser leurs processus d'apprentissage de manière de plus en plus autonome ou en collaboration avec d'autres apprenants. Ils peuvent produire de manière autonome les résultats de leur traitement de tâches d'apprentissage globales, les présenter aux autres apprenants, aux enseignants et éventuellement aussi à d'autres experts spécialisés et soumettre à la discussion. En résumé, le terme e-learning décrit un très large éventail de dispositifs d'apprentissage assistés par voie électronique. Des petits logiciels d'apprentissage aux grands systèmes de gestion de l'apprentissage, tout est concerné par ce terme. Ces systèmes peuvent se différencier très finement dans leur concept de base et être dotés de nombreuses fonctions différentes.

4 Outils et Systèmes e-learning

Divers outils et systèmes sont fournis pour assurer l'interaction entre différents utilisateurs. Ces systèmes sont considérés comme des canaux d'échange qui gèrent à la fois la communication synchrone et asynchrone et permettent de gérer les éléments pédagogiques. Les utilisateurs de ces systèmes sont les différents acteurs impliqués dans le processus e-learning, à savoir : les apprenants, les formateurs et les administrateurs. Selon des rôles et des droits d'accès bien définis, les outils de gestion de contenus éducatifs proposent des dispositifs d'accès à distance à des contenus éducatifs à l'aide de différents outils : email, forums, chat, visioconférence, tableaux blancs interactifs, ou autres. Le but de l'utilisation de ces outils est de fournir aux apprenants un environnement collaboratif afin de ne pas perdre la cohésion et la stimulation de la classe. Bien qu'ils aient tous les mêmes

objectifs, il existe plusieurs types de systèmes de gestion de l'apprentissage. Dans ce qui suit, nous définissons certains de ces types.

4.1 Learning management system (LMS)

Le LMS (Learning Management System) est une plateforme d'apprentissage qui permet de proposer et de gérer des cours en ligne complets. Des contenus d'apprentissage sont mis à disposition, qui peuvent être traités par l'apprenant et organisés par l'enseignant. Il offre également des fonctions qui permettent la communication entre l'apprenant et l'enseignant. Le LMS intègre des outils de communication tels que les forums, la messagerie électronique et la messagerie instantanée dans la plupart des cas, et est conçu pour assurer un espace de communication entre les étudiants et les enseignants afin de les aider à accomplir l'approche basée sur les tâches de l'apprentissage collectif en ligne[27].

La plate-forme open source Moodle est un exemple de LMS[28]. Son concept est de proposer des espaces de cours dans lesquels les utilisateurs peuvent s'inscrire pour participer à ce cours, qui est organisé et supervisé par un enseignant.

Ce système est considéré comme le cœur des dispositifs d'apprentissage en ligne. Il offre plusieurs fonctionnalités, notamment :

- Organisation et gestion de formations sur le réseau Internet ou intranet,
- diffusion de contenus pédagogiques,
- Communication entre les différents acteurs du processus e-learning,
- Gestion des inscriptions et connexion aux systèmes d'apprentissage en ligne,
- Gestion des activités d'apprentissage en ligne.

4.2 Course Management System (CMS)

Le CMS ou système de gestion de contenu est essentiellement conçu pour soutenir les cours éducatifs ou universitaires. Il permet à l'instructeur de créer un site web de cours, où des documents peuvent être téléchargés dans des formats populaires tels que Word, Power Point, etc, sans avoir besoin à les convertir dans

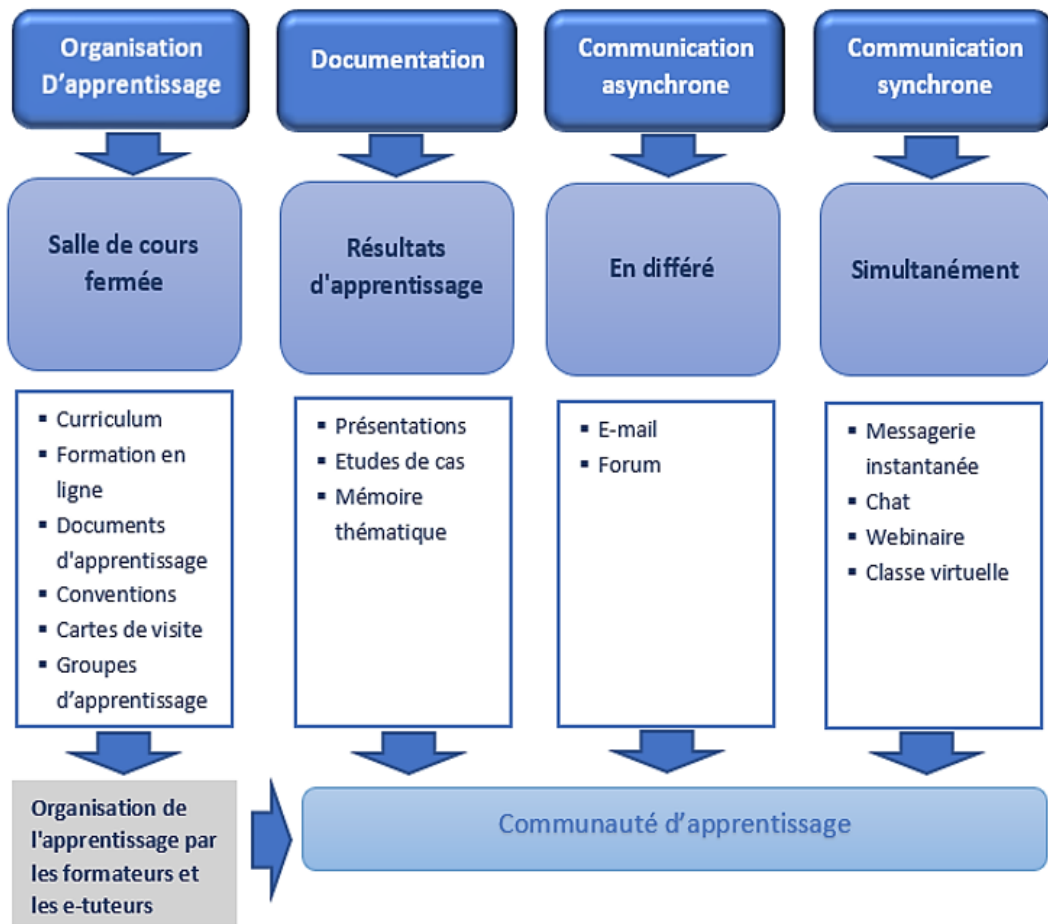


FIGURE 1 : Structure d'un LMS

un format web tel que le HTML. Cela nécessite peu de compétences spécialisées, ce qui fait d'un CMS le choix le plus courant des formateurs. Il soutient également efficacement l'apprentissage à distance grâce à son application robuste de forum de discussion. Les formateurs affichent l'essentiel du cours qui conduit les apprenants à travers diverses activités d'apprentissage, après, les formateurs supervisent les discussions du cours par le biais du forum de discussion[29].

En général, Le CMS fournit aux formateurs un cadre et un ensemble d'outils qui leur permettent de créer le contenu de cours en ligne, puis de gérer et d'enseigner ces cours, ce qui inclut les interactions avec les apprenants qui suivent ces cours[29]. D'après cette définition, il est clair que l'objectif du CMS est de fournir des fonctionnalités et des outils qui permettent, aux formateurs de créer

le contenu des cours en ligne et d'organiser l'enseignement et l'interaction avec leurs apprenants.

4.3 Learning Content Management System (LCMS)

Le « Learning Content Management System » (en abrégé : LCMS) est un système de gestion de contenu spécialement conçu pour le développement, la gestion et la distribution de contenu éducatif en ligne. Un LCMS permet aux auteurs de contenu et aux éducateurs de créer, stocker et gérer des ressources éducatives numériques telles que des cours, des modules d'apprentissage, des quiz et des évaluations. Il fournit également des fonctionnalités pour organiser, réviser et publier le contenu éducatif sur des plateformes en ligne. Le LCMS est utilisé dans les contextes de l'e-learning pour simplifier le processus de création et de gestion du contenu éducatif en ligne, permettant ainsi aux éducateurs et aux apprenants de tirer pleinement parti de l'apprentissage en ligne.

Le terme LCMS est souvent confondu avec LMS et considéré comme une version mise à jour de LMS. Cependant, les deux applications sont différentes car elles se complètent et se concentrent sur des fonctions différentes. Nous abordons par la suite la différence entre ces deux types. Les mots-clés qui montrent la différence entre ces deux termes peuvent être vus dans le seul mot qui les sépare, contenu. Paulsen[30] définit le LCMS comme un environnement dans lequel les enseignants/développeurs peuvent créer, stocker, gérer, réutiliser et diffuser du contenu d'apprentissage à partir d'un référentiel d'objets central. En règle générale, le LCMS traite du contenu basé sur le modèle d'objet d'apprentissage. D'autre part, Oakes[31] définit le LCMS comme une application pour "créer, stocker, assembler et diffuser un contenu d'apprentissage en ligne personnalisé sous la forme d'objets d'apprentissage". L'objectif principal de LCMS se concentre sur le contenu, car il prend en considération les obstacles de la création, la réutilisation, la gestion et la livraison du contenu [31]. Cependant, le LMS se concentre principalement sur l'organisation et les apprenants, sa principale préoccupation est de gérer les apprenants, les activités d'apprentissage et la cartographie des compétences d'organisation[18].

Généralement, on peut énumérer sept fonctions centrales d'un système de gestion de contenu :

- Acquisition et création de contenus
- Présentation et publication de contenus
- Préparation et mise à jour des contenus
- gestion et organisation des contenus
- distribution et intégration de contenus
- Traitement des contenus (workflow)
- Réutilisation de contenus

4.4 MOOC (Massive Open Online Courses)

MOOC signifie Massive Open Online Course, ce qui se traduit littéralement par cours en ligne massif et ouvert. Un MOOC est un cours en ligne accessible gratuitement ou moyennant des frais, proposé par une institution d'enseignement supérieur, une entreprise ou une organisation. Les MOOCs offrent une variété de contenus éducatifs, allant des cours de programmation et de marketing aux cours de langues et de sciences sociales [32]. Les MOOCs sont conçus pour accueillir un grand nombre d'étudiants du monde entier (des centaines de milliers d'étudiants). et proposent généralement des supports de cours interactifs tels que des vidéos, des forums de discussion, des quiz et des examens en ligne. Les MOOCs peuvent être complétés à votre rythme, et certains offrent des certificats de réussite ou des crédits universitaires. L'aspect de la massivité différencie les plates-formes MOOC à des systèmes de gestion d'apprentissage, car ces derniers ne seraient pas en mesure de gérer un grand nombre d'étudiants[33]. Les MOOCs se distinguent également des autres plates-formes en ligne par l'importance qu'ils accordent à la diffusion de contenus sous forme de vidéos. Ils sont fondés sur l'idée que l'apprentissage se déroule au sein d'un réseau, où les étudiants utilisent des plates-formes numériques telles que les blogs, les wikis et les plates-formes de médias sociaux pour établir des liens avec des contenus, des communautés d'apprentissage et d'autres apprenants afin de créer et de construire des connaissances.

Ainsi, la grande quantité de données générées par les MOOC nécessite l'utilisation de techniques d'exploration de données pour les traiter et les analyser.

Outre la masse, un autre trait qui caractérise les MOOC est l'ouverture, qui permet aux apprenants d'être libres d'apprendre, de participer et d'accéder aux cours sans contraintes, sans frais de scolarité, sans identification et sans conditions préalables ou certifications.

5 Normes et standard

5.1 Spécifications SCORM

SCORM (Sharable Content Object Reference Model) représente « comprehensive picture of how a Learning Management System (LMS) might serve up Web-based learning content to learners in a standard way »[34]. Cette définition décrit SCORM comme un moyen de fournir du contenu d'apprentissage sur différentes plates-formes. L'importance de SCORM est qu'il permet aux concepteurs pédagogiques de réutiliser le contenu d'apprentissage sur différentes plates-formes sans avoir à tout recréer[34]. En d'autres termes, il décrit une manière spécifique de diffuser du contenu d'apprentissage en ligne sur différentes plateformes. En fait, de nombreuses plateformes, commerciales ou open-source, supportent cette spécification[35]. En termes simples, SCORM est un ensemble de spécifications permettant de développer, de conditionner et de fournir des supports d'enseignement et de formation de haute qualité, à tout moment et en tout lieu. Les cours conformes à SCORM facilitent le développement de cours en garantissant que les cours conformes sont RAID[36] :

- **Réutilisabilité** : Facilité d'adaptation du contenu d'apprentissage avec d'autres outils et systèmes.
- **Accessibilité** : Disponibilité du contenu selon les besoins des différents utilisateurs du système.
- **Interopérabilité** : Possibilité de travailler avec les mêmes ressources pédagogiques sur différents systèmes et plateformes (par exemple sur différents dispositifs, différents systèmes d'exploitation, etc).
- **Durabilité** : Capacité de résistance avec les changements technologiques sans avoir besoin de faire des modifications coûteuses et profondes.

Les principaux composants de SCORM sont :

- Le CAM (Content Aggregation Model) qui définit un modèle de conditionnement des contenus d'apprentissage.
- Le RTE (RunTime Environment) qui définit une interface pour permettre la communication entre le contenu d'apprentissage et le système qui le lance (par exemple, un LMS).

Le composant Modèle d'agrégation de contenu (CAM) traite des actifs (asset), des objets de contenu partageable (SCO) et des paquets d'agrégation de contenu :

- Les actifs (asset) sont des objets individuels uniques tels que des médias ou des pages HTML, tandis que les SCO sont des collections d'actifs. Ils doivent être indépendants du contexte d'apprentissage et conçus comme de petites unités subjectives, de sorte qu'une réutilisation potentielle pour de multiples objectifs d'apprentissage soit possible.
- Les paquets d'agrégation de contenu comprennent un ou plusieurs SCO ou actifs, c'est-à-dire un ou plusieurs objets d'apprentissage. Ils doivent être structurés de telle sorte qu'ils soient prêts à être livrés à un apprenant. Le composant Environnement d'exécution comprend un adaptateur API (Application Programming Interface) et un service d'exécution.
- L'adaptateur API permet la communication entre le contenu d'apprentissage et le LMS qui le lance.
- Le service « Runtime » est fourni par le LMS et est responsable de la fourniture de l'interface utilisateur à l'apprenant.

La figure 2 montre ces principaux composants et la manière dont ils interagissent avec les plates-formes d'apprentissage en ligne telles que les LMS basés sur le Web. L'opération de base est lancée par un navigateur client qui demande un serveur LMS qui stocke les paquets SCORM. Cet accès LMS déclenche le service « Runtime » qui permet l'affichage du paquet SCORM dans le navigateur Web de l'utilisateur (par exemple, en utilisant une table des matières ou un ensemble de boutons de navigation). Une fois le service « Runtime » activé, l'utilisateur peut naviguer dans le contenu SCORM. En outre, les applications utilisateur peuvent obtenir ou déposer des informations sur le serveur LMS grâce aux fonctions d'adaptation de l'API[37].

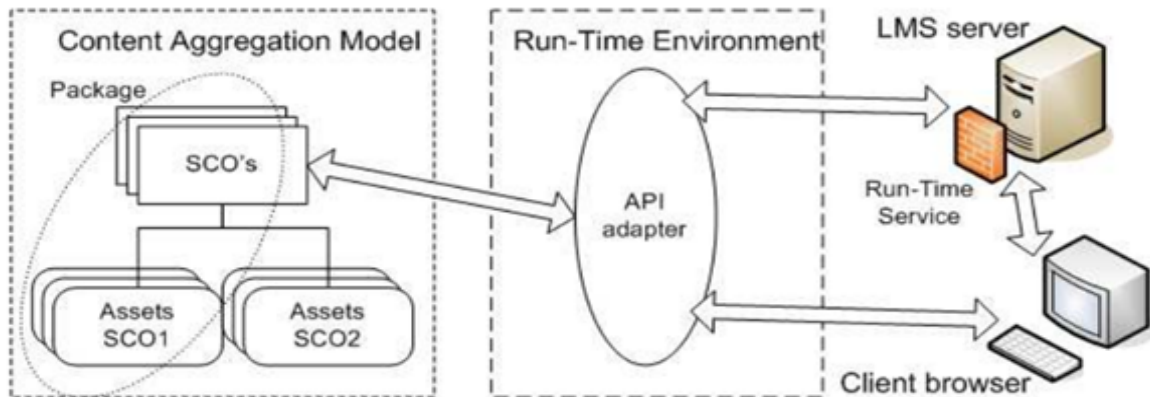


FIGURE 2 : Principes de base de SCORM [37]

5.2 Le standard LOM

Learning Object Metadata (LOM) est un standard ouvert internationalement reconnu, publié en 2002 par le « New York Institute of Electrical and Electronics Engineers Standards Association », pour décrire les "Learning Objects"[38]. Les propriétés pertinentes de l'objet d'apprentissage à décrire comprennent : le type d'objet, l'auteur, le propriétaire, les conditions de distribution, le format et les propriétés d'enseignement, telles que les méthodes d'enseignement ou d'interaction.

Le LOM définit un objet pédagogique comme « toute entité numérique ou non, utilisée dans un processus d'enseignement, de formation ou d'apprentissage »[39]. En fait, il est utilisé en général pour décrire des ressources numériques. Les métadonnées d'objet d'apprentissage sont un modèle de données, généralement encodé en XML (extensible markup language), qui décrit des objets d'apprentissage et des ressources numériques similaires utilisées pour soutenir l'apprentissage. Le but des métadonnées d'objets d'apprentissage est de soutenir la réutilisation de ces objets, de faciliter leur découverte et leur interopérabilité, généralement dans le contexte d'un système de gestion de l'apprentissage en ligne (LMS).

Les principales utilisations de LOM :

- La création de descriptions bien structurées des ressources d'apprentissage. Ces descriptions doivent faciliter la découverte, la localisation, l'évaluation et l'acquisition de ressources pédagogiques par les apprenants, les forma-

teurs ou les processus logiciels automatisés.

- Le partage des descriptions des ressources d'apprentissage entre les systèmes de découverte des ressources. Cela devrait permettre la réduction du coût de la fourniture de services basés sur des descriptions de ressources de haute qualité.
- Adaptation des descriptions de ressources pour répondre aux besoins spécifiques d'une communauté. Il peut s'agir de choisir des vocabulaires contrôlés appropriés pour la classification, de réduire le nombre d'éléments décrits ou d'en ajouter de nouveaux provenant d'autres schémas de description des ressources.
- Les créateurs et les éditeurs peuvent utiliser le LOM ainsi que d'autres spécifications pour "étiqueter" les ressources d'apprentissage avec une description qui peut être associée à la ressource. Cela fournira des informations dans un format standard similaire à celui que l'on trouve sur la couverture et la page de garde d'un manuel.

5.2.1 Structure de LOM

Le LOM est divisé en neuf catégories qui remplissent différentes fonctions. Les éléments contenus dans chaque catégorie sont présentés dans la figure 3 [40]. La sémantique d'un élément est déterminée par le contexte de la catégorie qui le contient.

1. La catégorie « General » regroupe les informations générales qui décrivent l'objet d'apprentissage dans son ensemble.
2. La catégorie « Lifecycle » regroupe les caractéristiques liées à l'histoire et à l'état actuel de cette matière d'apprentissage, ainsi que les personnes qui ont affecté cette matière d'apprentissage au cours de son évolution.
3. La catégorie « Meta-Metadata » regroupe les informations sur l'instance de métadonnées elle-même (plutôt que sur l'objet d'apprentissage que l'instance de métadonnées décrit).

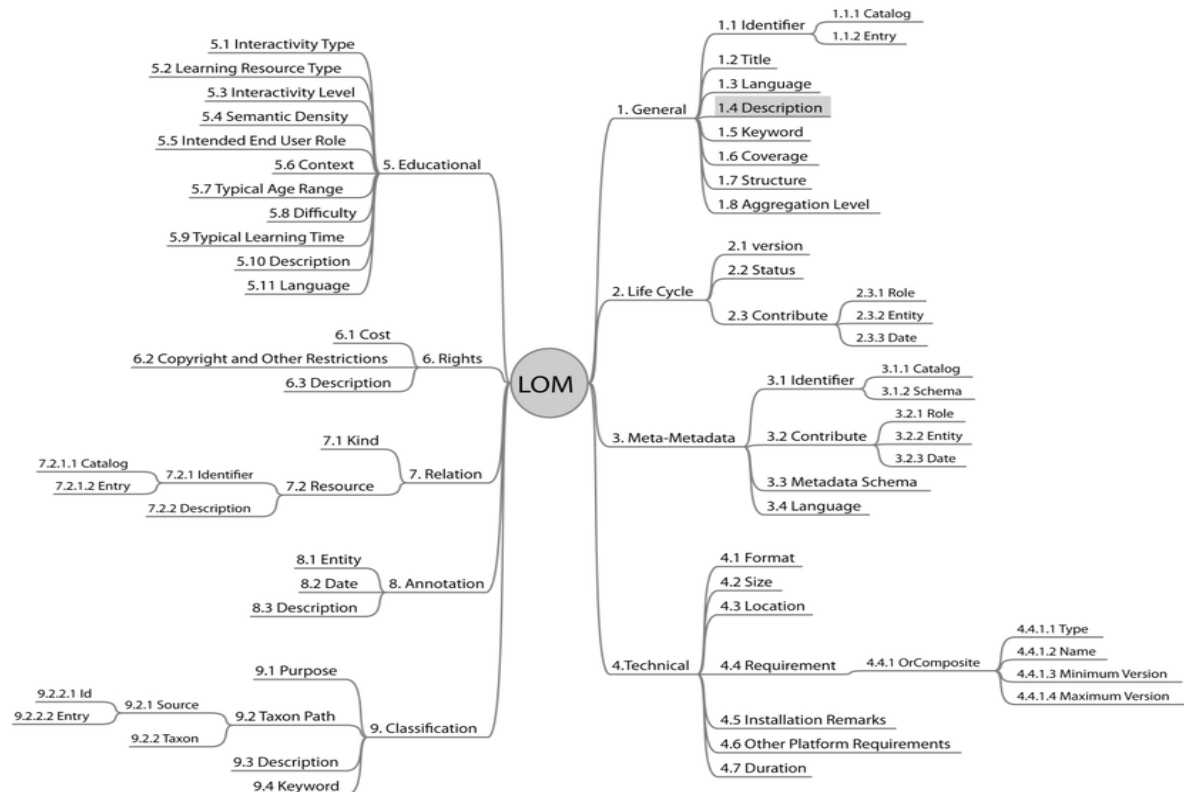


FIGURE 3 : représentation des éléments de la hiérarchie du modèle de LOM [40]

4. La catégorie « Technical » regroupe les exigences et les caractéristiques techniques de la classe d'apprentissage.
5. La catégorie « Educational » regroupe les caractéristiques éducatives et pédagogiques de la ressource d'apprentissage.
6. La catégorie « Rights » regroupe les droits de propriété intellectuelle et les conditions d'utilisation de la ressource didactique.
7. La catégorie « Relation » regroupe les caractéristiques qui définissent la relation entre l'objet d'apprentissage et d'autres objets d'apprentissage connexes.
8. La catégorie « Annotation » fournit des commentaires sur l'utilisation pédagogique de l'objet d'apprentissage et fournit des informations sur la date et l'auteur des commentaires.
9. La catégorie « Classification » décrit cet objet d'apprentissage par rapport

à un système de classification particulier.

5.3 IMS-LD

IMS (IMS) Learning Design [41] est une norme ouverte utilisée pour coder une grande variété de cours numériques (appelés "unités d'apprentissage" ou "unités d'étude") de manière formelle, sémantique, interopérable et lisible par machine. Par rapport à d'autres spécifications techniques d'apprentissage en ligne telles que SCORM (Sharable Content Object Reference Model), dans lesquelles un processus d'apprentissage est modélisé sous la forme d'une séquence de matériel d'apprentissage, LD (Learning design) prend en charge un large éventail d'approches pédagogiques modernes utilisées aujourd'hui, comme l'apprentissage actif, l'apprentissage collaboratif, l'apprentissage adaptatif et l'apprentissage basé sur les compétences. Il peut également être utilisé pour soutenir l'apprentissage informel qui a lieu dans les communautés de pratique et les communautés d'apprentissage.

La norme IMS/LD est une spécification dont l'objectif est de rendre la conception de processus d'enseignement/apprentissage formellement notable. IMS/LD intègre « Educational Modeling Language » (EML) avec les standards IMS déjà existants, notamment le Content Package Standard. L'élément central est ce que l'on appelle l'unité d'apprentissage (unit of learning), dans laquelle un paquet de contenu et un design d'apprentissage peuvent être reliés.

LD est divisé en 3 niveaux qui sont stratifiés en fonction de l'étendue et de la puissance des éléments linguistiques couverts :

- Le niveau A comprend le vocabulaire de base, qui contient l'ensemble le plus simple d'éléments linguistiques, mais qui permet une grande variété de scénarios pédagogiques. Les éléments de langage comprennent les activités du processus d'apprentissage, les rôles que les utilisateurs peuvent endosser et les services, tels que le chat ou le forum de discussion, qui peuvent être utilisés dans le processus d'apprentissage. Au niveau A, ce processus est spécifié de manière plutôt statique, les structures flexibles et dynamiques ne sont pas encore couvertes.
- Le niveau B ajoute des propriétés et des conditions qui peuvent être utilisées, entre autres, pour personnaliser et dynamiser les processus spécifiés.

- Le niveau C ajoute des notifications et offre ainsi la possibilité de varier le processus d'apprentissage en fonction des évènements.

IMS Learning Design utilise la métaphore d'une pièce de théâtre pour visualiser la manière de modéliser les unités d'apprentissage. Une pièce de théâtre est jouée par un certain nombre d'acteurs, qui peuvent jouer un certain nombre de rôles à différents moments de la pièce. De même, dans la conception de l'apprentissage, un apprenant peut jouer différents rôles à différentes étapes du processus d'apprentissage. A la fin de chaque acte, l'action s'arrête, tous les apprenants sont synchronisés, et un nouvel acte peut alors commencer [41].

La spécification IMS Learning Design a été développée pour répondre à certaines exigences spécifiques [42] :

1. Exhaustivité : La spécification doit être en mesure de décrire entièrement le processus d'enseignement dans une unité d'apprentissage, y compris les références aux objets et services numériques et non numériques nécessaires au cours du processus. Cela inclut :
 - L'intégration des activités des apprenants et des membres du personnel.
 - L'intégration des ressources (objets d'apprentissage et services de communication/collaboration) utilisées pendant l'apprentissage.
 - La prise en charge des modèles d'apprentissage à utilisateur unique ou multiple.
 - La prise en charge des modes mixtes (apprentissage mixte) ainsi que de l'apprentissage en ligne pur.
2. Expressivité pédagogique : la spécification doit être capable d'exprimer la signification et la fonctionnalité pédagogique des différents éléments de données dans le contexte d'une conception d'apprentissage. Bien qu'elle doive être suffisamment flexible pour décrire des conceptions d'apprentissage basées sur toutes sortes de pédagogies, elle doit éviter de privilégier une approche pédagogique spécifique.
3. Personnalisation : la spécification doit pouvoir décrire les aspects de personnalisation au sein d'une conception d'apprentissage, de sorte que le contenu

et les activités d'une unité d'apprentissage puissent être adaptés en fonction des préférences, du portefeuille, des connaissances préalables, des besoins éducatifs et des circonstances situationnelles des utilisateurs. En outre, elle doit permettre au concepteur, lorsqu'il le souhaite, de transmettre le contrôle du processus d'adaptation à l'apprenant, à un membre du personnel et/ou à l'ordinateur.

4. Compatibilité : la spécification doit permettre aux conceptions d'apprentissage d'utiliser et d'intégrer efficacement d'autres normes et spécifications disponibles lorsque cela est possible, telles que les spécifications IMS [43] et IEEE LTSC (Institute of Electrical and Electronics Engineers Learning Technology Standards Committee) [44].

La spécification de conception d'apprentissage d'IMS se compose de plusieurs éléments. Tout d'abord, elle consiste en un modèle conceptuel (une ontologie) pour la description des processus d'enseignement et d'apprentissage. Ce modèle est exprimé sous la forme d'un modèle UML (Unified Modeling Language) (voir la figure 4).

En substance, le modèle indique que les apprenants réalisent un ensemble d'activités d'apprentissage en utilisant des objets et des services d'apprentissage (qui se trouvent dans l'environnement de l'activité) afin d'atteindre certains objectifs d'apprentissage explicites ou implicites, qui peuvent ensuite être utilisés par d'autres dans leurs activités d'apprentissage ou de soutien (par exemple, un formateur peut fournir des commentaires sur un rapport rédigé par un apprenant).

Les enseignants, les autres membres du personnel ou les pairs peuvent réaliser des activités de soutien pour aider les apprenants en cas de besoin. La conception peut être statique ou adaptative, en tenant compte des compétences existantes, des besoins et des circonstances des personnes concernées.

5.4 Plateformes e-learning

Le domaine de l'e-learning est de plus en plus populaire, et ce pour les bonnes raisons. Au cours des dernières années, les plates-formes d'apprentissage en ligne

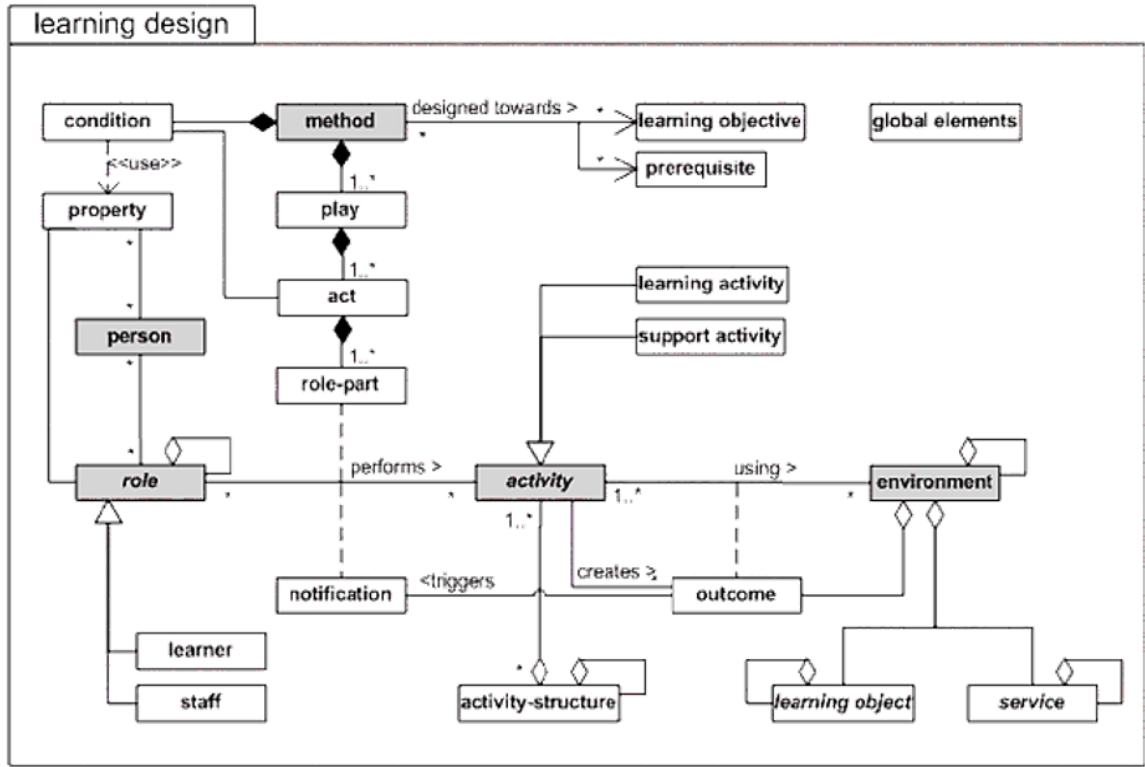


FIGURE 4 : Modèle conceptuel de la conception globale de l'apprentissage [43]

intuitives ont fait des bonds en avant en offrant une expérience d'apprentissage complète qui est amusante, interactive et finalement engageante.

Une plateforme d'apprentissage en ligne est une application logicielle électronique qui accompagne les processus d'apprentissage informatisés. Il organise le contact multimédia entre enseignants et apprenants. Une telle plateforme propose et gère diverses formations et peut mesurer et certifier leurs résultats d'achèvement.

Une plateforme d'apprentissage en ligne offre aux utilisateurs la possibilité de recevoir des commentaires immédiats qui peuvent être utilisés pour l'auto-réflexion en affichant les résultats d'apprentissage. De cette façon, les entreprises peuvent également suivre les certifications des employés et vérifier l'adéquation des compétences et des exigences du poste. Ces aspects impliquent la possibilité d'un meilleur développement individuel en termes de compétences et d'aptitudes ainsi qu'une planification optimale du déploiement du personnel.

On distingue deux grandes familles des plateformes d'apprentissage :

- Plateformes open source : dans ce type, on regroupe les plateformes ayant un code source ouvert et disponible pour contribution ou adaptation selon les besoins des utilisateurs. Par exemple : Moodle, Claroline Connect, Chamilo, etc. . .
- Plateformes propriétaires : cette famille de plateformes regroupe les plateformes n'ayant pas un code source accessible. Dans la plupart des cas, on trouve ces plateformes payantes. Exemples : Blackboard Learn, Desire2Learn, . . .
- Plateformes gratuites : ce sont des plateformes dont l'utilisation est gratuite et ouverte au public. Exemples : Acolad, Moodle, Claroline, . . .

Nous présentons ci-dessous, quelques plateformes dédiées à l'apprentissage en ligne.

5.5 Plateforme Moodle

Moodle [45] est une plateforme d'apprentissage en ligne servant à créer des communautés d'apprenants autour de contenus et d'activités pédagogiques. Le terme Moodle est l'acronyme de « Modular Object-Oriented Dynamic Learning Environment ». Dotée d'un système de gestion de contenu - (SGC) performant, Moodle a aussi des fonctions pédagogiques ou communicatives lui permettant de créer un environnement d'apprentissage en ligne. Ces fonctionnalités permettent de créer des interactions entre pédagogues, apprenants et ressources pédagogiques formant ainsi un réseau et parfois même une véritable communauté autour d'un thème choisi par les membres de la plateforme.

L'interface de Moodle se présente sous forme de différents blocs : le bloc central présente les documents et les activités du cours. Ils peuvent être classés selon plusieurs formats :

- Thématique : en fonction de thèmes ou de sujets du cours
- Hebdomadaire : en fonction d'un agenda ou du calendrier
- Informel : en fonction de sujets de discussion et de forums.

L'architecture d'une plateforme Moodle repose sur :

- Un serveur web
- Un système de gestion de bases de données : MariaDB / SQL
- Les technologies web : PHP5, HTML, CSS

5.6 Plateforme Blackboard

Blackboard [46] est une plateforme d'apprentissage exclusive développée par l'université de la Colombie-Britannique au Canada et répertoriée par Blackboard Inc depuis mai 1999. Blackboard Learn est une application d'apprentissage en ligne hautement personnalisable qui permet aux utilisateurs de suivre ou d'héberger des cours en ligne. Les étudiants et les enseignants peuvent interagir à l'aide de devoirs, de vidéoconférences, de groupes de discussion, de tests, etc. Les applications mobiles et les outils d'accessibilité de Blackboard permettent aux étudiants de participer à distance et de manière flexible aux cours.

Blackboard offre plusieurs moyens aux étudiants et aux enseignants de rester sur la bonne voie tout au long de leur cours. Les enseignants peuvent publier des annonces sur les modifications apportées à un programme, les dates d'échéance, les horaires des examens, etc. Les notifications peuvent apparaître sur la page "Accueil" du compte Blackboard d'un étudiant pour son établissement ou dans des classes spécifiques. Les enseignants et les étudiants peuvent également suivre les dates importantes grâce à l'outil calendrier, qui combine les horaires des cours, des individus et des établissements. Blackboard adhère également aux normes d'interopérabilité telles que IMS et SCORM[34].

5.7 Plateforme Chamilo

Le projet de la plateforme Chamilo a démarré en 2010 par la communauté Dokeos, qui est une extension de Claroline, et est soutenu par l'association Chamilo, qui vise l'amélioration et la promotion des logiciels utilisés dans le domaine de l'éducation [47].

Chamilo est une solution libre, basée sur le web, qui permet d'améliorer l'accès aux ressources d'apprentissage. D'un aspect pédagogique, c'est une plateforme

offrant aussi bien aux formateurs qu'aux apprenants un outil facile à prendre en main et adaptable selon leurs besoins. Chamilo peut être utilisé par des gens de différents âges, à partir de 6 et arrivant à 80 ans. Selon les statistiques affichées sur le site web officiel de la plateforme, Chamilo compte plus de 11.000.000 utilisateurs à travers le monde [47]. La facilité de l'outil ainsi que son intuitivité lui ont permis d'attirer de plus en plus d'utilisateurs.

D'un aspect technique, Chamilo peut être intégré avec plusieurs technologies, à savoir : Oracle ou IBM DB2. La norme SCORM est respectée ce qui assure une interopérabilité des contenus pédagogiques générés par Chamilo. C'est une plateforme qui traite également la problématique de sécurité en fournissant des codes sécurisés. Chamilo est une des plateformes gratuites et open-source, dont l'utilisation est simple. Cependant, elle représente quelques inconvénients comme l'absence de la messagerie interne et d'un dictaphone pour l'enregistrement d'audio en ligne.

6 M-learning : apprentissage mobile

Les technologies de l'information mobiles peuvent libérer les élèves des limites de la salle de classe traditionnelle et leur permettre d'apprendre en déplacement. Les élèves peuvent explorer l'information, collaborer avec leurs pairs, enregistrer leur apprentissage et recevoir des commentaires des enseignants à tout moment et en tout lieu [48]. Les innovations récentes dans les applications de programmes et les logiciels sociaux utilisant les technologies du Web 2.0, comme les blogs et les wikis ou les sites de réseaux sociaux, ont rendu les appareils mobiles plus dynamiques et omniprésents et promettent également un potentiel éducatif plus important [49].

Pour les apprenants d'aujourd'hui, les appareils mobiles font partie intégrante de leur vie personnelle et la plupart s'attendent à ce que les technologies mobiles fassent partie intégrante de leur apprentissage. Les possibilités offertes par les technologies mobiles doivent être évaluées en fonction de leur impact sur l'éducation, ce qui peut fournir des informations précieuses sur les implications pour l'apprentissage des étudiants du 21e siècle. Comme le posent Yu, Lee et Ewing [50], les concepteurs pédagogiques doivent adopter de nouvelles façons de facili-

ter l'apprentissage, en utilisant de multiples stratégies pédagogiques, pour donner aux étudiants l'accès à l'apprentissage ubiquitaire.

6.1 Quelques définitions

L'apprentissage mobile (également appelé Mobile Learning, M-Learning) décrit des cadres d'enseignement et d'apprentissage qui sont soutenus, étendus ou conçus de manière globale avec l'utilisation ciblée de technologies mobiles [51]. L'apprentissage mobile est défini comme un moyen d'accéder aux ressources et au contenu d'apprentissage à partir de n'importe quel appareil mobile.

L'apprentissage mobile désigne l'utilisation d'appareils mobiles ou sans fil dans le but d'apprendre en se déplaçant [49]. Les dispositifs d'apprentissage mobile comprennent les téléphones mobiles, les téléphones intelligents, les ordinateurs portables et les tablettes informatiques.

Selon Sharples [52], les dispositifs d'apprentissage mobile "permettent aux apprenants d'apprendre où qu'ils se trouvent et dans leur contexte personnel afin que l'apprentissage soit significatif." La disponibilité et la facilité d'utilisation accrues des dispositifs mobiles ont conduit à leur utilisation émergente dans l'environnement d'apprentissage mobile.

L'apprentissage mobile englobe tous les systèmes qui permettent aux apprenants d'accéder à des bases de données diffusées et de communiquer entre eux sans avoir à s'appuyer sur des réseaux électriques et de câbles de communication [53]. Les terminaux portables ne sont pas transportables, disposent de leur propre alimentation électrique et d'options de communication sans fil. Concrètement, y compris les ordinateurs portables, les tablettes et les smartphones.

6.2 Classement de l'apprentissage mobile

L'apprentissage mobile est décrit dans la littérature comme une extension ou un complément de l'apprentissage en ligne. Concrètement, l'apprentissage mobile est considéré comme une sous forme de l'apprentissage électronique (E-Learning), l'e-learning étant lui-même une sous-forme de l'enseignement à distance (Distance Learning).

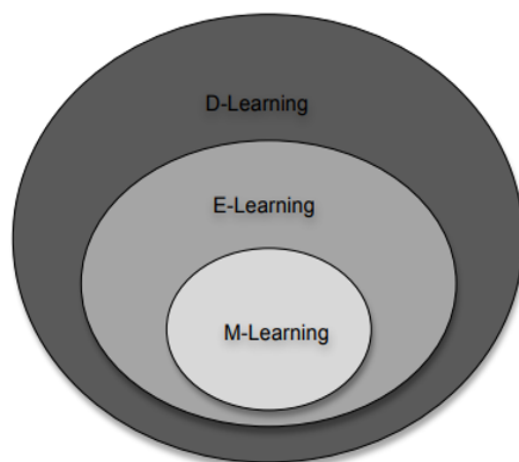


FIGURE 5 : Positionnement du m-learning

	E-learning	M-learning
Lieu	Salle de classe, Salles Internet	Indépendamment du lieu
Changement pédagogique	Plus de texte et de graphiques	Plus de voix, de graphiques, animations
Communication Enseignant<=>Etudiant	• Passif • Asynchrone • Planifié	• Immédiat • Synchrone • Spontané

	E-learning	M-learning
Communication Etudiant<=>Etudiant	• Face à face • Conférence téléphonique audio • E-mail	• Flexible • Audio et vidéo-conférence possible • Échange de messages 24h/24 et 7j/7
Tâches et tests	• En classe, à l'ordinateur • Période prédéfinie • Feedback retardé • Test standard	• Partout • À tout moment • Feedback immédiat • Test individuel

TABLE 1 : comparaison entre e-learning et m-learning

6.3 Avantages et défis du m-learning

6.3.1 Avantages du m-learning

Le m-learning, ou apprentissage mobile, présente plusieurs avantages, notamment :

- **Accessibilité accrue** : Les apprenants peuvent accéder aux cours et aux contenus pédagogiques à tout moment et n'importe où, à condition d'avoir un appareil mobile et une connexion Internet.

- Flexibilité : Les apprenants peuvent suivre des cours en fonction de leur propre emploi du temps, ce qui est particulièrement utile pour les apprenants adultes qui ont des engagements professionnels et familiaux.
- Personnalisation : Les cours et les contenus pédagogiques peuvent être adaptés en fonction des besoins individuels des apprenants, ce qui peut augmenter l'engagement et la motivation.
- Interactivité : Les technologies mobiles permettent une interactivité accrue avec les apprenants, grâce à des fonctionnalités telles que les forums de discussion, les quiz, les vidéos, les podcasts et les jeux éducatifs.
- Apprentissage en contexte : Les technologies mobiles permettent aux apprenants de se connecter avec le monde réel, en utilisant des technologies telles que la réalité augmentée et la géolocalisation, ce qui peut aider à faciliter l'apprentissage en contexte.
- Coût-efficacité : Les technologies mobiles peuvent être moins coûteuses que les approches traditionnelles d'apprentissage, car elles ne nécessitent pas d'infrastructures coûteuses telles que des salles de classe ou des laboratoires informatiques.

Ces avantages contribuent à faire du m-learning une option attrayante pour les apprenants et les éducateurs, en particulier dans un contexte où la demande de formation en ligne flexible et accessible est en constante augmentation.

6.3.2 Défis et limites du m-learning

Bien que l'apprentissage mobile présente de nombreux avantages, il existe également des inconvénients à prendre en compte, notamment :

- Limitations des appareils mobiles : Les appareils mobiles peuvent être limités en termes de taille d'écran, de capacité de stockage et de puissance de traitement, ce qui peut limiter leur capacité à prendre en charge certains types de contenus pédagogiques.
- Dépendance à la connectivité : Pour accéder aux contenus pédagogiques, les apprenants doivent avoir une connexion Internet fiable et stable, ce qui peut

être un défi dans certains contextes, notamment dans les régions rurales ou les pays en développement.

- **Distraction** : Les appareils mobiles peuvent être une source de distraction pour les apprenants, car ils peuvent être utilisés pour d'autres activités, telles que les réseaux sociaux ou les jeux.
- **Manque de contact humain** : L'apprentissage mobile peut limiter les interactions humaines entre les apprenants et les enseignants, ainsi qu'entre les apprenants eux-mêmes, ce qui peut avoir un impact sur l'engagement et la motivation.
- **Besoins en matière de compétences techniques** : Les apprenants et les enseignants doivent être à l'aise avec les technologies mobiles pour pouvoir les utiliser efficacement, ce qui peut nécessiter des compétences techniques supplémentaires.
- **Problèmes de sécurité** : L'utilisation d'appareils mobiles peut poser des problèmes de sécurité, tels que la vulnérabilité aux virus informatiques et aux cyberattaques. Ces inconvénients doivent être pris en compte lors de la planification et de la mise en œuvre de l'apprentissage mobile afin de maximiser ses avantages et de minimiser ses limitations.

Avantages	Inconvénients
Possibilité d'apprendre indépendamment du temps et du lieu	Les unités d'apprentissage sont souvent soumises à des interruptions
Les appareils mobiles sont très répandus	Un contenu complexe ne peut pas être transmis en peu de temps
Favorise l'apprentissage en situation	Investissement financier et en temps
Les unités d'apprentissage sont gérables, car elles se déroulent en petites unités	Manque d'interaction

Avantages	Inconvénients
Permet la communication et la recherche d'informations via Internet	Difficulté d'évaluation de l'apprentissage
Apprentissage dans un environnement	personnel Les problèmes de sécurité
Utilisation des moyens de communication	La dépendance technologique
Contenu multimédia	La nécessité de mises à jour régulières

TABLE 2 : Avantages et inconvénients de l'apprentissage mobile

6.4 Caractéristiques de l'apprentissage mobile

Pour utiliser l'apprentissage mobile dans l'enseignement, il est nécessaire de clarifier certaines caractéristiques de l'apprentissage mobile. Prieto et al. [54] énumère les principales caractéristiques :

- **Immédiateté** L'immédiateté dans le contexte de l'apprentissage mobile signifie généralement que les exercices de contenus peuvent être effectués directement sur l'appareil. En outre, il est également possible de publier des contenus via le terminal.
- **Multimédia** Alors qu'il n'est pas possible d'utiliser des contenus multimédias sans moyens techniques ou terminaux et qu'il faut donc recourir à des images statiques ou à des fiches. En revanche, les appareils mobiles permettent d'intégrer des éléments multimédias dans l'enseignement. Par contenu multimédia, on entend l'audio, la vidéo, les animations, les graphiques, etc.
- **Personnalisation** L'apprentissage mobile favorise un apprentissage efficace. Les appareils mobiles peuvent en effet être adaptés individuellement à leurs propres besoins et à l'environnement dans lequel l'appareil est utilisé, grâce à de multiples réglages.

6.5 Principes de l'apprentissage mobile

Sur la base des caractéristiques et propriétés de l'apprentissage mobile présentées ci-dessus, les principes didactiques suivants doivent être pris en compte conformément à la liste de Sanchez et. al [54] :

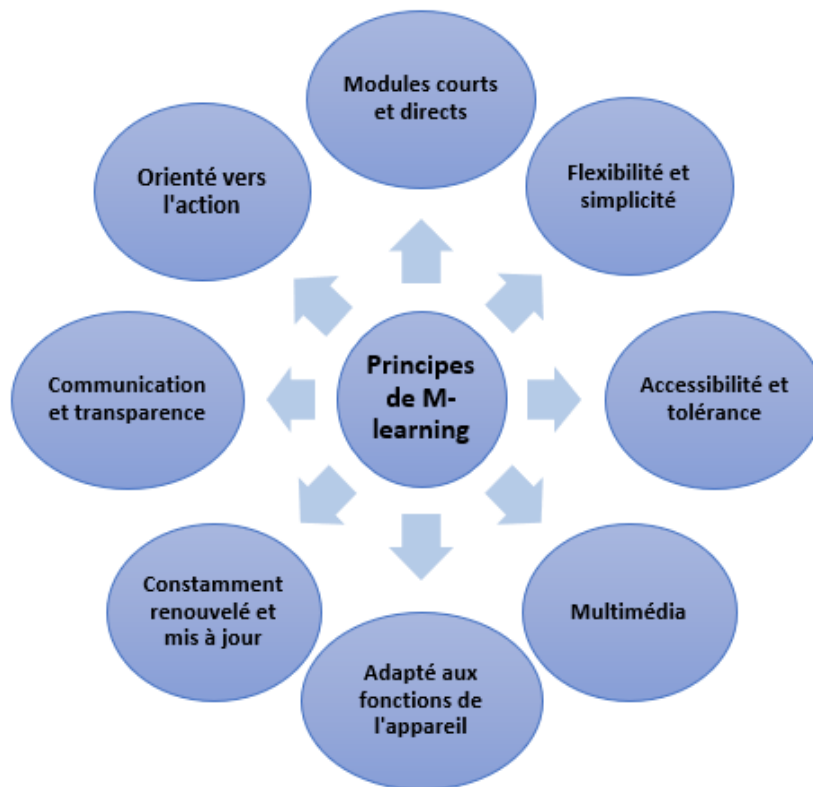


FIGURE 6 : les principes de M-learning[54]

7 Conclusion

De nos jours, nous pouvons constater que divers outils ont été développés pour améliorer les procédures d'enseignement et d'apprentissage. L'e-learning est devenu une tendance grâce à l'intégration de nouvelles technologies d'information et de communication dans l'enseignement. Ce changement a pour conséquence que de plus en plus de personnes possèdent au moins un appareil portable, qu'il

s'agisse d'une tablette, d'un smartphone, d'un ordinateur portable ou d'autres appareils mobiles tels qu'un lecteur de livre électronique. Si l'on suit la tendance des smartphones et des tablettes depuis 2010, on constate une forte augmentation de ces appareils mobiles dans le quotidien des gens. En raison de la multifonctionnalité des appareils mobiles, de nouveaux domaines et méthodes de transmission des connaissances apparaissent. Les systèmes d'apprentissage sont des outils qui permettent aux utilisateurs de disposer de canaux de communication et d'espaces pour gérer et stocker des contenus éducatifs. Ces derniers doivent se conformer aux normes pour assurer une meilleure interopérabilité entre les différents environnements techniques.

Dans ce chapitre, nous avons présenté une revue de la littérature sur l'utilisation des plateformes d'apprentissage en ligne dans les établissements scolaires. Nous avons examiné l'évolution des environnements d'apprentissage humain, de leur émergence à nos jours, en décrivons les grandes périodes de leur histoire. Nous avons ensuite défini ces environnements sous diverses formes, telles que l'apprentissage à distance (D-learning), l'apprentissage en ligne (E-learning), l'apprentissage mobile (M-learning) et les MOOC. Nous avons également décrit les avantages, les inconvénients et les différents systèmes et applications qui offrent différentes capacités pour chaque type d'apprentissage.

Dans le prochain chapitre, nous présentons le concept Big Data, ses caractéristiques et les différentes technologies qui y sont associées. Nous examinons également les systèmes de recommandations, en détaillant leurs approches et leurs critères d'évaluation. En outre, nous analysons l'évolution de ces technologies ainsi que leur application à l'e-learning, en mettant en lumière leur impact sur les plateformes d'apprentissage.

Chapitre II

Big data et systèmes de recommandation au service de l'apprentissage

Sommaire

1	Introduction	43
2	Big Data	45
3	Les systèmes de recommandation	67
4	Conclusion	93

1 Introduction

Les dernières années ont connu une révolution dans le domaine de formation grâce à l'apparition de l'apprentissage à distance (E-learning). Ce type d'apprentissage a permis aux apprenants et aux formateurs d'avoir accès à l'information sans avoir besoin d'une présence physique, grâce à des systèmes conçus et développés pour cet objectif. Ces systèmes sont largement utilisés et sont devenus une nouvelle tendance, ils génèrent un grand volume de données éducatives capturées à partir des activités des étudiants pendant le processus d'apprentissage en ligne. La quantité de données disponibles est tellement énorme que les techniques de traitement traditionnelles ne peuvent pas être utilisées pour les traiter. Par conséquent, des techniques d'analyse sophistiquées sont nécessaires pour gérer et analyser la grande quantité de données afin d'améliorer les expériences

d'apprentissage en ligne des apprenants.

Le traitement de la grande quantité de données existantes dans le domaine de l'éducation a été rendu possible grâce au développement des nouvelles technologies d'information et de communication (TIC). Ce développement a conduit diverses institutions d'enseignement à analyser les données existantes issues des interactions de leurs apprenants et à tirer ainsi des conclusions qui amélioreront l'environnement de travail, généreront de nouvelles structures d'organisations éducatives et de nouvelles procédures d'apprentissage. Le Big Data offre la possibilité de traiter efficacement les données d'apprentissage volumineuses afin d'ajouter la valeur aux plates-formes d'apprentissage en ligne.

L'un des domaines les plus intéressants de l'apprentissage en ligne est l'application des systèmes de recommandation en utilisant les Big Data. Les systèmes de recommandation sont un type d'intelligence artificielle conçu pour analyser les données sur le comportement des utilisateurs et faire des recommandations personnalisées aux apprenants. Ces systèmes utilisent l'exploration de données, l'apprentissage automatique et d'autres techniques avancées pour identifier des modèles et des relations dans les données éducatives, et utilisent ces informations pour suggérer du contenu, des cours ou d'autres ressources susceptibles d'intéresser les apprenants. Cela peut améliorer considérablement l'expérience d'apprentissage en ligne, en aidant les apprenants à trouver les ressources dont ils ont besoin plus rapidement et plus efficacement.

Les systèmes de recommandation ont le potentiel d'améliorer considérablement l'impact de l'apprentissage en ligne, en aidant les apprenants à tirer la meilleure partie du processus d'apprentissage et en rendant l'expérience éducative plus personnalisée et plus attrayante. En analysant les grandes quantités de données générées par les activités d'apprentissage en ligne, les systèmes de recommandation peuvent fournir aux apprenants des recommandations personnalisées basées sur leurs besoins et préférences d'apprentissage uniques. Cela peut contribuer à améliorer l'efficacité du processus d'apprentissage, à accroître la satisfaction des étudiants et à promouvoir l'adoption de l'enseignement à distance à plus grande échelle.

Dans ce chapitre, nous présentons le concept Big Data, ses caractéristiques

et ses différentes technologies. Nous abordons également les systèmes de recommandations, leurs approches, les critères d'évaluations et leur utilisation dans les systèmes e-learning. Nous expliquons ensuite l'impact de ces technologies sur l'apprentissage, et leur rôle pour améliorer les méthodes d'enseignement et d'apprentissage.

2 Big Data

La révolution technologique et la numérisation de la grande majorité des processus ont conduit à une augmentation exponentielle des données et des informations. Néanmoins, les techniques classiques de traitement et de stockage ne peuvent pas gérer ce volume énorme de données. Par conséquent, une bonne analyse et interprétation de ces données massives peuvent faciliter la prise de décision et la mise en œuvre de propositions susceptibles pour une meilleure solution. C'est dans ce but que le Big Data a vu le jour [55]. Il s'agit d'un ensemble de technologies et outils qui permettent de gérer, d'analyser, d'interpréter et d'évaluer de grandes quantités de données et d'informations générées dans différents contextes [56]. La valeur du Big Data réside dans la possibilité d'analyser une plus grande quantité d'informations d'une manière efficace, afin de découvrir de nouvelles connaissances qui doivent être prises en compte dans la prise de décision. Le Big Data peut-être appliqué à de nombreux domaines, tels que la médecine, l'ingénierie et bien d'autres spécialités. Dans notre recherche on se focalise sur l'application du Big Data dans le domaine de la formation à distance vue l'importance de ce dernier surtout avec l'essor qu'il a connu après la crise de COVID 19.

2.1 Définition

Le terme "Big Data" désigne l'évolution et l'utilisation des technologies qui fournissent les bonnes informations à partir d'une masse de données en croissance exponentielle. Le défi consiste non seulement à traiter des volumes de données en augmentation rapide, mais aussi à gérer des formats de plus en plus hétérogènes ainsi que des données complexes et interconnectées [57]. Le Big Data se présente comme une solution destinée à fournir à tous les intervenants dans la formation

à distance un accès en temps réel à des bases de données géantes.

Il s'agit d'un domaine émergent, Le Big Data est un concept très difficile à définir précisément, car la notion même de "big" en terme de volume de données varie d'un domaine à autre. En effet, la quantité totale de données créées, capturées, copiées et consommées dans le monde augmente rapidement, atteindront 97 zettaoctets en 2022 (figure 7). La croissance a été plus élevée que prévu, en raison de l'augmentation de la demande due à la pandémie de COVID-19, du fait que la plupart des personnes travaillaient et apprenaient à domicile et utilisaient plus souvent les options de divertissement à domicile.

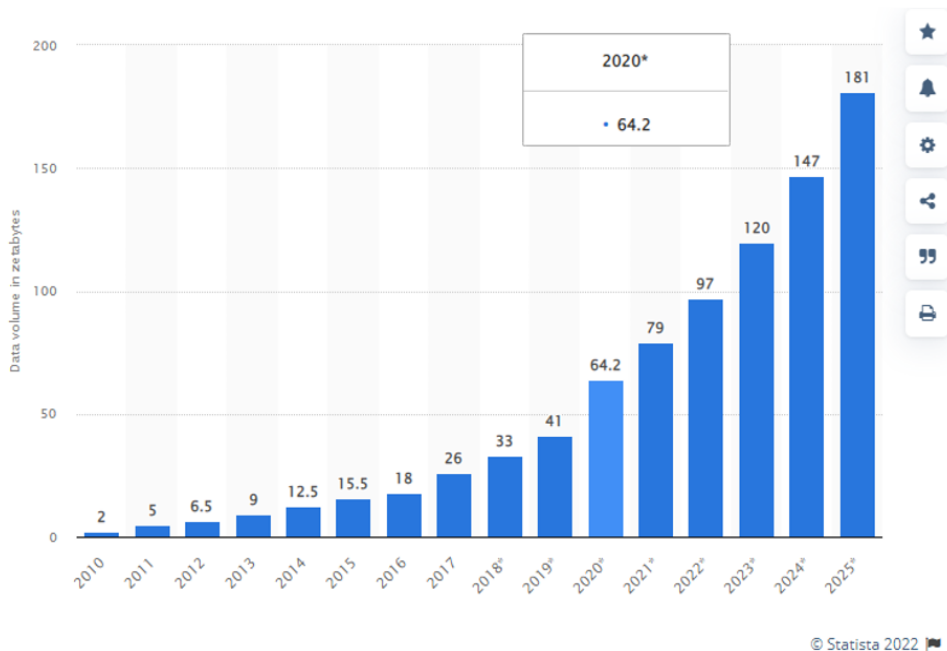


FIGURE 7 : le volume des données massive d'après Statista 2022 [58]

La définition telle que vue par Banica Logica [59], est la suivante : «Les Big Data sont une collection massive de données partageables provenant de tout type de sources numériques privées ou publiques, ce qui représente à lui seul une source de découverte, d'analyse, de veille stratégique et de veille économique».

Le terme "Big Data" désigne généralement des données qui dépassent les capacités de stockage, de traitement et de calcul habituelles des bases de données et des techniques d'analyse de données classiques. En tant que ressource, le Big Data nécessite des outils et des méthodes qui peuvent être appliqués pour analyser et

extraire des modèles à partir de données à grande échelle [60]. Selon l'Institut national des normes et de la technologie, le Big Data fait référence à la nécessité de paralléliser le traitement des données dans les applications à forte intensité de données.

L'analyse des données volumineuses désigne le processus de collecte, d'organisation et d'analyse de grands ensembles de données afin de découvrir différents modèles et d'autres informations utiles. L'analytique des Big Data est un ensemble de technologies et de techniques qui nécessitent de nouvelles formes d'intégration pour révéler de grandes valeurs cachées à partir de grands ensembles de données qui sont différents des habituels, plus complexes et d'une échelle énorme. Elle vise principalement à résoudre de nouveaux problèmes ou d'anciens problèmes de manière plus efficace [61].

2.2 Brève histoire

Le concept des big data est relativement récent, mais les ensembles de données volumineux remontent aux années 1960 et 1970, lorsque le monde des données débutait avec les premiers centres de données et le développement de la base de données relationnelle [62].

Autour de 2005, les gens ont commencé à réaliser à quel point les utilisateurs généraient de données à travers Facebook, YouTube et d'autres services en ligne. Dans la même année, Hadoop, un Framework open-source créé spécifiquement pour stocker et analyser les grands ensembles de données, a été développé [63]. NoSQL a également commencé à gagner en popularité à cette époque. Le développement de cadres open-source tels que Hadoop (et plus récemment Spark) a été essentiel pour la croissance des big data [63], car ils facilitent le travail avec les big data et les rendent moins chers à stocker. Depuis lors, le volume de big data a explosé. Les utilisateurs continuent de générer d'énormes quantités de données, mais ce ne sont pas seulement les humains qui le font. Avec l'avènement de l'Internet des objets (IoT), de plus en plus d'objets et de dispositifs sont connectés à internet, collectant des données sur les habitudes d'utilisation des clients et les performances des produits [64]. L'émergence de l'apprentissage automatique a produit encore plus de données. Bien que les big data aient beaucoup progressé, leur utilité ne fait que commencer. Le cloud computing a encore plus élargi les

possibilités des big data. Le cloud offre une élasticité véritablement scalable, où les développeurs peuvent simplement créer des grappes ad hoc pour tester un sous-ensemble de données [65].

2.3 Caractéristiques de Big Data

Les caractéristiques fondamentales qui déterminent le Big Data sont décrit par le modèle 5V [66], chaque V représente un élément clé : Volume, Vitesse, Variété, véracité et la valeur.

2.3.1 Volume

Représente la quantité volumineuse des données disponibles pour le stockage, le traitement, et l'analyse par des outils et technologies Big Data afin d'extraire des informations précieuses. En effet, les entreprises sont inondées de données de tous types qui ne cessent de croître, accumulant facilement des téraoctets, voire des pétaoctets d'informations.

2.3.2 Vitesse

Fait référence au temps dans lequel les Big Data peuvent être traitées. Certaines activités sont très importantes et nécessitent des réponses immédiates, c'est pourquoi un traitement rapide maximise l'efficacité. Pour les processus sensibles au facteur temps, tels que la détection des fraudes, les flux de Big Data doivent être analysés et utilisés dès leur arrivée dans les organisations afin de maximiser la valeur de l'information.

2.3.3 Variété

Représente le type de données que le Big Data peut comprendre. Le Big Data consiste en tout type de données, y compris des données structurées et non structurées telles que du texte, des données de capteurs, des audio, des vidéos, des flux de clics, des fichiers journaux, etc.

2.3.4 Véracité

Désigne le degré auquel un leader se fie aux informations pour prendre une décision. Par conséquent, trouver les bonnes corrélations dans le Big Data est

très important pour l'avenir des entreprises.

2.3.5 Value

C'est Le dernier V des 5 V du big data, il définit la valeur ajoutée que les données collectées peuvent apporter au processus. Il ne s'agit pas seulement de la quantité de données que nous stockons ou traitons. Il s'agit en fait de la quantité de données précieuses, fiables et dignes de confiance qui doivent être stockées, traitées et analysées pour trouver des informations. La capacité à tirer la valeur des big data est une exigence, car la valeur des big data augmente considérablement en fonction des informations que l'on peut en tirer.

2.4 Technologies pour l'analyse Big Data

Avec la progression de la technologie informatique, les entreprises peuvent désormais traiter de grandes quantités de données qui auraient auparavant nécessité l'utilisation d'ordinateurs coûteux. La gestion du Big Data est devenue cruciale pour des entreprises telles que Yahoo!, Google et Facebook qui se sont rendus compte qu'ils avaient besoin d'outils pour monétiser les quantités énormes de données générées par leurs services. Pour faire face à ces défis, ces entreprises sont à la recherche de technologies innovantes qui leur permettent de stocker, accéder et analyser ces données en temps réel. Cette analyse en temps réel est nécessaire pour maximiser les avantages des données provenant des utilisateurs. Les solutions qui en découlent ont eu un impact sur le marché de la gestion des données en général. Les technologies telles que MapReduce, Hadoop et Big Table ont montré qu'elles permettent une nouvelle génération de gestion des données. Elles aideront les entreprises à résoudre l'un des défis les plus fondamentaux, celui de traiter de grandes quantités de données de manière efficace, rentable et rapide. Pour traiter et analyser efficacement ces données, plusieurs technologies sont utilisées :

2.4.1 Hadoop

Une des solutions pour gérer les Big Data est le framework logiciel Hadoop [67]. C'est un projet open source développé par Apache Foundation. Hadoop fournit un environnement informatique fiable, évolutif et distribué. Hadoop peut

stocker une taille de données pratiquement illimitée et peut traiter ces données dans des environnements distribués. Il s'agit d'une solution libre, à source ouverte et relativement simple à mettre à l'échelle.

La plate-forme Hadoop contient les modules suivants (voir figure 8) :

- Hadoop Common - Ce sont des bibliothèques et des utilitaires Java requis par d'autres modules Hadoop. Ces bibliothèques fournissent des abstractions au niveau du système de fichiers et du système d'exploitation et contiennent les fichiers Java et les scripts nécessaires au démarrage de Hadoop ;
- HDFS - Système de fichiers distribué Hadoop : Système de fichiers distribué qui fournit un accès à haut débit aux données des applications
- YARN - un cadre pour la planification des travaux et la gestion des ressources de cluster ;
- MapReduce - un système basé sur YARN pour le traitement parallèle de grands ensembles de données.

Dans un environnement Hadoop distribué, chaque partie des données est stockée en plusieurs copies (généralement au moins trois) sur différents ordinateurs (c'est à dire des nœuds de cluster). Une des hypothèses les plus importantes est que les données sont traitées localement, Cela signifie que les données sont traitées là où elles sont stockées (sur le même ordinateur / nœud), ce qui minimise les transferts sur le réseau.

De plus, Hadoop ne s'appuie pas sur le matériel pour assurer la tolérance aux pannes et la haute disponibilité, mais la bibliothèque Hadoop elle-même a été conçue pour détecter et gérer les défaillances au niveau de la couche application. Ainsi, les serveurs peuvent être ajoutés ou supprimés du cluster de manière dynamique et Hadoop continue à fonctionner sans interruption, de telles situations sont gérées par le système lui-même, de sorte que l'utilisateur n'a pas besoin d'entreprendre d'autres actions.

Hadoop s'appuie sur deux serveurs [68] :

- JobTracker : il n'y a qu'un seul JobTracker par cluster Hadoop. Il reçoit les tâches Map/Reduce à exécuter et organise leur exécution sur le cluster. Ce

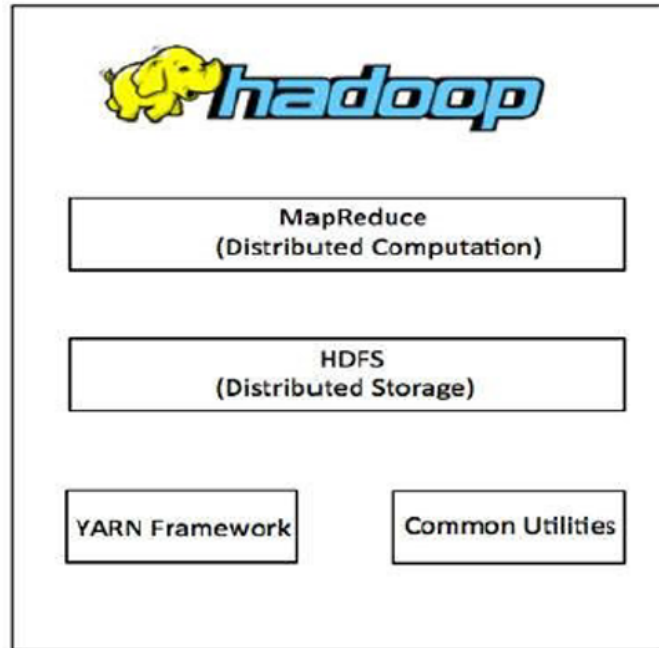


FIGURE 8 : Architecture du framework Hadoop [67]

plan d'exécution comprend la détermination des nœuds qui contiennent les données à exploiter, la disposition des nœuds pour qu'ils correspondent aux données, surveiller les tâches en cours d'exécution, et relancer les tâches si elles échouent.

- TaskTracker : plusieurs par cluster. Exécute le travail Map/Reduce lui-même (en tant que tâche Map et Reduce avec les données d'entrée associées).

Le serveur JobTracker est en communication avec HDFS ; il sait où se trouvent les données d'entrée du programme Map/Reduce et où doivent être stockées les données de sortie. Il peut ainsi optimiser la distribution des tâches en fonction des données associées.

Nous donnons maintenant une brève description des principaux éléments de la plate-forme Hadoop : HDFS et MapReduce.

2.4.2 HDFS – Hadoop Distributed File System

Le HDFS est un système de fichiers tolérant aux pannes, capable de stocker d'énormes quantités d'informations, et de survivre aux pannes de stockage sans perdre de données. Comme le montre la figure 9, il utilise une architecture NameNode et DataNode pour mettre en œuvre un système de fichiers distribué qui fournit un accès performant aux données sur des clusters Hadoop hautement évolutifs. Les clusters Hadoop sont construits avec des ordinateurs peu coûteux. Si un ordinateur (ou nœud) tombe en panne, le cluster peut continuer à fonctionner sans perdre de données ni interrompre le travail en redistribuant simplement le travail aux autres machines du cluster. HDFS gère le stockage sur le cluster en divisant les fichiers en petits blocs et en stockant des copies dupliquées de ceux-ci sur l'ensemble des nœuds. La figure 9 illustre comment un ensemble de données est généralement stocké sur un cluster de cinq nœuds. Dans cet exemple, l'ensemble des données sera toujours disponible, même si deux des serveurs sont tombés en panne.

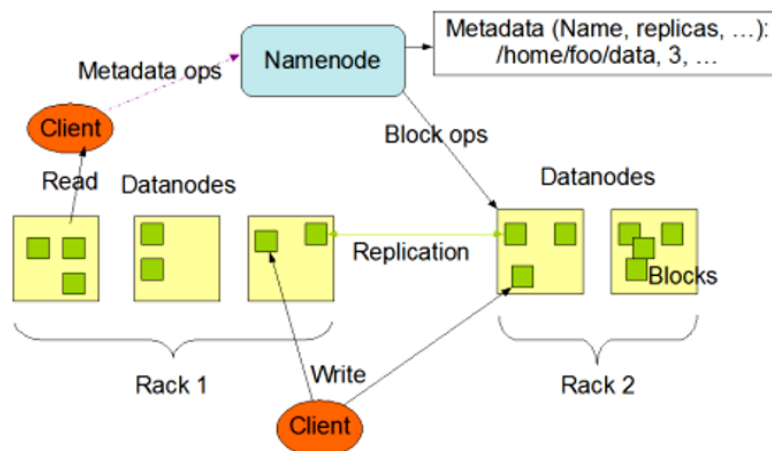


FIGURE 9 : Schéma de traitement dans l'architecture HDFS [67]

HDFS s'appuie sur deux serveurs :

- **NameNode** : unique sur le cluster. Il stocke les informations sur les noms de fichiers et leurs caractéristiques. C'est le maître du HDFS qui contrôle les **DataNode** esclaves.

- Secondary NameNode : le NameNode secondaire surveille l'état du cluster HDFS et prend des "instantanés" des données contenues dans le NameNode. Si le NameNode tombe en panne, le NameNode secondaire peut être utilisé à la place du NameNode.
- DataNode : multiple par cluster. Stocke le contenu des fichiers eux-mêmes, fragmenté en blocs (64KB par défaut).

2.4.3 MapReduce

MapReduce [69] est un modèle de programmation pour les applications qui traitent des grandes quantités de données structurées et non structurées en parallèle sur des centaines ou des milliers de serveurs standard d'un cluster Hadoop, de manière fiable et tolérante aux pannes. MapReduce fait en réalité référence à deux tâches distinctes accomplies par le programme Hadoop :

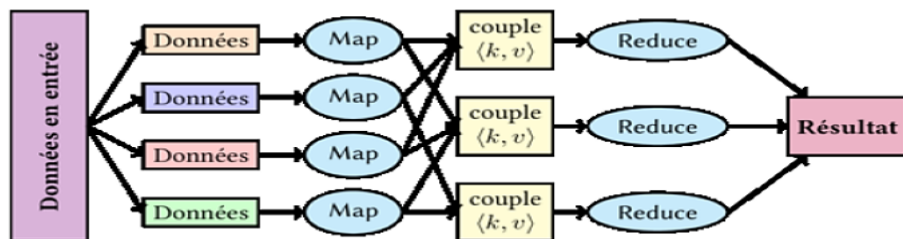


FIGURE 10 : Modèle MapReduce [70]

- La tâche MAP, il prend des ensembles de données et les convertit en d'autres sous-ensembles de données dans lesquels des éléments individuels sont décomposés en tuples (paires clé/valeur).
- La tâche REDUCE prend la sortie de MAP en tant qu'entrée et combine ces paires de valeurs en un ensemble plus petit de tuples. Comme le nom MapReduce indique, la tâche de réduction est effectuée après le travail de MAP (voir figure 10).

2.4.4 Apache spark

Apache Spark [71] est un framework de traitement unifié et un ensemble de bibliothèques pour le traitement parallèle des données sur des grappes d'ordinateurs. Il s'agit du moteur open-source le plus activement développé pour cette tâche, ce qui en fait un outil standard pour tout développeur ou spécialiste en sciences des données "data scientist" intéressé par le big data. Spark prend en charge plusieurs langages de programmation largement utilisés (Python, Java, Scala et R), comprend des bibliothèques pour diverses tâches allant du SQL au streaming et à l'apprentissage automatique, et Spark fonctionne partout, d'un ordinateur portable à un cluster de milliers de serveurs. Il s'agit donc d'un système facile à utiliser au départ et à mettre à l'échelle pour le traitement des données volumineuses ou à une échelle incroyablement grande.

- Composants d'apache spark

Apache Spark se compose de Spark Core Engine, Spark SQL, Spark Streaming, MLlib, GraphX et Spark R [72]. on peut utiliser Spark Core Engine avec l'un des cinq autres composants mentionnés ci-dessus. Il n'est pas nécessaire d'utiliser tous les composants Spark ensemble. Selon le cas d'utilisation et l'application, un ou plusieurs de ces composants peuvent être utilisés avec Spark Core.

- Spark SQL : agit comme une bibliothèque au-dessus de Spark Core. Il fournit un support pour les données structurées. Les développeurs Spark peuvent tirer parti de la puissance des requêtes déclaratives et de l'optimisation du stockage en exécutant des requêtes de type SQL sur des données Spark, présentes dans des RDD et d'autres sources externes. Il permet d'interroger les données via SQL (Structured Query Language) ainsi que la variante Apache Hive de SQL appelée HQL (Hive Query Language). Il prend en charge les connexions JDBC et ODBC qui établissent une relation entre les objets Java et les bases de données existantes, les entrepôts de données et les outils de veille stratégique. Il prend également en charge diverses sources de données comme les tables Hive, Parquet et JSON. Spark SQL facilite le processus d'extraction et de fusion de divers ensembles de données afin que ces derniers soient prêts à être utilisés pour l'apprentissage automatique.

- Spark Streaming : est une API qui permet aux développeurs d'effectuer facilement le traitement par lots et le streaming de données, dans la même application. Il utilise un flux continu de données d'entrée (RDD : Resilient Distributed Dataset) pour traiter les données en temps réel. Spark Streaming tire parti de la capacité d'ordonnancement rapide d'Apache Spark Core pour effectuer des analyses en continu en ingérant des données par mini-batches. Les transformations sont appliquées à ces mini-lots de données. Les données dans Spark Streaming sont ingérées à partir de diverses sources de données et de flux en direct comme Twitter, Apache Kafka, Akka Actors, IoT Sensors, Amazon Kinesis, Apache Flume, etc. dans des applications événementielles, tolérantes aux pannes et sûres.
- MLib : signifie Machine Learning Library. C'est une bibliothèque d'apprentissage automatique qui contient divers algorithmes d'apprentissage automatique, et peut être appelée depuis les langages de programmation comme Scala, Python, R et Java. Il s'agit notamment de corrélations et de tests d'hypothèses, de classification et de régression, de regroupement et d'analyse en composantes principales. MLib est simple à utiliser, évolutive, compatible avec divers langages de programmation et peut être facilement intégrée à d'autres outils.
- GraphX : pour les graphes et le traitement parallèle des graphes, Apache Spark fournit une autre API appelée GraphX. Le terme "graphe" ne désigne pas ici des graphiques, des lignes ou des diagrammes à barres, mais des graphes en informatique, comme les réseaux sociaux, qui sont constitués de sommets où chaque sommet représente un utilisateur individuel du réseau social où de nombreux utilisateurs sont reliés les uns aux autres par des arêtes. Ces arêtes représentent la relation entre les utilisateurs du réseau. Pour manipuler les graphes, GraphX prend en charge divers opérateurs fondamentaux comme le sous-graphe, la jointure des sommets et l'agrégation des messages.
- SparkR : le langage de programmation R est largement utilisé par les "Data scientists" en raison de sa simplicité et de sa capacité à exécuter des algorithmes complexes. Mais R souffre d'un problème : sa capacité de traitement des données est limitée à un seul nœud. Cela rend R inutilisable lors

du traitement d'une énorme quantité de données. Le problème est résolu par SparkR qui est un paquetage R dans Apache Spark. SparkR fournit une mise en œuvre de cadre de données qui prend en charge des opérations telles que la sélection, le filtrage, l'agrégation, etc. sur de grands ensembles de données distribuées. SparkR prend également en charge l'apprentissage automatique distribué à l'aide de Spark MLlib.

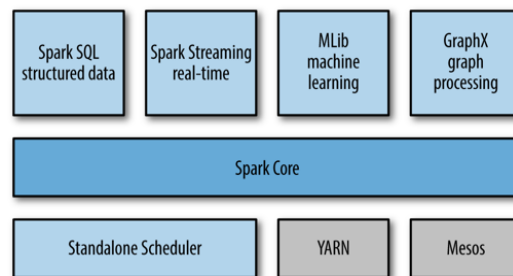


FIGURE 11 : Composants d'apache Spark[72]

Les composants ci-dessus font d'Apache Spark le meilleur moteur de traitement des Big data. Tous ces composants sont fournis tels quels, et nous pouvons les utiliser séparément ou ensemble. Spark nous aide également à décomposer les tâches intensives et à forte intensité de calcul en tâches plus petites et plus concises qui sont ensuite exécutées par les nœuds de travail. Il permet également de traiter des données en temps réel ou archivées grâce à son architecture de base.

2.4.5 Apache YARN

Apache Hadoop YARN est la technologie de gestion des ressources et de planification des tâches du cadre de traitement distribué open source Hadoop. L'un des principaux composants d'Apache Hadoop, YARN est responsable de l'allocation des ressources système aux diverses applications exécutées dans un cluster Hadoop et de la planification des tâches à exécuter sur les différents nœuds du cluster [73]. YARN est l'abréviation de Yet Another Resource Negotiator (négociateur de ressources supplémentaires). La technologie est devenue un sous-projet d'Apache Hadoop au sein de l'Apache Software Foundation (ASF) en 2012 et a été l'une des principales fonctionnalités ajoutées dans Hadoop 2.0, qui a été publié

pour être testé cette année-là et est devenu généralement disponible en octobre 2013.

Apache Hadoop YARN décentralise l'exécution et le suivi des tâches de traitement en séparant les différentes responsabilités dans ces composants :

- Un ResourceManager global qui accepte les soumissions de tâches des utilisateurs, planifie les tâches et leur alloue des ressources.
- Un esclave NodeManager installé sur chaque nœud et qui fonctionne comme un agent de surveillance et de rapport du ResourceManager.
- Un ApplicationMaster qui est créé pour chaque application afin de négocier les ressources et de travailler avec le NodeManager pour exécuter et surveiller les tâches.
- Des conteneurs de ressources qui sont contrôlés par les NodeManagers et auxquels sont attribuées les ressources système allouées aux applications individuelles.

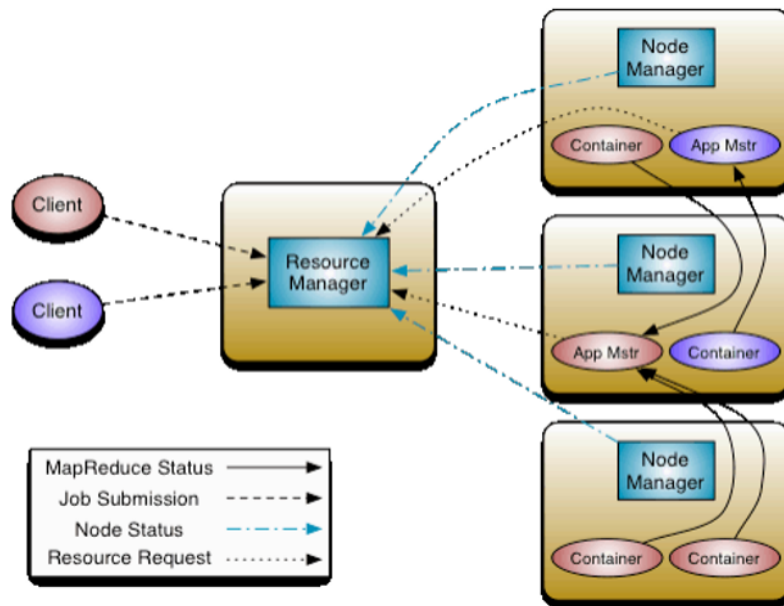


FIGURE 12 : composants d'apache Hadoop YARN [73]

2.4.6 Apache Flume

Apache Flume est un système distribué, fiable et disponible permettant de collecter, d'agréger et de déplacer efficacement de grandes quantités de données de journaux provenant de nombreuses sources différentes vers un magasin de données centralisé. Apache Flume est un projet de haut niveau de Apache Software Foundation [74].

L'utilisation d'Apache Flume ne se limite pas à l'agrégation de données de journaux. Les sources de données étant personnalisables, Flume peut être utilisé pour transporter des quantités massives de données d'événements, y compris les données de trafic réseau, les données générées par les médias sociaux, les messages électroniques et à peu près toutes les sources de données possibles. En général, l'architecture d'Apache Flume est composée des éléments suivants, comme le montre l'architecture suivante :

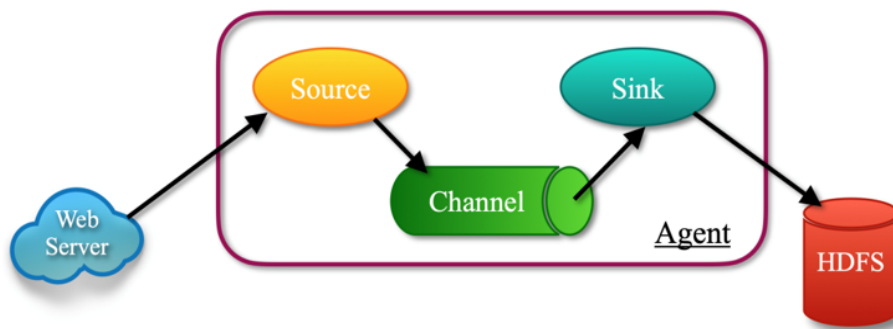


FIGURE 13 : architecture d'apache Flume [74]

1. Source Flume : Une source Flume est présente sur les générateurs de données comme Facebook ou Twitter. La source collecte les données du générateur et les transfère au canal Flume sous la forme d'événements Flume. Flume supporte différents types de sources.
2. Flume Channel : est appelé canal du flux, c'est un système de stockage intermédiaire qui met en mémoire tampon les événements envoyés par la source du flux jusqu'à ce qu'ils soient consommés par le Sink. Le canal agit comme un pont intermédiaire entre la source et le Sink. Les canaux Flume sont transactionnels par nature. Flume fournit un support pour le

canal Fichier et le canal Mémoire. Le canal File est durable par nature, ce qui signifie qu'une fois que les données sont écrites dans le canal, elles ne seront pas perdues, même si l'agent redémarre. Dans le canal Mémoire, les événements du canal sont stockés en mémoire, il n'est donc pas durable mais très rapide par nature.

3. Flume Sink : un Flume Sink est présent sur les dépôts de données comme HDFS, HBase. Flume sink consomme les événements du canal et les stocke dans les dépôts de destination comme HDFS. Il n'existe pas de règle selon laquelle le Sink doit livrer les événements au système de stockage, mais nous pouvons le configurer de manière à ce qu'un Sink puisse livrer des événements à un autre agent. Flume supporte différents Sink comme le Sink HDFS, le Sink Hive, le Sink Thrift, le Sink Avro.
4. Flume agent : un agent Flume est un processus Java à long terme qui fonctionne sur la combinaison Source - Channel - sink. Flume peut avoir plus d'un agent. Nous pouvons considérer Flume comme une collection d'agents Flume connectés qui sont distribués par nature.
5. Flume event : un événement est l'unité de données transportée dans Flume. La représentation générale de l'objet de données dans Flume est appelée événement. L'événement est composé d'une charge utile d'un tableau d'octets avec des en-têtes optionnels.

2.4.7 Apache Scoop

Apache Sqoop [72] est un outil conçu pour transférer efficacement des données en masse entre Apache Hadoop et des bases de données structurées telles que des bases de données relationnelles. Les entreprises qui utilisent Hadoop trouvent qu'il est nécessaire de transférer certaines de leurs données des systèmes traditionnels de gestion de bases de données relationnelles (SGBDR) vers l'écosystème Hadoop. Sqoop, qui fait partie intégrante d'Hadoop, peut effectuer ce transfert de manière automatisée. De plus, les données importées dans Hadoop peuvent être transformées avec MapReduce avant d'être réexportées vers le SGBDR. Sqoop peut également générer des classes Java pour interagir de manière programmatique avec les données importées. Sqoop utilise une architecture basée sur des

connecteurs qui lui permet d'utiliser des plugins pour se connecter à des bases de données externes.

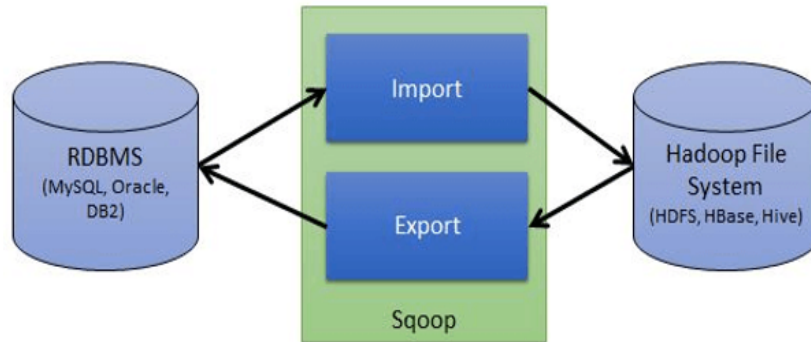


FIGURE 14 : Architecture de Apache Sqoop [72]

2.5 Travaux connexes

La littérature dans le domaine d'application des Big Data dans l'enseignement contient quelques travaux de recherche très récents. Cette revue de la littérature est axée sur l'utilisation de l'analyse de Big Data dans l'enseignement supérieur. Le tableau ci-dessous représente les articles relatifs au domaine d'application des Big Data dans l'enseignement par auteur, titre, objectif, approche et année publiée :

Auteur/Année	Titre	Objectif	Approche
Liang-qiu Meng 2014 [75]	“Application of Big Data in Higher Education”	Définir les caractéristiques du Big Data et ses applications dans l’enseignement supérieur	Présentation de quatre applications de Big Data dans l’enseignement supérieur : • Prévission • Fournir une nouvelle plateforme pour l’éducation • Développer un nouveau mode d’enseignement
Ben Daniel et Russell Butson 2014 [76]	“Foundations of Big Data and Analytics in Higher Education”	Décrire le rôle que jouent les Big Data, les opportunités et les défis associés à la mise en œuvre de l’analyse Big Data dans l’enseignement supérieur.	Framework conceptuel qui fournit une base pour comprendre les composants fondamentaux du Big Data dans l’enseignement supérieur.
Tristan Denley, 2014 [77]	“How Predictive Analytics and Choice Architecture Can Improve Student Success”	Explorer les défis auxquels les étudiants sont confrontés dans la structure du programme d’études postsecondaires, et comment un système d’analyse prédictive et une architecture de choix peuvent jouer un rôle.	Une structure d’évaluation solide pour évaluer l’impact du Degree Compass, un système de recommandation de cours qui associe avec succès les étudiants actuels aux cours qui correspondent le mieux à leurs talents et à leur programme d’études pour les prochains semestres.

Auteur/Année	Titre	Objectif	Approche
Katrina Sin et Loganathan Muthu 2015 [78]	“Application of Big Data in education data mining and learning analytics – a literature review”	Examiner des applications récentes des technologies Big Data dans l'éducation et présenter une revue de la littérature disponible sur l'exploration de données pédagogiques et l'analyse de l'apprentissage.	Étude systématique de la littérature publiée sur le domaine d'application du « Data mining » et « Learning analytics » en éducation et en apprentissage.
Shakeel Ahmed et Hemant Kumar Mehta 2015 Saudi Arab , [79]	“On Applying Big Data and Cloud Computing for Quality Improvement in Higher Education”	Présenter une approche basée sur le Cloud et le Big Data pour améliorer la qualité de l'enseignement supérieur.	Framework basé sur le Cloud et les Big Data, Ce cadre fonctionne comme un système de recommandation et suggère les modifications à apporter au plan de cours et au programme.
Banica Logica, Radulescu Magdalena 2015 [59]	“Using Big Data in the Academic Environment”	Traiter les aspects relatifs à l'évolution des technologies Big Data, à la manière de les appliquer à la formation en ligne et à leur influence sur l'environnement universitaire.	Une architecture système en trois étapes pour un consortium d'universités, reposant sur des solutions logicielles réelles, ayant pour objectif d'analyser, d'organiser et d'accéder à d'énormes ensembles de données dans l'environnement en nuage.

Auteur/Année	Titre	Objectif	Approche
Ahmed Ashraf, Hazem El-Bakry, Samir M. Abd El-razek, Yehia El-Mashad 2015 [80]	“Handling Big Data in E-Learning”	Présenter un nouveau modèle d'apprentissage en ligne afin d'améliorer les performances du traitement des Big Data dans les systèmes éducatifs.	Un modèle qui peut déterminer le comportement de tous les apprenants dans le système d'apprentissage en ligne en traitant des données volumineuses éducatives à l'aide de hadoop et de NoSQL.
Ben Daniel 2015 [81]	“Big Data and analytics in higher education : Opportunities and challenges”	Identifier les défis contemporains auxquels sont confrontés les établissements d'enseignement supérieur du monde entier et explore le potentiel du Big Data pour relever ces défis.	Description d'un certain nombre d'opportunités et de défis associés à la mise en œuvre du Big Data dans le contexte de l'enseignement supérieur, ainsi que les orientations futures relatives à l'élaboration et à la mise en œuvre d'un projet institutionnel sur le Big Data.
Zijiao Tang 2017 [82]	“On Study of Application of Big Data and Cloud Computing Technology in Smart Campus”	Améliorer le processus de création d'informations dans les collèges et universités.	Construction d'un campus intelligent en combinant la technologie Big Data et la technologie Cloud Computing pour créer un système de base de données et un système de gestion de campus.

Auteur/Année	Titre	Objectif	Approche
Marouane Birjali, Abderrahim Beni-Hssane, Mohammed Erritali 2018 [83]	“A novel adaptive e-learning model based on Big Data by using competence-based knowledge and social learner activities”	Décrire comment la technologie Big Data transforme le paradigme de l'apprentissage en ligne	Un modèle d'apprentissage en ligne adaptatif pour fournir le contenu d'apprentissage le mieux adapté à chaque apprenant. Les résultats expérimentaux montrent que l'environnement Big Data convergent beaucoup mieux que ceux des travaux traditionnels.
Yi Li, et Xiaoning Zhai 2018 [84]	“Review and Prospect of Modern Education using Big Data”	Fournir la revue et la perspective de l'éducation moderne à l'aide du Big Data implémentée dans le lycée affilié à l'Université Renmin de Chine	Organigramme des systèmes d'analyse d'apprentissage, les données et l'expérimenté seraient une information utile pour effectuer des recherches précises affiliés à l'Université Renmin de Chine.
Mohammad Shorfuzzaman et al. 2018 [85]	“Harnessing the power of Big Data analytics in the Cloud to support learning analytics in mobile learning environment”	Adopter l'analyse de Big Data pour un traitement efficace des données d'apprentissage volumineuses afin d'ajouter de la valeur aux environnements d'apprentissage mobiles.	Framework d'apprentissage mobile basé sur le Cloud qui utilise la technique d'analyse de Big Data pour extraire les valeurs d'un volume énorme de données d'apprenants mobiles.
Aleksandra Nikolovska, Aleksandar et al. 2018 [86]	“Framework for Big Data Analytics of Moodle Data Using Hadoop in the Cloud”	Analyser la base de données de la plateforme d'apprentissage en ligne Moodle dans le but d'améliorer le processus éducatif	Elaboration d'un bon cadre pour l'analyse des données massives collectées auprès du LMS (système de gestion de l'apprentissage) Moodle.

Auteur/Année	Titre	Objectif	Approche
--------------	-------	----------	----------

TABLE 3 : Travaux connexes

2.6 Intégration du big data dans les systèmes d'apprentissage en lignes

L'adoption des technologies Big Data dans les systèmes d'apprentissage en ligne améliorera considérablement l'ensemble du processus de collecte, de stockage, de gestion et d'analyse des grands ensembles de données [87]. La figure 15 illustre la conception du Framework big data pour les systèmes d'apprentissage en ligne.

Le Framework est une solution complète pour gérer les données volumineuses dans l'écosystème big data [69]. Il prend en charge toutes les étapes de la gestion, du traitement et de l'analyse des données structurées et non structurées. Il est conçu pour aider les utilisateurs à comprendre les différentes composantes de l'écosystème big data et les outils utilisés à chaque étape.

Le Framework est constitué de quatre modules principaux : le module de stockage, le module de traitement, le module d'analyse et le module de visualisation. Chacun de ces modules est conçu pour gérer une partie spécifique de l'écosystème big data.

1. Acquisition et intégration des données : dans cette phase, les données sont collectées et capturées à partir des systèmes d'apprentissage en ligne (LMS ou CMS...etc.) qui contiennent des informations sur les apprenants, des données d'évaluation et de test, etc. Ces données jouent un rôle très important dans l'identification des styles d'apprentissage et des connaissances de l'apprenant afin d'adapter l'apprentissage au profil de chaque utilisateur. Cette étape comprend également la préparation, l'extraction et l'intégration des données à l'aide des outils ETL à partir de sources multiples (textes, vidéos, images, audios, multimédia et tutoriels...).
2. Gestion des données : pour faciliter le processus de chargement des données, nous utilisons apache Flume pour stocker, agréger et transférer les données du journal LMS depuis le serveur de base de données. Pour importer et

exporter tous les types de données entre les bases de données relationnelles et Hadoop HDFS, nous utilisons apache Scoop.

3. Analyse des données et apprentissage automatique : il s'agit d'une étape cruciale dans le framework, dans laquelle on trouve tous les algorithmes d'apprentissage automatique mettant en œuvre des modèles d'analyse e-learning, tels que l'apprentissage supervisé et non supervisé, les règles d'association, la logique floue, la régression, le clustering, la classification, le filtrage collaboratif, TF-IDF...etc. Dans cette étape, nous mettons en œuvre le traitement en streaming et en temps réel des données de logs d'e-learning en utilisant Hadoop et Spark.
4. Visualisation et tableaux de bord : dans cette phase, les résultats analytiques des données d'apprentissage sont visualisés efficacement à l'aide d'outils et de logiciels avancés. Ces nouvelles connaissances seront adoptées par les éducateurs pour la prise de décision, ces décisions comprennent : la détection des comportements, l'estimation des compétences des étudiants, la détection des risques d'attrition, l'analyse prédictive, le système d'enseignement intelligent et la recommandation de cours.

En résumé, cette section présente et explore l'intégration des technologies big data dans les systèmes d'apprentissage en ligne. Le big data améliorera considérablement le processus d'apprentissage en fournissant un large ensemble d'outils et de systèmes pour le traitement parallèle de stockage distribué des données d'apprentissage massives produites par les apprenants. Le framework servira de modèle de référence pour les professionnels de l'apprentissage en ligne afin de faciliter la mise en œuvre d'une nouvelle plate-forme d'apprentissage basée sur le big data, il offre des outils avancés et des API précieuses pour l'analyse des données éducatives massives, afin d'améliorer la qualité de l'apprentissage et le processus de prise de décision, et aussi de découvrir de nouvelles stratégies d'apprentissage à distance.

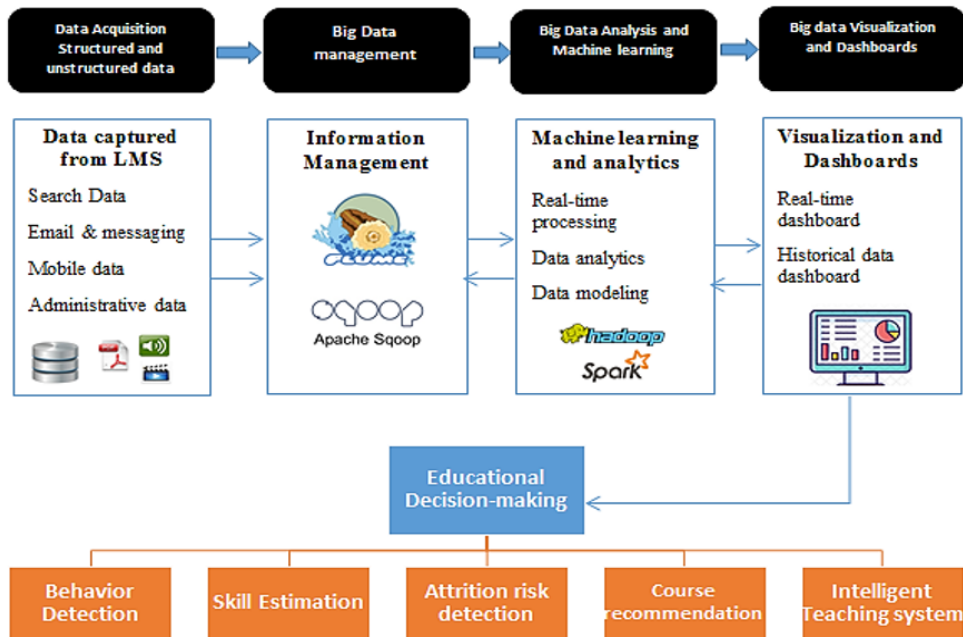


FIGURE 15 : Framework Big Data pour les systèmes d'apprentissage en ligne [69]

3 Les systèmes de recommandation

3.1 Définition

Les systèmes de recommandation ont été définis à l'origine comme des systèmes dans lesquels " les personnes fournissent des recommandations en entrée, que le système agrège ensuite et dirige vers les destinataires appropriés" [88]. Robin Burke a décrit un système de recommandation comme « tout système qui produit des recommandations individualisées en sortie ou qui a pour effet de guider l'utilisateur de manière personnalisée vers des objets intéressants ou utiles dans un large espace d'options possibles » [89]. Les systèmes de recommandation peuvent utiliser différentes sources de données, comme les préférences des utilisateurs, les interactions entre utilisateurs et les objets, les caractéristiques des objets, et d'autres contextes. Ces systèmes peuvent être basés sur des données explicites (comme les notes ou les évaluations) ou implicites (comme les données de navigation) et utilisent des techniques telles que le filtrage collaboratif, la factorisation de matrices et l'apprentissage automatique pour comprendre les pré-

férences des utilisateurs et les utiliser pour effectuer des recommandations. Les systèmes de recommandation sont utilisés dans nombreux domaines différents, tels que les médias en ligne, les systèmes de commerce électronique, les réseaux sociaux, les systèmes de recherche, et les systèmes de recommandation de soins de santé.

3.2 Brève histoire

Les premières approches de systèmes de recommandation ont vu le jour en 1992, Belkin et Croft [90] ont analysé et comparé le filtrage de l'information et la recherche d'information, la recherche d'information étant la technologie fondamentale du moteur de recherche, et le système de recommandation étant principalement basé sur la technologie du filtrage de l'information. La même année, Goldberg et al. [91] ont proposé le système Tapestry, qui est le premier système de filtrage de l'information basé sur le filtrage collaboratif par évaluation humaine. Inspirés par cette étude, des chercheurs du Massachusetts Institute of Technology (MIT) et de l'Université du Minnesota (UMN) ont développé un service de recommandation des articles de presse personnalisés, appelé GroupLens [92], dont le composant clé est un modèle de filtrage collaboratif utilisateur-utilisateur. Le professeur John Riedl a fondé un laboratoire de recherche à l'UMN, également nommé GroupLens, qui est le pionnier des études sur les systèmes de recommandation. Pour la musique et la vidéo, des technologies de recommandation similaires ont été appliquées par le système Ringo [93] et Video Recommender [94].

En 1997, le laboratoire de recherche GroupLens a lancé le projet MovieLens [95] et a entraîné la première version du modèle de recommandation avec le jeu de données EachMovie. Par la suite, plusieurs jeux de données MovieLens ont été publiés en continu de 1998 à 2019, et sont devenus l'un des jeux de données les plus populaires pour les études de recommandation.

Auteur	Approche	Pincipe
[Goldberg et al., 1992]	Tapestry	Tapestry repose sur l'utilisation de la méthode de filtrage collaboratif pour rassembler les évaluations d'utilisateurs pour un document donné et les utiliser pour recommander des documents pertinents à d'autres utilisateurs.
[Resnick et al., 1994]	GroupLens	GroupLens est un système de recommandation développé par des chercheurs du MIT et de l'Université du Minnesota. Il utilise un modèle de filtrage collaboratif utilisateur-utilisateur pour fournir des recommandations de films aux utilisateurs.
[Shardanand et Maes, 1995]	Ringo	Ringo est un système de recommandation de musique qui a été développé pour aider les utilisateurs à découvrir de nouveaux morceaux de musique en utilisant des techniques de filtrage collaboratif.
[Herlocker, 1999]	MovieLens	MovieLens est un système de recommandation de films similaire à GroupLens, qui utilise également un modèle de filtrage collaboratif utilisateur-utilisateur pour fournir des recommandations personnalisées.

TABLE 4 : Les premiers systèmes de recommandation

3.3 Les domaines d'application

Les systèmes de recommandation sont utilisés dans de nombreux domaines pour aider les utilisateurs à naviguer et à trouver des informations pertinentes. Voici quelques exemples de domaines d'application des systèmes de recommandation, avec des exemples de systèmes couramment utilisés [96] :

1. E-commerce : Les systèmes de recommandation peuvent être utilisés pour suggérer des produits pertinents aux clients en fonction de leur historique d'achat, de leur comportement de navigation sur le site et des préférences exprimées. Amazon et Netflix sont des exemples courants de systèmes de recommandation d'e-commerce.
2. Médias en ligne : Les systèmes de recommandation peuvent être utilisés pour suggérer des articles de presse, des vidéos ou de la musique en fonction des préférences des utilisateurs et de leur historique de consommation de contenu. YouTube et Spotify sont des exemples courants de systèmes de recommandation de médias en ligne.
3. Réseaux sociaux : Les systèmes de recommandation peuvent être utilisés pour suggérer des amis ou des pages à suivre sur les réseaux sociaux en fonction des intérêts et des relations existantes des utilisateurs. Facebook est un exemple courant de système de recommandation de réseau social.
4. Jeux vidéo : Les systèmes de recommandation peuvent être utilisés pour suggérer des jeux en fonction des préférences et des antécédents de jeu des utilisateurs. Steam est un exemple courant de système de recommandation de jeux vidéo.

3.4 Classification des systèmes de recommandation

Il existe plusieurs méthodes pour classer les systèmes de recommandation, mais l'une des classifications les plus courantes est basée sur la façon dont ils recueillent et utilisent les données pour générer des recommandations [97]. Figure 16 décrit les trois principales classifications du système de recommandation, à savoir : le système basé sur le contenu, le système basé sur le filtrage collaboratif et le système hybride.

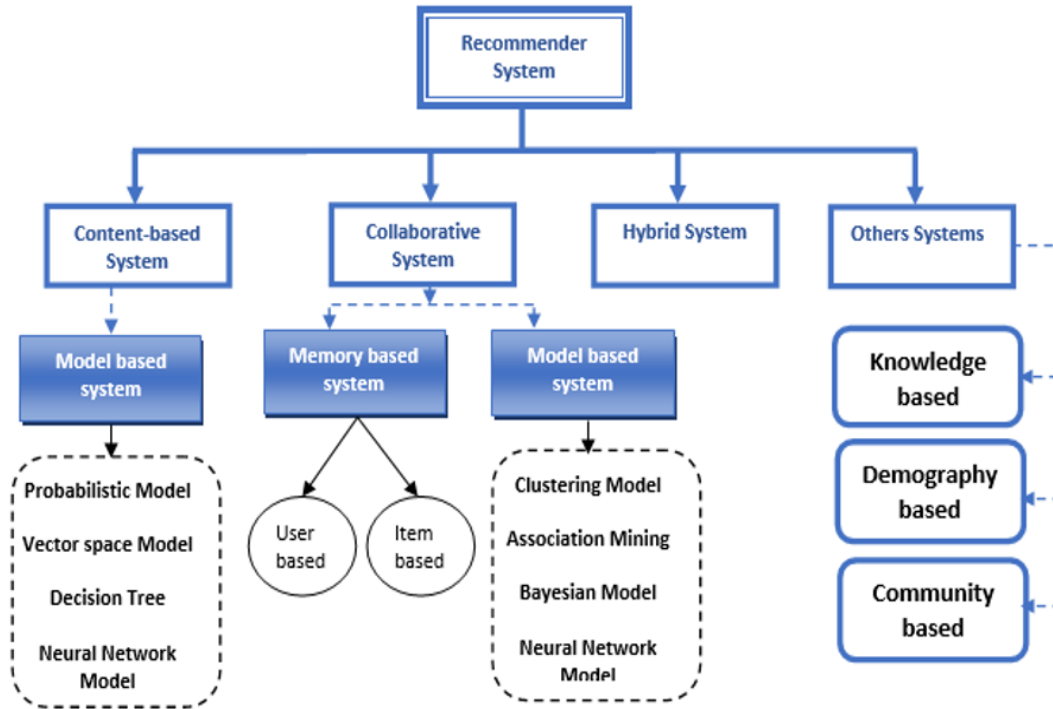


FIGURE 16 : Classification des systèmes de recommandation [97]

3.4.1 Filtrage basé sur le contenu

Le filtrage basé sur le contenu est un mécanisme de recommandation qui repose sur l'analyse des caractéristiques d'un article afin d'identifier les articles similaires et les recommander aux utilisateurs. Cette technique utilise des algorithmes spécifiques pour examiner le contenu de ces articles, tels que le texte, les images, les commentaires et les métadonnées associées. Elle peut également utiliser des informations sur les utilisateurs, telles que leur historique de navigation ou les articles qu'ils ont aimés ou achetés [98].

Le processus de recommandation est divisé en trois étapes distinctes, chacune étant gérée par un composant dédié : l'analyseur de contenu, l'apprentissage de profil et le filtrage [100].

- L'analyseur de contenu est responsable de l'extraction des caractéristiques pertinentes des items à partir de leur contenu. Il peut utiliser des techniques d'analyse de texte pour extraire des informations telles que les mots-clés, les thèmes, les genres, etc. Ces caractéristiques sont ensuite utilisées pour

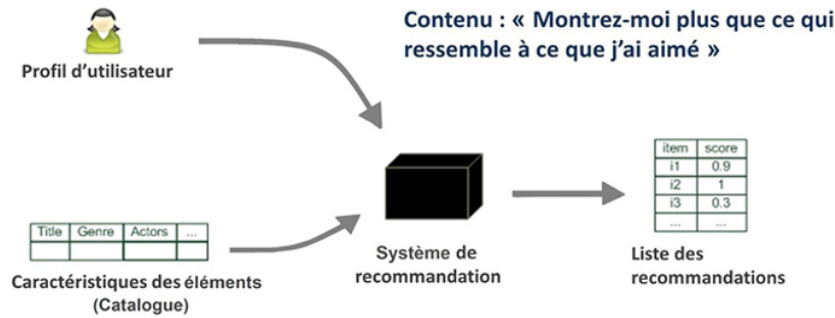


FIGURE 17 : Filtrage basé sur le contenu [99]

construire les profils d'items qui seront utilisés pour effectuer les recommandations.

- L'apprentissage de profil consiste à créer des profils utilisateurs en utilisant les caractéristiques extraites par l'analyseur de contenu. Il utilise les évaluations précédemment données par l'utilisateur, comme les items qu'il a aimés ou consultés, pour identifier les préférences de l'utilisateur et établir des correspondances entre les items et les utilisateurs.
- Le composant de filtrage est responsable de recommander les items les plus pertinents pour chaque utilisateur en utilisant les profils utilisateurs et les profils d'items. Il peut utiliser différentes méthodes pour effectuer les recommandations, comme la similarité cosinus, la distance euclidienne, etc. Les recommandations sont ensuite présentées à l'utilisateur pour consultation.

Il existe plusieurs techniques et algorithmes utilisés dans le filtrage basé sur le contenu, chacun ayant ses propres avantages et inconvénients. Voici quelques-uns des algorithmes les plus couramment utilisés :

- Algorithmes de similarité de contenu : ces algorithmes examinent les caractéristiques des articles pour déterminer leur similarité. Les algorithmes les plus couramment utilisés incluent la similarité cosinus, la similarité de Jaccard, et la similarité de Pearson [101]. Ces algorithmes peuvent utiliser des techniques telles que la vectorisation de termes et la réduction de dimension pour améliorer les résultats.

- Algorithmes de clustering : ces algorithmes regroupent des articles similaires en clusters ou groupes. Les algorithmes les plus couramment utilisés incluent l'algorithme de k-means et l'algorithme de mélange de gaussiennes[102].
- Algorithmes de réseaux de neurones : ces algorithmes utilisent des réseaux de neurones pour prédire les articles qui seront pertinents pour un utilisateur donné. Les réseaux de neurones peuvent être entraînés pour prédire les articles similaires à ceux que l'utilisateur a aimés ou achetés auparavant.

Exemples de systèmes de filtrage basés sur le contenu

Voici quelques exemples réels de systèmes de recommandation basés sur le contenu :

- Amazon : Amazon utilise des systèmes de recommandation pour suggérer des produits similaires ou complémentaires aux articles que les utilisateurs ont consultés ou achetés. Les systèmes de recommandation d'Amazon utilisent des algorithmes de similarité de contenu et de filtrage collaboratif pour identifier les articles pertinents [103].
- Netflix : Netflix utilise des systèmes de recommandation pour suggérer des films et des émissions de télévision qui pourraient intéresser les utilisateurs. Les systèmes de recommandation de Netflix utilisent des algorithmes de similarité de contenu et de filtrage collaboratif pour identifier les contenus pertinents.[104]
- YouTube : YouTube utilise des systèmes de recommandation pour suggérer des vidéos similaires ou complémentaires aux vidéos que les utilisateurs ont consultées. Les systèmes de recommandation de YouTube utilisent des algorithmes de similarité de contenu et de filtrage collaboratif pour identifier les vidéos pertinents.[105]
- Spotify : Spotify utilise des systèmes de recommandation pour suggérer des chansons similaires ou complémentaires aux chansons que les utilisateurs écoutent. Les systèmes de recommandation de Spotify utilisent des algorithmes de similarité de contenu et de filtrage collaboratif pour identifier les chansons pertinentes. [106]

- TikTok : TikTok utilise des systèmes de recommandation pour suggérer des vidéos similaires ou complémentaires aux vidéos que les utilisateurs ont consultées.

Avantages et inconvénients

Les systèmes de recommandation basés sur le contenu ont des avantages et des inconvénients. Les principaux avantages de ces approches sont [107] :

- Autonomie de l'utilisateur : les techniques de recommandation basées sur le contenu traitent les utilisateurs de manière indépendante. Seules les évaluations de chaque utilisateur sont prises en compte pour construire son profil utilisateur et faire la recommandation, ce qui n'est pas le cas pour les approches utilisant le filtrage collaboratif.
- Prise en compte immédiate d'un nouvel item : les systèmes basés sur le contenu peuvent recommander des items sans avoir besoin d'une évaluation préalable de la part d'un utilisateur. Cela permet de recommander des items récents ou peu connus.
- Flexibilité dans la sélection des items : les systèmes basés sur le contenu peuvent recommander des items en fonction de n'importe quelle caractéristique du contenu, comme le genre, les acteurs, les mots-clés, etc. Cela permet de personnaliser les recommandations pour chaque utilisateur en fonction de ses intérêts.
- Pas de besoin de données d'évaluation de la part d'autres utilisateurs : contrairement aux systèmes basés sur le filtrage collaboratif, les systèmes basés sur le contenu n'ont pas besoin de données d'évaluation de la part d'autres utilisateurs pour fonctionner. Cela permet de recommander des items à des utilisateurs nouveaux ou qui n'ont pas encore évalué beaucoup d'items.

Cependant, il y a aussi des limites à prendre en compte, comme :

- La difficulté à identifier les préférences cachées des utilisateurs : les systèmes basés sur le contenu peuvent rencontrer des difficultés à identifier les préférences des utilisateurs qui ne sont pas explicitement exprimées dans les évaluations.

- La complexité de l'analyse de contenu : l'analyse de contenu des items peut être complexe et coûteuse en termes de temps et de ressources. Il peut y avoir des erreurs dans l'extraction des caractéristiques ou des biais dans les profils d'items.
- Le risque d'une recommandation trop similaire : les systèmes basés sur le contenu peuvent recommander des items trop similaires entre eux, ce qui peut devenir ennuyeux pour l'utilisateur.
- La sur-spécialisation (Over-specialization) : La sur-spécialisation est un des inconvénients potentiels des systèmes de recommandation basés sur le contenu. Les systèmes ne peuvent recommander que des items similaires aux préférences et aux intérêts connus de l'utilisateur, ce qui peut limiter la diversité des recommandations reçues par l'utilisateur. Cela peut réduire l'expérience utilisateur car elle ne pourra pas découvrir de nouveaux items qui sortent de sa zone de confort. Il est important pour les systèmes de recommandation basés sur le contenu de disposer d'algorithmes et de données suffisamment robustes pour détecter les similarités entre les items de différents domaines, afin d'éviter la sur-spécialisation [98].

3.4.2 Filtrage collaboratif

Les approches de filtrage collaboratif sont des techniques utilisées pour recommander des éléments (comme des films, de la musique, des livres, etc.) à un utilisateur en se basant sur les préférences et les comportements d'autres utilisateurs similaires. Il existe deux principaux types d'approches de filtrage collaboratif : le filtrage basé sur les utilisateurs et le filtrage basé sur les items.

Les techniques de filtrage collaboratif sont généralement divisées en deux catégories : les méthodes basées sur la mémoire et les méthodes basées sur le modèle. Les méthodes basées sur la mémoire ne se basent que sur la matrice de notation utilisateur-élément et peuvent être mises à jour avec de nouvelles notations. Les méthodes basées sur le modèle, comme les réseaux de neurones, créent un modèle qui apprend à partir des informations de notation utilisateur-élément et recommande de nouveaux éléments [108].

Filtrage basé sur la mémoire

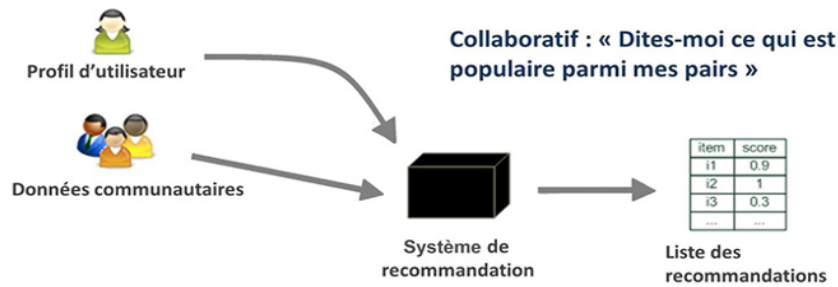


FIGURE 18 : Filtrage collaboratif [99]

- Le filtrage basé sur les utilisateurs : cette approche utilise les préférences et les comportements d'autres utilisateurs similaires pour recommander des éléments à un utilisateur donné. Par exemple, si un utilisateur a aimé des films de genre similaire à un autre utilisateur, il est probable qu'il appréciera également les films que cet autre utilisateur a aimés. Il existe plusieurs mesures de similarité couramment utilisées dans les méthodes de filtrage collaboratif basées sur l'utilisateur. Les plus courantes sont [109] :
 - La similarité cosinus : cette mesure calcule la similarité entre deux utilisateurs en comparant leurs vecteurs de notation. Elle mesure l'angle entre ces vecteurs, où un angle plus petit signifie une plus grande similarité.
 - La similarité de Pearson : cette mesure calcule la corrélation entre les notations de deux utilisateurs. Elle prend en compte la tendance générale des utilisateurs à noter les éléments de manière similaire.
 - La similarité de Jaccard : cette mesure calcule la similarité entre deux utilisateurs en comparant les éléments qu'ils ont notés en commun. Elle est souvent utilisée pour les données binaires où une notation est simplement un indice d'existence ou d'absence d'un élément.
 - La similarité de distance Euclidienne : cette mesure calcule la distance entre les notations de deux utilisateurs. Elle est souvent utilisée pour les données numériques.
 - La similarité de Manhattan : cette mesure calcule la distance entre les notations de deux utilisateurs en utilisant la somme des distances de

chaque notation à la notation correspondante.

Il existe d'autres mesures de similarité, chacun avec ses propres avantages et inconvénients. Le choix de la mesure de similarité dépendra de la nature des données et des besoins de la tâche de recommandation.

- Le filtrage basé sur les items : cette approche utilise les préférences et les comportements des utilisateurs pour les éléments similaires pour recommander des éléments à un utilisateur donné. Par exemple, si un utilisateur a aimé un film, il est probable qu'il appréciera également les films qui ont des caractéristiques similaires (genre, acteurs, réalisateur, etc.).

Filtrage collaboratif basées sur le modèle

Les méthodes de filtrage collaboratif basées sur le modèle, quant à elles, construisent un modèle pour prédire les notations des éléments pour un utilisateur. Ces méthodes sont basées sur des techniques de factorisation de matrices, où la matrice de notation utilisateur-élément est factorisée en représentations latentes d'utilisateurs et d'éléments. Ces représentations latentes capturent les caractéristiques sous-jacentes des utilisateurs et des éléments, qui sont ensuite utilisées pour faire des prédictions [110]. La méthode basée sur le modèle la plus populaire est la factorisation de matrices, qui est basée sur la décomposition en valeurs singulières (SVD). Ces méthodes sont plus puissantes et peuvent mieux gérer les données éparées et les problèmes d'évolutivité que les méthodes basées sur la mémoire. Cependant, elles sont plus coûteuses en termes de calcul et nécessitent plus de données pour entraîner le modèle.

En résumé, les méthodes basées sur la mémoire sont simples et faciles à mettre en œuvre, mais présentent des limitations lorsqu'il s'agit de données éparées et de problèmes d'évolutivité. Les méthodes basées sur le modèle sont plus puissantes, mais nécessitent plus de données et de ressources de calcul pour entraîner le modèle.

Il existe plusieurs algorithmes de filtrage collaboratif basés sur le modèle, qui utilisent des exemples pour prédire les préférences des utilisateurs pour des items non encore évalués.

Voici quelques exemples couramment utilisés :

- L'algorithme de filtrage collaboratif basé sur les k plus proches voisins (k-NN) : cet algorithme utilise la similarité entre les utilisateurs pour trouver les k utilisateurs les plus similaires à un utilisateur donné, et utilise ensuite les évaluations de ces utilisateurs similaires pour prédire les préférences de l'utilisateur donné pour des items non encore évalués.
- L'algorithme de filtrage collaboratif basé sur la factorisation de matrices (MF) : cet algorithme utilise une technique de factorisation de matrices pour approximer la matrice des évaluations des utilisateurs en deux matrices de facteurs latents, une matrice d'utilisateurs et une matrice d'items. Ces facteurs latents sont utilisés pour prédire les évaluations manquantes des utilisateurs pour les items non encore évalués.
- L'algorithme de filtrage collaboratif basé sur les réseaux de neurones : ces algorithmes utilisent des réseaux de neurones pour modéliser les interactions entre les utilisateurs et les items, et utilisent ensuite ces modèles pour prédire les évaluations manquantes des utilisateurs pour les items non encore évalués.

Il est important de noter que ces algorithmes sont souvent utilisés en combinaison avec d'autres techniques de filtrage collaboratif pour améliorer la précision des prédictions. Il est important de choisir le bon modèle en fonction des données et des besoins de l'application.

Avantages et limites

Les méthodes de filtrage collaboratif sont des techniques populaires pour fournir des recommandations personnalisées aux utilisateurs. Ces méthodes ont plusieurs avantages importants, notamment :

- Personnalisation : les systèmes de filtrage collaboratif peuvent utiliser les préférences des utilisateurs similaires pour fournir des recommandations personnalisées à un utilisateur donné. Cela permet de fournir des recommandations plus pertinentes et plus adaptées aux intérêts de l'utilisateur.
- Utilisation des informations implicites : les systèmes de filtrage collaboratif peuvent utiliser des informations implicites (comme les articles consultés ou les achats effectués) pour fournir des recommandations, plutôt que de

nécessiter des informations explicites (comme les notes ou les évaluations) de la part des utilisateurs.

- Prise en compte des interactions entre les utilisateurs et les items : les systèmes de filtrage collaboratif peuvent prendre en compte les interactions entre les utilisateurs et les items pour fournir des recommandations.
- Utilisation dans divers domaines : les systèmes de filtrage collaboratif peuvent être utilisés dans divers domaines tels que la recommandation de produits, de contenus, de musique, de films, de livres, etc.
- Combinaison avec d'autres techniques : les systèmes de filtrage collaboratif peuvent être combinés avec d'autres techniques de filtrage pour améliorer les prédictions.

Les principales limites de filtrage collaboratif sont :

- Problème de démarrage à froid (cold start problem) : il s'agit d'un problème courant dans les systèmes de filtrage collaboratif, qui se produit lorsque les systèmes doivent recommander des items pour des utilisateurs ou des items qui n'ont pas encore été évalués. Ce problème peut être résolu en utilisant des techniques de filtrage hybrides qui combinent des informations de contenu et des informations sociales pour générer des recommandations.
- Sparsity : il s'agit d'un autre problème courant dans les systèmes de filtrage collaboratif, qui se produit lorsque les matrices d'évaluation ont de nombreux éléments manquants. Cela peut rendre difficile la génération de recommandations de qualité pour certains utilisateurs et certains items. Il existe plusieurs techniques pour contourner cette limitation comme l'utilisation de techniques de complétion de matrices pour estimer les évaluations manquantes.
- Dépendance de la qualité des données d'évaluation : les systèmes de filtrage collaboratif dépendent de la qualité des données d'évaluation pour les utilisateurs et les items pour générer des recommandations. Si les données sont bruitées ou incomplètes, les recommandations peuvent être de mauvaise qualité.

- Recommandation d'items populaires : les systèmes de filtrage collaboratif peuvent avoir tendance à recommander des items qui sont populaires mais qui ne sont pas nécessairement pertinents pour l'utilisateur. Cela peut être résolu en utilisant des techniques de pondération pour atténuer l'influence des items les plus populaires sur les recommandations.

Exemples de systèmes collaboratifs

Exemples de systèmes populaires basés sur le filtrage collaboratif :

- Netflix, un service de streaming vidéo qui propose des recommandations de films et de séries en fonction de l'historique et des préférences de visionnage des utilisateurs.
- Amazon, un revendeur en ligne qui fournit des recommandations de produits en fonction des achats et de l'historique de recherche des utilisateurs.
- Spotify, un service de streaming musical qui propose des recommandations de chansons en fonction de l'historique d'écoute et des préférences des utilisateurs.
- MovieLens, un service de recommandation de films géré par le GroupLens Research de l'Université du Minnesota qui utilise le filtrage collaboratif pour faire des recommandations de films.

3.4.3 Systèmes hybrides

Chaque méthode de recommandation présente ses propres avantages et inconvénients, ce qui a conduit les chercheurs à les combiner pour obtenir un système de recommandation qui tire parti de ces forces et surmonte la plupart des limites. Enfin, les chercheurs ont eu l'idée d'utiliser une méthode hybride pour les systèmes de recommandation. Une méthode de filtrage hybride peut utiliser une combinaison de filtrage collaboratif et de filtrage démographique ou de filtrage collaboratif et de filtrage basé sur le contenu pour obtenir de meilleurs résultats [111].

Burke a défini plusieurs types d'hybridation dans son article "Hybrid Web Recommender Systems" [112] :

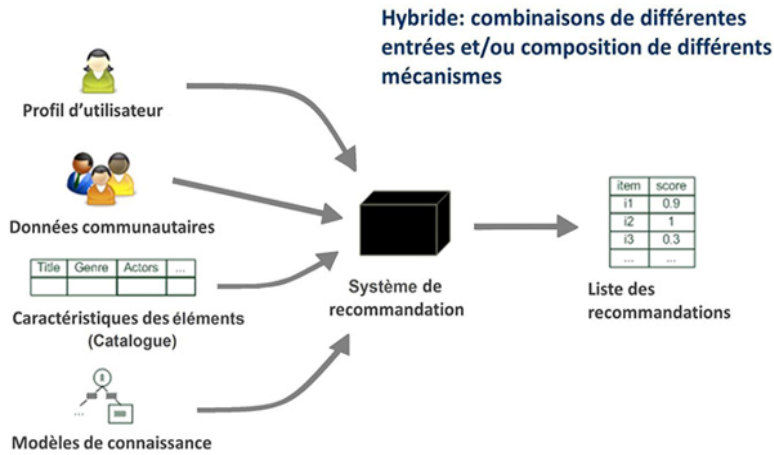


FIGURE 19 : Système de recommandation hybride [99]

- Pondéré (Weighted) : Les scores (ou votes) de plusieurs techniques de recommandation sont combinés ensemble pour produire une recommandation unique.
- Commutation (Switching) : Le système passe d'une technique de recommandation à l'autre en fonction de la situation actuelle.
- Mixte (Mixed) : Les recommandations de plusieurs systèmes de recommandation sont présentées ensemble.
- Combinaison de caractéristiques (Feature Combination) : Les caractéristiques dérivées de différentes sources sont combinées et transmises à un seul algorithme de recommandation.
- Augmentation de caractéristiques (Feature Augmentation) : Une technique de recommandation est utilisée pour calculer des caractéristiques, qui sont ensuite utilisées en tant qu'entrée pour la prochaine technique.
- Cascade (Cascade) : Dans ce type d'hybridation, une technique de recommandation est utilisée pour produire un premier classement des items candidats et une deuxième technique affine ensuite la liste des recommandations.
- Niveau méta (Meta-level) : Une technique de recommandation est appliquée pour produire un modèle, qui est ensuite utilisé en tant qu'entrée par la prochaine technique.

Avantages et limites

Les systèmes de recommandation hybrides combinent deux ou plusieurs techniques de recommandation pour obtenir de meilleures performances avec moins d'inconvénients que les techniques individuelles. Parmi les avantages des systèmes de recommandation hybrides :

- Amélioration de la précision : En combinant différentes techniques de recommandation, les systèmes hybrides peuvent souvent atteindre une plus grande précision par rapport aux systèmes à méthode unique.
- Gestion des problèmes de démarrage à froid et de sparsité : Les systèmes hybrides peuvent gérer les problèmes de démarrage à froid, lorsqu'il y a peu d'informations sur un nouvel utilisateur, et de sparsité, lorsqu'il y a peu d'interaction entre les utilisateurs et les articles.
- Diversité des recommandations : En utilisant plusieurs méthodes, les systèmes hybrides peuvent fournir un ensemble diversifié de recommandations aux utilisateurs.

Les systèmes de recommandation hybrides peuvent présenter certaines limites :

- Complexité accrue : Les systèmes hybrides sont souvent plus complexes que les systèmes à méthode unique et peuvent nécessiter plus de temps et de ressources pour leur développement et leur maintenance.
- Difficultés d'intégration : La combinaison de plusieurs méthodes peut être un défi et peut entraîner des difficultés d'intégration et une diminution des performances si elle n'est pas effectuée correctement.
- Augmentation des frais de calcul : Les systèmes hybrides peuvent être plus gourmands en ressources informatiques que les systèmes à méthode unique.

3.4.4 Recommandation communautaire

Un système de recommandation basé sur la communauté est un système de recommandation qui utilise les informations fournies par la communauté, telles

que les évaluations et les préférences, pour générer des recommandations personnalisées pour les utilisateurs.

Le principe de fonctionnement de ces systèmes repose sur l'idée que les personnes qui ont des intérêts et des goûts similaires ont tendance à fournir des évaluations similaires pour les produits ou les services. Le système analyse les modèles dans les évaluations de la communauté et les utilise pour faire des prédictions sur les articles qu'un utilisateur particulier pourrait aimer[113].

Parmi les exemples de systèmes de recommandation communautaires, citons les sites Web de recommandation de films et de livres en ligne, tels que IMDb, Netflix et Goodreads, où les utilisateurs peuvent évaluer et commenter des produits et où le système peut générer des recommandations sur la base des préférences et des évaluations d'utilisateurs similaires. Un autre exemple est la fonction "Les clients qui ont acheté ceci ont également acheté" d'Amazon, qui utilise des recommandations basées sur la communauté pour suggérer des produits liés aux articles qu'un utilisateur a précédemment achetés.

Facebook se base également sur un système communautaire. Il utilise divers signaux provenant de ses utilisateurs, tels que les "j'aime", les partages, les commentaires et l'activité en ligne, pour recommander du contenu à ses utilisateurs. Les recommandations sont générées en fonction des préférences et des comportements de la communauté, l'idée étant que les personnes qui ont des intérêts et des goûts similaires auront tendance à s'intéresser à un contenu similaire[114].

Par exemple, le fil d'actualité de Facebook utilise un algorithme de recommandation pour afficher le contenu le plus pertinent pour chaque utilisateur. L'algorithme prend en compte le contenu qu'un utilisateur a aimé ou avec lequel il a interagi dans le passé, ainsi que le contenu que ses amis ont aimé ou partagé. Cela permet à Facebook de fournir des recommandations personnalisées basées sur les préférences et le comportement du réseau social de l'utilisateur.

3.4.5 Recommandation basée sur le filtrage démographique

Les systèmes de recommandation basés sur la démographie utilisent des informations démographiques, telles que l'âge, le sexe, la profession, le revenu, le niveau d'éducation, etc., pour faire des recommandations aux utilisateurs. L'idée

derrière cette approche est que les personnes ayant des profils démographiques similaires ont tendance à avoir des goûts, des centres d'intérêt et des préférences similaires. Ces systèmes utilisent les informations démographiques pour regrouper les utilisateurs en segments et ensuite faire des recommandations en fonction des préférences des utilisateurs dans chaque segment. Par exemple, un système de recommandation basé sur la démographie pour les films pourrait recommander des films d'action à de jeunes utilisateurs masculins et des comédies romantiques à de jeunes utilisatrices féminines.

3.4.6 Recommandation à base de connaissance

Un système basé sur la connaissance est un type de système de recommandation qui utilise des connaissances sur l'utilisateur et l'article pour faire des recommandations[115]. Ces connaissances peuvent prendre la forme de préférences explicites ou d'autres informations sur l'utilisateur, et sont utilisées pour déterminer quel article répondra le mieux aux besoins de l'utilisateur. Ces systèmes reposent sur une prise de décision basée sur les données et peuvent être considérés comme une forme d'intelligence artificielle ou d'apprentissage automatique, car ils utilisent des algorithmes pour analyser et interpréter les données disponibles. Cependant, la principale différence entre un système basé sur la connaissance et les autres types de systèmes de recommandation est qu'il est conçu pour se fonder sur une compréhension plus approfondie des besoins et des motivations de l'utilisateur, plutôt que sur son comportement passé ou d'autres facteurs superficiels.

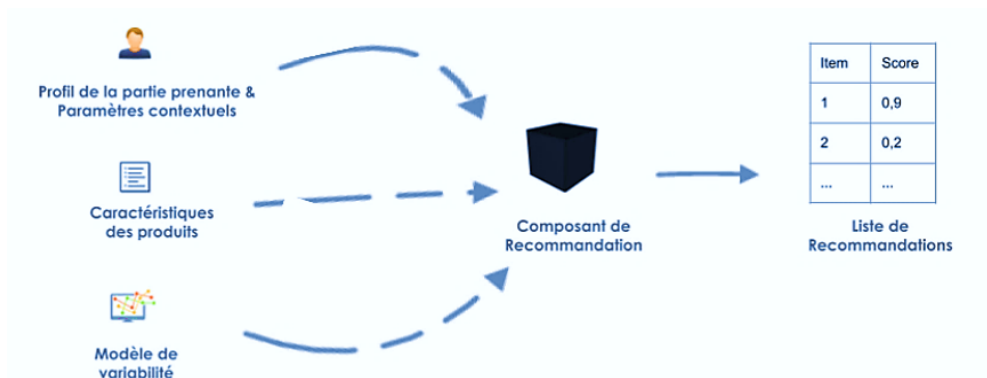


FIGURE 20 : Recommandation à base de connaissance [116]

3.5 Critères d'évaluation des systèmes de recommandation

Les systèmes de recommandation doivent être évalués pour de nombreuses raisons, comme la comparaison de la qualité des techniques. Les mesures d'évaluation jouent un rôle important dans la comparaison de plusieurs solutions pour un même problème. Les mesures d'évaluation sont classées en quatre groupes principaux [98] :

- les mesures de prédiction, telles que la précision
- les mesures de recommandation d'ensemble, telles que la précision et le rappel
- les mesures de recommandation de rang, telles que la demi-vie et
- la diversité et la nouveauté.

3.5.1 Qualité de la prédiction

La qualité de la prédiction est le premier critère qui a été utilisé pour comparer les systèmes de recommandation. L'une des métriques de prédiction les plus utilisées est l'erreur absolue moyenne (MAE) et d'autres métriques dérivées de celle-ci telles que l'erreur quadratique moyenne et l'erreur absolue moyenne normalisée.

- La Précision mesure la proportion d'éléments recommandés qui sont pertinents pour l'utilisateur. La formule est la suivante :

$$\text{Précision} = \frac{\text{Nombre d'éléments pertinents recommandés}}{\text{Nombre total d'éléments recommandés}} \quad (\text{Chapitre II.1})$$

- Mean Absolute Error (MAE) est une métrique couramment utilisée pour évaluer la qualité de la prédiction en machine learning. Il mesure la différence moyenne entre les valeurs prédites et les valeurs réelles pour un ensemble de données. Le MAE est calculé en prenant la somme absolue de la différence entre la valeur prédite P_i et la valeur réelle A_i pour chaque élément i dans l'ensemble de données, divisé par le nombre total d'éléments N . MAE peut être calculée à l'aide de la formule suivante :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |P_i - r_i| \quad (\text{Chapitre II.2})$$

- Root Mean Square Error (RMSE) est une autre métrique couramment utilisée pour évaluer la qualité de la prédiction en machine learning. Il mesure la moyenne des erreurs au carré entre les valeurs prédites et les valeurs réelles pour un ensemble de données. Le RMSE est calculé en prenant la racine carrée de la somme des erreurs au carré divisée par le nombre total d'éléments. Formule mathématique pour le RMSE :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - P_i)^2} \quad (\text{Chapitre II.3})$$

Où n est le nombre total d'éléments dans l'ensemble de données.

Le RMSE est un indicateur fiable de la qualité de la prédiction car il tient compte de la gravité des erreurs. Cependant, il peut surestimer la gravité des erreurs lorsque les différences entre les valeurs prédites et réelles sont considérables. Un RMSE plus faible indique une meilleure performance du système.

3.5.2 Qualité de l'ensemble des recommandations

La qualité de l'ensemble des recommandations est importante pour certains utilisateurs. La précision, le rappel et le F1 sont les métriques les plus pertinents pour évaluer cette qualité. La précision mesure la proportion d'articles pertinents parmi les articles recommandés, tandis que le rappel mesure la proportion d'articles pertinents parmi l'ensemble des articles pertinents. Le F1 est une combinaison des deux. On calcule ces métriques en effectuant n recommandations de test pour un utilisateur u et en comparant les recommandations (X_u) avec l'ensemble des n recommandations (Z_n) en utilisant un seuil .

$$\text{precision} = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i \in Z_u | r_{u,i} \geq \theta\}|}{n} \quad (\text{Chapitre II.4})$$

Le rappel est un ratio qui mesure la proportion d'articles pertinents que l'algorithme de recommandation a réussi à inclure dans les recommandations faites

à l'utilisateur.

$$\text{recall} = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i \in Z_u | r_{u,i} \geq \theta\}|}{|\{i \in Z_u | r_{u,i} \geq \theta\}| + |\{i \in Z_u^c | r_{u,i} \geq \theta\}|} \quad (\text{Chapitre II.5})$$

Où $\{i \in Z_u^c | r_{u,i} \geq \theta\}$ indique le nombre d'éléments pertinents qui n'ont pas été recommandés.

Le score F1 est une métrique utilisée pour évaluer les performances d'un classificateur binaire. Il est calculé en tant que moyenne harmonique de la précision et du rappel. La formule pour le score F1 est :

$$\text{F1 measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (\text{Chapitre II.6})$$

Où Précision est le nombre de vrais positifs divisé par le nombre de vrais positifs et de faux positifs et Rappel est le nombre de vrais positifs divisé par le nombre de vrais positifs et de faux négatifs.

Le score F1 prend en compte à la fois la précision et le rappel d'un classificateur et les équilibre pour donner un score global qui indique les performances du classificateur. Le score F1 varie entre 0 et 1, avec 1 indiquant un classificateur parfait et 0 indiquant un classificateur qui ne fait aucune prédiction correcte.

Dans la figure suivante, nous montrons le rôle du rappel et de la précision pour évaluer les systèmes de recommandation.

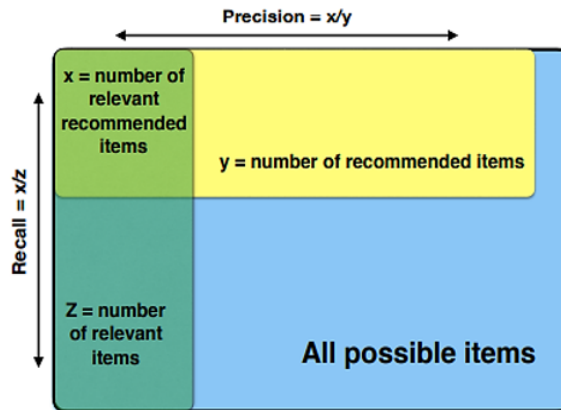


FIGURE 21 : Précision et Rappel concernant tous les Items

3.6 Utilisation des systèmes multi-agents dans les systèmes de recommandation

Les systèmes multi-agents peuvent être utilisés dans les systèmes de recommandation pour modéliser les interactions entre différents acteurs tels que les utilisateurs, les articles et les fournisseurs de recommandations. Chaque acteur peut être modélisé en tant qu'agent indépendant doté de ses propres objectifs et comportements. Les agents peuvent interagir entre eux pour influencer les recommandations produites par le système. De plus, les systèmes multi-agents peuvent permettre une mise à jour en temps réel des recommandations en fonction des préférences et des comportements changeants des utilisateurs. Dans les systèmes de recommandation multi-agents, chaque agent peut avoir une responsabilité spécifique, telles que la collecte de données sur les préférences des utilisateurs, l'analyse de ces données pour produire des recommandations, ou la présentation de ces recommandations aux utilisateurs. Les agents peuvent utiliser des techniques telles que l'apprentissage automatique et la reconnaissance de patrons pour effectuer leurs tâches.

Les interactions entre les agents peuvent être modélisées à l'aide de contrats ou de règles de négociation, qui décrivent comment les agents peuvent coopérer pour produire des recommandations de qualité. Cela permet une flexibilité accrue dans la façon dont les recommandations sont produites et permet également de prendre en compte les différences entre les utilisateurs. En utilisant les systèmes multi-agents, les systèmes de recommandation peuvent également être plus robustes et tolérants aux erreurs, car les différents agents peuvent surveiller et corriger les erreurs commises par d'autres agents. De plus, les systèmes multi-agents peuvent facilement être adaptés pour prendre en compte de nouveaux types de données ou de nouveaux types d'interactions entre les agents. En conclusion, les systèmes multi-agents peuvent fournir une approche souple et efficace pour les systèmes de recommandation, en permettant une interaction flexible et en temps réel entre les différents acteurs impliqués.

Il existe de nombreux travaux de recherche sur l'utilisation des systèmes multi-agents dans les systèmes de recommandation. Voici quelques exemples de travaux pertinents :

- "A Multi-Agent Recommender System" de A. Jorge Morais [117] : Cette étude propose un système de recommandation multi-agent qui utilise la collaboration entre les agents pour améliorer la qualité des recommandations produites.
- " A Multi-Agent Recommender System Using Social Networks " [118] de F. Siala : Ce travail de recherche explore l'utilisation des systèmes multi-agents pour la recommandation dans les réseaux sociaux, en prenant en compte les relations sociales entre les utilisateurs.
- "MARST : Multi-Agent Recommender System for e-Tourism Using Reputation Based Collaborative Filtering" de Bedi et al [119] : présente un système de recommandation multi-agents pour l'e-tourisme pour recommander des services touristiques aux utilisateurs. Ce système utilise l'algorithme de filtrage collaboratif basé sur la réputation qui ajoute la réputation à l'approche collaborative existante pour générer des recommandations pertinentes et gérer le problème du démarrage à froid des nouveaux utilisateurs dans le domaine du tourisme.
- " IoT and Context-Aware Mobile Recommendations Using Multi-agent Systems " : de B. Twardowski et al [120] : Ce travail présente un système multi-agent pour la recommandation contextuelle, qui utilise des informations sur le contexte, acquises de l'Internet des objets (IoT), pour produire des recommandations plus précises sur des dispositifs mobiles à l'aide de données contextuelles.
- "Social Recommendation System Based on Multi-agent Deep Reinforcement Learning" de L. Yinggang et al [121] Cette étude propose un système de recommandation multi-agent qui utilise l'apprentissage profond par renforcement pour produire des recommandations de haute qualité. Plus précisément, un agent est chargé de capturer les préférences dynamiques des utilisateurs en fonction de leur interaction avec le système. L'autre agent exploite un réseau social omniprésent pour minimiser les démarrages à froid et la rareté des données pour certains utilisateurs.

Ces travaux montrent comment les systèmes multi-agents peuvent être utilisés pour résoudre les défis courants dans les systèmes de recommandation, tels que la

prise en compte des préférences changeantes des utilisateurs et la recommandation en fonction du contexte.

3.7 Notre Approche de recommandation basé sur les systèmes multi-agents

3.7.1 Méthodologie

Le nombre d'étudiants utilisant les plateformes d'apprentissage en ligne augmente de façon explosive de jour en jour, de même que le nombre de cours en ligne proposés. Il devient donc essentiel d'utiliser un système de recommandation qui les aide à trouver les cours optimaux et appropriés pour les apprenants.

Dans cette section, nous présentons notre proposition d'architecture d'un système de recommandation dans une plate-forme d'apprentissage en ligne avec un système multi-agents, le système est basé sur une méthode basée sur le contenu pour recommander des cours universitaires pertinents à un étudiant, cette méthode utilise uniquement des informations sur la description et les caractéristiques des cours que les utilisateurs ont précédemment consultés. Nous calculons la similarité entre les cours, pour obtenir la liste des N meilleurs cours, puis nous utilisons deux agents, le premier pour les commentaires positifs et le second pour les commentaires négatifs (messages reçus du chat, forum et étiquettes de toutes les plateformes). Ces agents vont améliorer la liste de recommandation du Top N en fonction des valeurs extraites des commentaires. La figure ci-dessous (figure 3) montre l'architecture structurelle de notre système de recommandation proposé pour l'e-learning, les étudiants visualisent et évaluent les documents, ils peuvent également interagir par des "likes" ou des commentaires. La sortie du système est le Top-N des cours similaires à l'apprenant actif [122]. Nous décrivons ci-dessous les étapes du workflow du système de recommandation de cours :

1. Importation des données des apprenants et des cours depuis la base de données.
2. Le prétraitement des données comprend l'intégration, la collecte et la préparation des données.
3. Pondération des données en utilisant la mesure TF-IDF.

4. Le calcul de la similarité en utilisant la similarité cosinus et la génération de la liste de recommandations Top N.
5. Extraction des commentaires positifs et négatifs de chaque cours en utilisant un système multi-agents composé de deux agents, le premier pour les commentaires positifs et le second pour les commentaires négatifs.
6. Reclassification de la liste Top N en tenant compte des valeurs obtenues par le système multi-agents.
7. Exportation de la liste des cours recommandés vers l'apprenant actif.

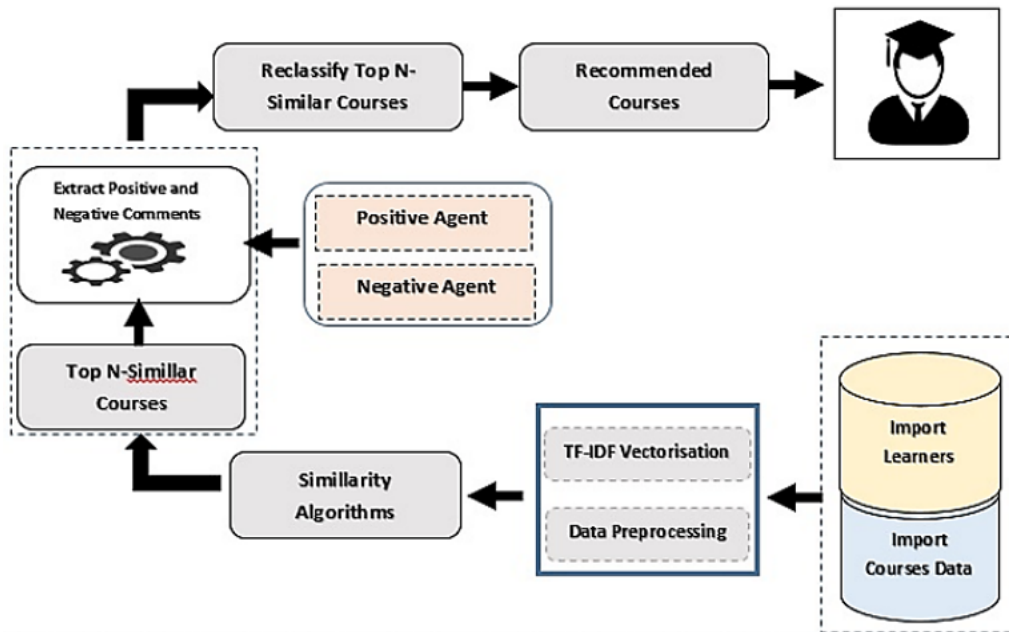


FIGURE 22 : Proposition d'architecture de système de recommandation pour l'apprentissage en ligne avec un système multi-agents [122]

3.7.2 Résultats expérimentaux

Pour tester les systèmes de recommandation, les jeux de données sont disponibles sur MovieLens [95], mais ces données sont moins appropriées pour tester les systèmes de recommandation en e-learning car les jeux de données sont basés sur le classement des films.

Pour l'expérience, nous utilisons un échantillon de jeux de données de cours

universitaires, qui contiennent des informations (Id du cours, Titre du cours, Description...) sur 8012 cours comme indiqué dans cette capture.

	CourseTitle	Description
1063	Cisco CCENT (640-822 ICND1)	This Cisco CCENT training course covers the in...
4861	Linux Server Skills for Windows Administrators	As more and more Internet server operations mo...
1842	SQL Server 2012-2014 Native File Streaming	This course teaches SQL Server developers the ...
2869	Animating an Acrobatic Fight Scene in Maya	Throughout these lessons, we will create a fig...
5399	Working with Polymer.js Elements	Polymer.js is one of the most promising framew...

FIGURE 23 : L'en-tête des cours "DataFrame" [122]

Après avoir prétraité les données et choisi les caractéristiques pertinentes pour notre problème, qui sont le titre et la description du cours, nous commençons à calculer les scores TF-IDF, et nous calculons la similitude en cosinus entre les cours. L'étape suivante consiste à retourner les cours les plus similaires sur la base du score de similitude en cosinus. Après avoir développé et implémenté notre système de recommandation, nous avons comme résultat la liste des cours recommandés.

Par exemple, les dix cours les plus similaires à "ASP.NET 3.5 Fundamentals" sont les suivants :

Identifiant du cours	Titre du cours	Score
8	ASP.NET 3.5 Working With Data	0.76159542824619
5325	Building a Web App with ASP.NET Core, MVC 6,	0.76159455399578
6	ASP.NET Ajax Fundamentals	0.76159455176471
6174	ASP.NET Identity 2 Fundamentals	0.76159454419870
565	Introduction to ASP.NET Web Services	0.76159452630515

Identifiant du cours	Titre du cours	Score
6687	Upgrading Your ASP.NET Core 1.x Web API to 2.0	0.76159452612046
1882	ASP.NET 5 : First Look	0.76159451259603
5917	Implementing and Securing an API with ASP.NET	0.76159449291658
10	ASP.NET MVC Fundamentals	0.76159447973701
791	One ASP.NET From Scratch	0.76159447610679

TABLE 5 : Les dix premiers cours similaires de “ASP.NET 3.5 Fundamentals”

3.7.3 Discussion

Les systèmes de recommandation jouent un rôle important dans l'amélioration des systèmes d'apprentissage en ligne ; ils tentent de recommander intelligemment des cours bénéfiques à un apprenant en fonction de ses préférences et de ses besoins. Notre approche proposée donne une analyse des systèmes de recommandation utilisés dans le contexte de l'e-learning, d'après notre analyse, nous constatons que la méthode de filtrage collaboratif est la technique la plus utilisée dans les systèmes de recommandation de cours. Cependant, elle a quelques inconvénients comme le problème de démarrage à froid et la rareté des données, pour cette raison, notre approche proposée est basée sur le filtrage basé sur le contenu et le système multi-agent, peut traiter ces défis, les résultats de l'expérience montrent que les cours recommandés sont très similaires aux préférences de l'utilisateur.

4 Conclusion

En conclusion, les systèmes de recommandation et les technologies Big Data sont des éléments clés pour le développement des systèmes e-learning. Dans ce chapitre, nous avons présenté les différentes caractéristiques et technologies des Big Data, ainsi que son impact sur le domaine d'apprentissage en ligne. Les

Big Data permettent aux systèmes e-learning de collecter, stocker et analyser de grandes quantités de données pour améliorer les décisions concernant la formation et le développement de programmes d'apprentissage. Les technologies associées aux Big Data, telles que le traitement en temps réel, l'analyse de données en nuage et la visualisation de données, offrent une grande flexibilité pour traiter les données et les utiliser pour prendre des décisions. Nous avons défini également les systèmes de recommandation, leurs approches et leurs critères d'évaluation. Les systèmes de recommandation, quant à eux, utilisent les Big Data pour personnaliser l'expérience d'apprentissage des apprenants en fournissant des recommandations basées sur leur historique académique, leur progression dans les cours, leur style d'apprentissage et leurs centres d'intérêt. Les algorithmes d'apprentissage automatique, tels que les algorithmes de filtrage collaboratif et les algorithmes de contenu, sont utilisés pour construire des modèles de données qui peuvent être utilisés pour prédire les intérêts futurs des utilisateurs en matière d'apprentissage. Les systèmes de recommandation peuvent aider à offrir une expérience d'apprentissage personnalisée à chaque utilisateur, ce qui peut améliorer leur engagement et leur réussite académique.

En fin de compte, les systèmes de recommandation et les technologies Big Data sont des domaines en constante évolution qui permettent aux systèmes e-learning de mieux comprendre les besoins et les intérêts des utilisateurs et de fournir des programmes d'apprentissage plus pertinents et adaptés à chaque personne. Les progrès continus dans ces domaines offrent de nouvelles opportunités pour améliorer les expériences d'apprentissage et stimuler la croissance et le développement des systèmes e-learning.

Dans le chapitre suivant, nous nous concentrons sur la mise en œuvre d'un modèle de recommandation d'e-learning à base d'ontologie et du web sémantique. Ce modèle de recommandation utilise des technologies sémantiques pour acquérir, représenter et inférer des connaissances à partir de données d'apprentissage. Dans cette section, nous explorons les principes fondamentaux de l'ontologie et du web sémantique, et leur pertinence pour les systèmes de recommandation en e-learning. Ensuite, nous discutons des différentes étapes de la construction de notre modèle de recommandation, y compris l'acquisition de connaissances, la représentation de connaissances et l'inférence. Enfin, nous évaluons l'efficacité de

notre modèle de recommandation en utilisant différentes mesures d'évaluation.

Chapitre III

Mise en œuvre d'un modèle de recommandation E-learning à base d'ontologie et du web sémantique

Sommaire

1	Introduction	96
2	Le web sémantique et l'apprentissage en ligne	98
3	Les ontologies	103
4	Approche proposée	111
5	Résultats expérimentaux	125
6	Conclusion	132

1 Introduction

Ces dernières années, les systèmes de recommandation ont été reconnus comme une solution utile au problème de surcharge de données. Dans l'apprentissage en ligne, ces systèmes sont conçus pour fournir des données adaptées aux besoins spécifiques des apprenants. Cela favorise le développement de leurs connaissances et améliore leurs capacités cognitives. Au départ, les systèmes de recommandation ont été conçus pour les sites Web commerciaux, où ils ont été utilisés pour prédire les préférences des utilisateurs et leur recommander des produits ou services pertinents. Cependant, avec l'expansion de l'utilisation d'Internet, ces systèmes se sont étendus à d'autres domaines, tels que l'apprentissage en ligne, les médias sociaux et les systèmes de santé. De nos jours, les systèmes de recommandation

sont devenus un outil de prédiction largement utilisé dans divers secteurs. Ils analysent les données des utilisateurs, telles que leur historique de navigation, leur comportement d'achat ou leurs interactions sur les réseaux sociaux, pour générer des prédictions pour l'avenir. A cet égard, Les systèmes de recommandation jouent un rôle important dans l'apprentissage en ligne en aidant les apprenants à trouver les ressources d'apprentissage les plus adaptées à leurs besoins.

Cependant, contrairement aux systèmes de recommandation utilisés dans le commerce électronique, les systèmes de recommandation pour l'apprentissage en ligne doivent tenir compte de trois éléments clés : l'apprenant, l'objet d'apprentissage et l'activité d'apprentissage. En outre, l'objet d'apprentissage a la possibilité d'appartenir aux différentes catégories, ce qui n'est pas offerte par les systèmes de E-commerce. En général, les systèmes d'apprentissage sont conçus pour fonctionner avec des bases de données statiques, qui ne changent pas dynamiquement. Chose qui s'articule comme un problème donnant lieu à des systèmes moins efficaces. Dans ce contexte, les ontologies dynamiques peuvent être une solution efficace à ce genre de problème [123]. Prenant en considération, d'une part, la nature des plateformes e-learning en libre accès, et d'autre part, la difficulté à personnaliser des contenus dynamiques et à créer des profils utilisateurs dynamiques, le développement d'un système de recommandation e-learning avec des ontologies dynamiques (ERSDO) [124] basé sur le comportement d'apprentissage de chaque apprenant, présente pour nous un véritable défi.

Comme il est mentionné par Devedziu et al [125], la personnalisation est le fait d'adapter l'expérience d'apprentissage à chaque apprenant en analysant ses connaissances, ses compétences et ses préférences d'apprentissage individuelles. Dans un environnement d'apprentissage, plusieurs éléments sont nécessaires pour développer un système efficace. Il s'agit notamment du style d'apprentissage de l'apprenant, de la structure de l'information (ontologie) et de l'approche technique utilisée (système de recommandation). Ces éléments sont fondamentaux pour garantir le succès d'un système d'apprentissage.

Le style d'apprentissage, en tant que concept, décrit la manière dont les apprenants découvrent, assimilent et mémorisent l'information [126]. Dans la littérature, les théories des styles d'apprentissage expliquent comment les apprenants peuvent être classés en fonction de leur style d'apprentissage (Visuel, Auditif, Ki-

nesthésique...). Bien qu'il existe plus que soixante-dix modèles de styles d'apprentissage dans la littérature, seuls sept modèles sont généralement utilisés dans les systèmes d'apprentissage en ligne. Ces modèles sont attribués à David Kolb, Peter Honey, Alan Mumford, Neil Fleming avec son célèbre VAK/VARK, le NASSP d'Anthony Gregorc et celui de Felder Silverman[127].

Ce travail tend vers un double objectif, premièrement, il vise à étudier la capacité de construire des ontologies dynamiques pour les apprenants et leurs contenus. Deuxièmement, il vise à suggérer une recommandation améliorée correspondant le mieux aux besoins de l'apprenant. Dans cette optique, afin de réaliser un travail qui répond aux normes et aux standards internationaux, l'utilisation des techniques du web sémantique et du web sémantique social seront des éléments indispensables dans notre proposition.

2 Le web sémantique et l'apprentissage en ligne

2.1 Le Web sémantique

Le Web sémantique est une amélioration de la vision classique du World Wide Web dans laquelle toutes les données sont structurées et organisées sémantiquement dans une ontologie sous-jacente. L'objectif ultime du web sémantique vise à créer un Web lisible pour les machines, dans lequel l'ordinateur peut automatiquement interpréter le sens de l'information. Le concept de web sémantique a été relevé la première fois par Tim Berners-Lee en 2001 [128], où l'a défini comme suit :

"Le web sémantique est une extension du web actuel dans laquelle les informations sont dotées d'une signification bien définie, ce qui permet à l'ordinateur et aux personnes de travailler en coopération." [128].

Le Web sémantique fait appel à de nombreuses technologies se focalisant sur des tâches précises et qui se basent en partie les uns sur les autres. En 2000 Berners Lee a illustré cela par un modèle en couches représentant la structure théorique du web sémantique, représenté dans la figure 24.

Les techniques XML (Extensible Markup Language) et URI (Uniform Resource Identifier) constituent la base du Web sémantique. D'un côté, XML est

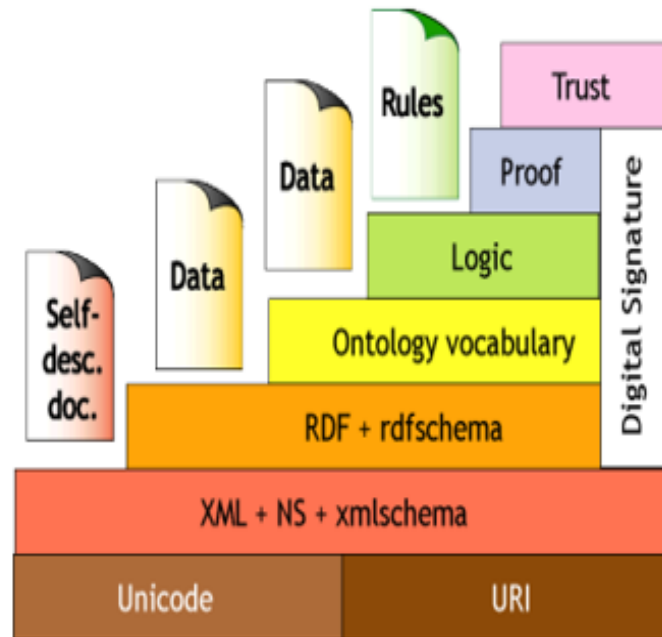


FIGURE 24 : Le modèle en couches original selon Tim Berners-Lee [128]

un langage simple qui permet de structurer les données. D'un autre côté, les URI sont utilisés essentiellement pour identifier des ressources. Ci-dessous se trouvent les autres techniques du web sémantique utilisées par Tim Berners-Lee dans son modèle en couche :

- Le niveau RDF (Resource Description Framework), pour représenter les relations entre les données.
- Le niveau ontologique, qui définit la norme formelle sur la signification des données.
- Le niveau logique, qui permet un raisonnement intelligent avec les données structurées.
- Contrairement au niveau RDF, considéré comme un langage de communication, les deux autres couches ontologique et logique de modèle de Berners-Lee s'intéressent davantage à la normalisation et la structuration des données.
- Les couches finales du modèle de Proof et Trust sont des couches de vérification de données, dans la mesure où la couche contrôle vérifie si les

conclusions établies dans la couche logique sont correctes. Ainsi, ce modèle vérifie comment les liaisons ont été faites. Ainsi qu'à partir de quelles données les relations ont été établies.

- La dernière couche 'Trust' vérifie, en revanche, la provenance des données en rapport avec le degré de la sécurité de la source de ces données [129]. Au fil des années, le modèle en couches de Berners-Lee a été affiné et étendu par Steve Bratt en 2007 [130]. La figure 25 illustre le détail de ce modèle.

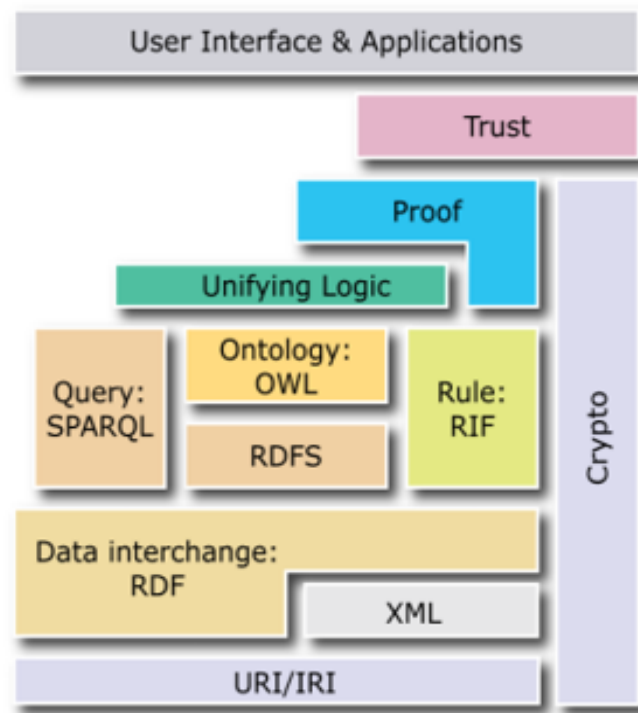


FIGURE 25 : Une version révisée du modèle en couches de Steve Bratt [130]

Dans le nouveau modèle de Steve Bratt on constate que d'autres langues standardisées ont été ajoutées pour faciliter la manipulation des données, à titre d'exemple :

- SPARQL (Simple Protocol and Rdf Query Language) comme langage d'interrogation pour les documents RDF

- OWL (Web Ontology Language) en tant que langage de description d'ontologies ainsi que la spécification de description de règles RIF (Rule Interchange Format).

Ces langages ne sont toutefois qu'une recommandation du W3C, d'où provient ce modèle, permettant l'accès à la possibilité d'utiliser d'autres spécifications en fonction de la nature des données. Au sein des couches logiques Proof et Trust, aucun langage ou technologie standardisé ne s'est encore développé, même au moment de la présente élaboration. Chose tendant vers la conclusion sur laquelle porte ce travail dans le sens où l'accent est surtout mis sur les couches RDF/ontologie et logique.

2.2 Le Web sémantique social

Le terme Web Sémantique Social désigne un concept de création, de traitement et de diffusion d'informations par l'association des technologies et des approches du Web sémantique ainsi que du Web social. À ce propos, le Web sémantique tente de donner un sens aux données et de les rendre lisibles par les machines grâce à l'utilisation de métadonnées et d'ontologies. En revanche, l'approche du web social consiste à créer une plateforme où les utilisateurs peuvent se rencontrer, interagir et créer des contenus. A titre d'exemple, des wikis, des blogs et des services de partage de photos/vidéos [131].

En 2004, Cahier et Zacklad ont conféré pour la première fois le concept "Web socio-sémantique" où ils ont expliqué que l'homme ne devrait pas être exclu du processus d'extraction de données, mais d'autre côté contraire, il doit être impliqué dans l'ensemble de déroulement. Dans leur article, ils ont formulé ce concept comme suit :

"Les applications du "Web socio-sémantique" permettent aux communautés de coproduire et d'utiliser des artefacts organisationnels symboliques, tels que des "cartes", des "nuages de tags" ou des index partagés rendant les connaissances et les actions/activités collectives à la fois plus visibles et plus réflexives." [132]

Selon Gruber et al [133], le web sémantique permet aux machines de collecter et de structurer les données afin de leur donner une signification claire et globale. Tandis que, ces données peuvent être facilement partagées entre différents sys-

tèmes via le web social. Gruber met l'accent sur l'importance du web sémantique social dans la phase d'extraction des données personnelles sur les utilisateurs des systèmes. Gruber illustre ces concepts sous forme des exemples de systèmes intitulés "collective knowledge systems".

2.3 Le web sémantique social et L'e-learning

Depuis l'émergence du web sémantique, un ensemble des chercheurs ont essayé de le donner place dans les plateformes e-learning. En 2002, Nilsson et al [134] ont proposé une conception d'application d'e-learning à l'aide des technologies web sémantique. Dans leur article, ils ont montré que l'utilisation de métadonnée est essentielle pour décrire un contexte d'apprentissage compréhensible et significatif pour l'apprenant. Ainsi, Les idées et les techniques du web sémantique social ont également été rapidement utilisées dans le domaine de l'apprentissage en ligne. De cette manière, Gruber [127] aborde les avantages du web social sémantique dans le domaine de l'e-learning. Selon lui, le web sémantique social permet d'obtenir une grande quantité de données.

Par ailleurs, Torniai et al [131] ont souligné que le web sémantique n'est pas encore largement utilisé malgré ses aspects prometteurs. Cela peut être dû, entre autres, à la difficulté de créer et de maintenir une bonne ontologie prenant en charge le processus d'ajout d'annotations sémantiques aux données. Tous les chercheurs tendent à soulignent que le web sémantique social est un morceau essentiel dans le développement l'e-learning, où Leur vision s'articule comme "E-learning based on Social Web Semantic", dans lequel les activités d'apprentissage axés sur la pédagogie seraient faciles à créer, à partager et à utiliser par les apprenants et les enseignants, sans avoir besoin des connaissances approfondies ou des compréhensions avancées des technologies.

Plus les étudiants sont attachés à leurs appareils mobiles, plus ils demandent d'avoir accès à tous les types d'information à tout moment. Les contraintes des longues distances pour le cas des étudiants habitant loin des établissements d'enseignement, constituent un obstacle devant l'obtention des informations relatives aux parcours de formation. La consultation des notes, des emplois de temps ou encore la demande des pièces administratives sont des services importants à rap-

procher à l'étudiant se trouvant à n'importe quel emplacement.

Partant de ces points, nous avons constaté que les plateformes x-learning offrent des services basiques par rapport aux attentes des utilisateurs. D'où l'intégration des nouveaux services permettant un rapprochement des services aux étudiants s'avère nécessaire. L'objectif d'obtenir une plateforme complète ne peut être atteint qu'après une vraie réflexion autour des nouveaux services.

3 Les ontologies

Le terme ontologie (du grec : Doctrine de l'être) vient à l'origine de la philosophie désignant une discipline qui traite de la classification des types de base des entités du domaine et de leurs relations structurelles. Cette branche philosophique est, également, largement connue sous le nom de "métaphysique générale" [134]. Le terme 'ontologie' remonte à un recueil de textes d'Aristote appelé Métaphysique se représentant comme l'un des premiers à entrer en contact avec ce domaine. Par ailleurs, le terme a été utilisé dans le domaine de l'informatique pour désigner un modèle formel de structure de connaissances et ses entités et relations associées [135]. Dans un autre article scientifique de Tom Gruber, l'ontologie est présentée en tant qu'elle est utilisée dans le domaine de l'informatique, par le fait qu'elle est définie comme une "spécification explicite et formelle d'une conceptualisation" [136]. Dans ce contexte, le terme conceptualisation est défini en tant qu'une tentative de représentation simple et abstraite de la réalité. Selon Gruber, une ontologie est définie par un ensemble de concepts et de relations représentant les objets et les relations entre eux dans le domaine considéré, ainsi que par des axiomes formels qui déterminent l'interprétation et l'utilisation correcte de ces concepts. Guarino et Giaretta en 1995 [135] ont remis en question cette définition de Gruber, plus précisément, le volet portant sur la signification de la notion de conceptualisation. Selon Gruber, ce dernier concept ne représente qu'une image d'une seule situation d'objet. Cependant, Guarino et Giaretta montrent qu'une conceptualisation est mieux définie comme une sorte de grille conceptuelle qui peut être appliquée à de nombreuses situations. A ce propos, dans un exemple illustré à la figure 26, deux couches d'objets sont représentées. Toutes ces deux avec les objets a, b, c, d, e. Dans la première définition de Gruber, la table de gauche et la table de droite montrent deux conceptualisations différentes avec les

mêmes objets.

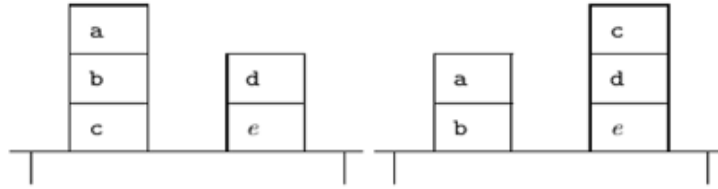


FIGURE 26 : Un exemple à partir d'un monde en blocs, avec les objets a,b,c,d,e dans deux couches d'objets différentes [135]

Selon la définition de Guarino et Giaretta, on parle de la même conceptualisation appliquée à des situations différentes, où Guarino, à la base de cette explication, a déduit sa propre définition d'une ontologie en tant que :

" (...) une théorie logique rendant compte du sens voulu d'un vocabulaire formel, c'est-à-dire de son engagement ontologique envers une conceptualisation particulière du monde. Les modèles prévus d'un langage logique utilisant un tel vocabulaire sont contraints par son engagement ontologique. Une ontologie reflète indirectement cet engagement (et la conceptualisation sous-jacente) en se rapprochant de ces modèles prévus." [135]

En général, Une ontologie représente un ensemble de concepts abstraits symbolisés avec des relations sémantiques. Elle est donc un domaine d'une tentative de formalisation complète et détaillée d'un certain volet de connaissances au moyen d'un schéma conceptuel. Ainsi un tel schéma, habituellement, consiste en une structure de données hiérarchique contenant toutes les classes d'objets pertinentes et leurs relations ainsi que les règles (contraintes) acceptées dans ce domaine. Dans ce sens, en 1993, La première définition de l'ontologie dans le domaine de l'éducation a été présentée par Tom Gruber. Cette définition est considérée comme la pierre angulaire de la recherche dans ce domaine, car elle décrit l'ontologie comme une spécification formelle d'une conceptualisation dans un contexte donné. Cette conceptualisation est une description des concepts, ainsi que de toutes les informations liées à ces concepts (telles que les propriétés, les relations, les contraintes, les axiomes et les assertions), nécessaires pour décrire et résoudre les problèmes dans le domaine choisi. En d'autres termes, le modèle conceptuel est un système de concepts et de relations d'un domaine, et cette

approche est essentielle pour comprendre la complexité des problèmes dans ce domaine. Un concept est une notion reflétant un objet concret ou abstrait du monde réel. En qualité de cause formelle, l'ontologie se compose d'un ensemble fini de concepts (notions) du domaine X , d'un ensemble fini de relations entre concepts R et d'un ensemble fini de fonctions d'interprétation définies sur les concepts et/ou les relations F ; $O = \langle X, R, F \rangle$. Lorsque $R=0$ et $F=0$, l'ontologie est transformée en un simple dictionnaire.

3.1 Structure d'une ontologie

En terme général, la structure d'une ontologie est un ensemble d'éléments, qui se compose de quatre catégories : concepts (ou classes), relations (ou propriétés, attributs), les axiomes, et les instances [137].

3.1.1 Les concepts

Dans une ontologie, un concept (ou une classe) représente une catégorie générale qui peut être utilisée pour regrouper des entités individuelles ayant des propriétés communes. Par exemple, dans une ontologie médicale, le concept de "maladie" peut être défini comme une classe qui inclut toutes les maladies, tous les troubles et tous les syndromes connus de la science médicale. Chaque maladie individuelle (par exemple, le diabète, le cancer ou la maladie d'Alzheimer) peut alors être considérée comme une instance de la classe "Maladie".

Les concepts sont souvent organisés de manière hiérarchique dans une ontologie, avec des catégories plus générales en haut et des catégories plus spécifiques en bas. Par exemple, la classe "Maladie" peut être une sous-classe de la classe plus générale "Condition médicale", qui à son tour peut être une sous-classe de la classe encore plus générale "Santé et médecine". Cette organisation hiérarchique permet un raisonnement plus précis et plus efficace sur les relations entre les différentes catégories d'entités.

Chaque concept est généralement défini par un ensemble d'attributs qui décrivent ses propriétés et ses caractéristiques. Par exemple, la classe "Maladie" peut avoir des attributs tels que "Symptômes", "Traitement" et "Pronostic", qui décrivent les propriétés communes de toutes les maladies de la classe. Ces at-

tributs peuvent ensuite être utilisés pour aider à distinguer une maladie d'une autre, et pour identifier les relations entre différentes maladies sur la base de leurs propriétés communes ou divergentes.

3.1.2 Les relations

Dans une ontologie, les concepts peuvent être reliés entre eux par différents types de relations, qui décrivent la nature du lien qui les unit. L'une des relations les plus utilisées dans toutes les ontologies est la relation de catégorisation, également connue sous le nom de relation taxonomique, classe-sous-classe, générique ou hyponyme-hyperonyme.

La relation de catégorisation est utilisée pour lier un concept plus général (la superclasse ou l'hyperonyme) à un ou plusieurs concepts plus spécifiques (les sous-classes ou les hyponymes) qui lui appartiennent. Par exemple, dans une ontologie biologique, le concept "Animal" peut être une superclasse qui est liée à des sous-classes plus spécifiques telles que "Mammifère", "Oiseau" et "Poisson". Chacune de ces sous-classes peut à son tour être liée à des sous-classes encore plus spécifiques, telles que "Primate" sous "Mammifère".

Un autre exemple de la relation de catégorisation peut être observé dans une ontologie musicale, où le concept "Instrument" peut être une superclasse liée à des sous-classes plus spécifiques telles que "Percussion", "Corde" et "Cuivre". Chacune de ces sous-classes peut à son tour être liée à des sous-classes encore plus spécifiques, telles que "Drum" sous "Percussion".

En utilisant la relation de catégorisation, les ontologies peuvent être structurées de manière hiérarchique, avec des concepts plus généraux au sommet et des concepts plus spécifiques à la base. Cela permet un raisonnement efficace sur les relations entre les différentes classes, et facilite la recherche d'informations sur les entités d'une catégorie particulière.

3.1.3 Les axiomes

Dans une ontologie, les axiomes sont utilisés pour exprimer des déclarations explicites qui relient les concepts et les relations. Un axiome peut être compris comme une déclaration introduite dans une ontologie qui sert de point de départ à

partir duquel d'autres déclarations peuvent être dérivées. Les axiomes définissent les conditions de mise en relation des catégories et des relations, et permettent d'effectuer d'autres déductions au sein de l'ontologie. Par exemple, considérons une ontologie géographique qui inclut les concepts "Pays" et "Ville", et la relation "se-trouve", qui décrit la relation entre une ville et le pays dans lequel elle est située. Un axiome pourrait être introduit pour spécifier qu'une ville doit être située dans exactement un pays. Cet axiome fournirait une contrainte sur la relation "se-trouve", permettant de déduire automatiquement le pays dans lequel une ville est située sur la base de l'axiome.

Les axiomes peuvent également être utilisés pour définir des relations plus complexes entre les concepts. Par exemple, considérons une ontologie qui comprend les concepts "Personne" et "Parent", et la relation "à-un-enfant", qui décrit la relation entre une personne et son enfant. Un axiome pourrait être introduit pour spécifier que si la personne A est le parent de la personne B, et que la personne B est le parent de la personne C, alors la personne A est également le grand-père de la personne C. Cet axiome permettrait la dérivation automatique de nouvelles informations sur les relations entre les différentes entités de l'ontologie.

3.1.4 Les instances

Dans une ontologie, les instances représentent les membres individuels d'une classe d'entités ou de phénomènes. Les instances sont des éléments spécifiques d'une catégorie et peuvent être désignées par divers noms tels que ; instances concrètes, individus ou instances individuelles.

Par exemple, considérons une ontologie dans le domaine des livres, qui comprend le concept "Livre". L'ontologie pourrait avoir de multiples instances du concept "Livre", comme "To Kill a Mockingbird" de Harper Lee, "1984" de George Orwell, et "The Catcher in the Rye" de J.D. Salinger. Chacune de ces instances possède un ensemble unique de propriétés, telles que l'auteur, la date de publication et le genre.

Les instances sont importantes dans une ontologie car elles représentent des entités ou des phénomènes du monde réel auxquels on peut se référer et qu'on peut analyser. Elles fournissent des exemples concrets des concepts définis dans

l'ontologie et peuvent être utilisées pour démontrer les relations entre différents concepts et instances.

Les instances peuvent également être utilisées pour générer de nouvelles connaissances et de nouvelles idées au sein d'une ontologie. Par exemple, si une ontologie comprend les concepts "Animal" et "Oiseau", et qu'une instance d'un oiseau est ajoutée à l'ontologie avec la propriété "Peut voler", on peut en déduire que tous les oiseaux ont la propriété "Peut voler". Cela permet de générer de nouvelles connaissances au sein de l'ontologie et de faire des déductions sur la base des propriétés et des relations entre les différentes instances.

3.2 Langages de description d'ontologie

Un langage de description d'ontologie est un langage formel utilisé pour coder une ontologie, qui peut être utilisée pour représenter et raisonner sur les connaissances dans un domaine. Il existe plusieurs langages de description d'ontologie, chacun ayant ses propres forces et faiblesses. Certains des langages de description d'ontologie les plus utilisés sont OWL, KIF, CL et CycL [138].

- OWL (Web Ontology Language) : OWL est un langage de description d'ontologie basé sur RDF (Resource Description Framework) et est utilisé pour représenter des ontologies sur le Web. Il fournit un ensemble riche de primitives pour définir des concepts, des relations et des axiomes logiques. OWL est utilisé dans de nombreuses applications, telles que les systèmes de recommandation, la recherche sémantique, l'intelligence artificielle, etc [139]
- KIF (Knowledge Interchange Format) : KIF est un langage de description d'ontologie basé sur la logique de premier ordre. Il fournit un ensemble de syntaxes pour décrire des concepts, des relations, des axiomes logiques, des règles d'inférence, etc. KIF est utilisé dans de nombreuses applications, telles que la représentation des connaissances, la vérification des spécifications, la planification, etc [140].
- CL (Common Logic) : CL est un langage de description d'ontologie basé sur la logique de premier ordre. Il fournit un ensemble de syntaxes pour décrire des concepts, des relations, des axiomes logiques, des règles d'inférence, etc.

CL est utilisé dans de nombreuses applications, telles que la représentation des connaissances, la vérification des spécifications, la planification, etc.

- CycL (Cyc Language) : CycL est un langage de description d'ontologie basé sur la logique du premier ordre. Il fournit un ensemble de syntaxes pour décrire des concepts, des relations, des axiomes logiques, des règles d'inférence, etc. CycL est utilisé dans de nombreuses applications, telles que la représentation des connaissances, la vérification des spécifications, la planification, etc.

La représentation ontologique des domaines d'objets peut être présentée différemment selon le langage de description d'ontologie utilisé et les technologies utilisées pour travailler avec les ontologies. Le même domaine peut être représenté de différentes manières, et la question de savoir si le domaine est bien représenté et s'il dessine le modèle de réalité souhaité est une décision subjective.

3.3 Travaux connexes sur les systèmes de recommandation basé sur l'ontologie dans l'E-learning

Les systèmes de recommandation pour l'apprentissage en ligne jouent un rôle essentiel dans la création de contenus pertinents à partir des nombreuses informations disponibles sur Internet. Diverses approches ont été utilisées pour développer des systèmes de recommandation, telles que le filtrage basé sur le contenu, le filtrage collaboratif et le filtrage hybride qui surmonte les inconvénients d'une approche individuelle. Récemment, de nombreux chercheurs ont développé des systèmes de recommandation E-learning basés sur l'ontologie, qui représente la connaissance dans le domaine de l'éducation. Par exemple, Mohammed Ibrahim et al [141] ont développé un système de recommandation de cours pour les étudiants universitaires, qui suggère les meilleurs cours répondant à leurs besoins en fonction de l'ontologie. Dans cette étude, les auteurs représentent les connaissances sur le profil des étudiants, le contenu des cours et le domaine d'apprentissage avec des ontologies statiques. Dans [115], les auteurs proposent un nouveau système de recommandation basé sur la connaissance utilisant l'ontologie pour modéliser la connaissance du domaine du profil de l'apprenant et des ressources d'apprentissage, le système adopté dans ce travail a appliqué SPM (Sequential Pattern Mining) avec la méthode de filtrage collaboratif pour générer les N meilleures

recommandations. Une autre approche hybride de recommandation, basée sur l'ontologie, a été proposée par les auteurs dans [142], cette approche combine le filtrage collaboratif et le filtrage basé sur le contenu. Soutenu par la similitude d'ontologie entre les profils des éléments et les profils des utilisateurs, Bouihi et Bahaj dans [143], ont utilisé l'ontologie pour représenter la connaissance du contenu et du contexte, et explorent les règles SWRL (Semantic Web Rule Language) comme technique de recommandation basée sur la pertinence et la pondération des objets d'apprentissage. Cette méthode aide les apprenants à obtenir le matériel d'apprentissage approprié pour avoir une expérience d'apprentissage réussie. L'ontologie peut également être utilisée afin de structurer la représentation sémantique du profil Mooc et fournir par la suite des recommandations personnalisées pour chaque apprenant. En 2018, Kahina et Latifa [144] ont proposé un système de recommandation Mooc basé sur la combinaison du filtrage collaboratif et de l'ontologie. Dans ce travail, les auteurs utilisent les connaissances ontologiques pour atténuer le problème du démarrage à froid et intégrer les propriétés de l'apprenant et du Mooc dans le processus de recommandation. Les auteurs dans [145] ont développé un système de recommandation personnalisé basé sur l'ontologie et le filtrage collaboratif pour recommander des contenus d'apprentissage appropriés et utiles aux apprenants. Ce système surmonte le problème de la personnalisation et suggère des éléments pertinents en fonction de la compétence de l'apprenant, de son niveau et de ses connaissances. Cependant, tous les travaux connexes mentionnés ci-dessus souffrent des problèmes de démarrage à froid et d'éparpillement (Sparsity), qui sont des inconvénients bien connus dans le domaine des systèmes de recommandation en E-learning, car ils ne prennent pas en compte les changements de style d'apprentissage et le niveau de l'apprenant. Pour cette raison, nous avons pensé à l'ontologie dynamique dans le contexte de la recommandation. De plus, dans un environnement dynamique, toutes les informations sur les apprenants seront mises à jour et stockées dans l'ontologie de l'apprenant. Ainsi, elles seront intégrées dans notre système de recommandation pour fournir les matériaux d'apprentissage les plus appropriés aux apprenants. En outre, avec l'ontologie dynamique, le système de recommandation peut atténuer les problèmes de démarrage à froid et d'éparpillement à l'aide de l'énorme quantité d'informations collectées à partir de multiples ressources. Cette proposition peut améliorer, à la fois, la performance et la qualité des recomman-

dations, et produire de meilleurs résultats par rapport aux autres approches.

L'E-learning et la structuration des connaissances (ontologie) ont un lien natif, étant donné que l'E-learning se base essentiellement sur la création, la transmission et la réception des connaissances, l'ontologie joue un rôle décisif dans l'organisation, la structuration et la mise en disposition des liens logiques entre ces connaissances créées. La création, l'organisation et le stockage des contenus d'apprentissage sont l'objectif principal derrière l'utilisation de l'ontologie. L'objectif est de rendre la création aussi simple que possible pour l'utilisateur, de sorte qu'il n'ait pas besoin de connaître les techniques informatiques avancées. Souvent, dans le domaine de l'E-learning, le contenu de l'ontologie s'adapte à une plate-forme d'apprentissage spécifique, de sorte que le contenu ne peut être créé que pour ce système. Cela est dû aux différents modèles de contenu sur lesquels reposent les plates-formes d'apprentissage. Cependant, il est également possible d'utiliser le contenu sur plusieurs systèmes et de l'échanger entre les plates-formes d'apprentissage en utilisant le même modèle de référence pour le contenu sur ces systèmes. Un tel modèle est proposé par SCORM (Sharable Content objet référence Model), qui offre un ensemble de normes et de spécifications. Il définit notamment la manière dont les contenus d'apprentissage doivent être structurés et échangés entre les systèmes. En E-learning une ontologie combine ces deux concepts et offre, en plus de la gestion de contenu via des dispositifs informatiques, une estrade d'apprentissage dans laquelle les systèmes peuvent utiliser ces différents modules d'apprentissage sous forme d'objets pour de plus grandes unités pédagogiques des cours, des présentations ou de la documentation.

4 Approche proposée

Dans le cadre du développement d'un système d'apprentissage dynamique basé sur une ontologie qui prendra en charge l'échange personnalisé et adaptatif du contenu d'apprentissage, les informations cognitives des apprenants collectées sous forme d'attributs via le Web sémantique seront un atout très important pour mener à bien cette action. L'ontologie de l'e-learning que nous avons proposé pour achever cet objectif est conçue comme un ensemble de ressources ontologiques homogènes qui ont été identifiées pour être utilisées dans multiples environnements d'apprentissage (figure 27). Cette ontologie pourrait être expliquée à partir de

deux perspectives, une perspective modulaire (figure 28) et une perspective en couches qui approfondit la mise en œuvre de la vue modulaire (figure 29). La perspective modulaire fournit une couverture détaillée de l'architecture nommée ERSDO (E-learning Recommender System based on Dynamic Ontology). Cette architecture contient les éléments fonctionnels du système selon des étapes consécutives. Afin de modéliser complètement un système basé sur l'ontologie, tout le domaine (y compris l'apprenant, enseignant, objets d'apprentissage, évaluations) ont été modélisé à un niveau abstrait de granularité. Chacun de ces aspects fait l'objet d'une modélisation détaillée sous forme d'ontologies.

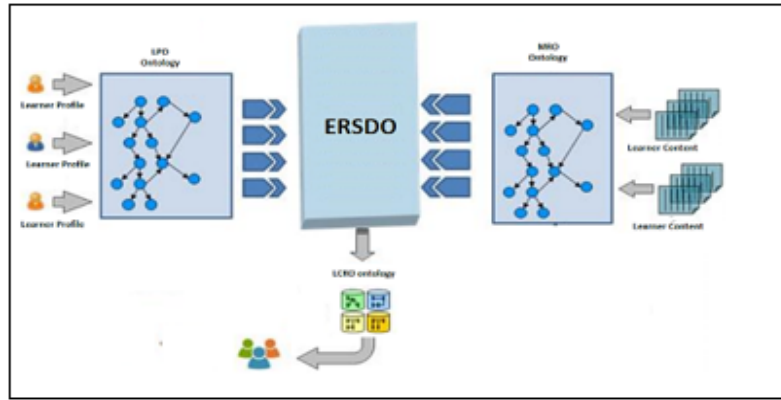


FIGURE 27 : les ontologies utilisées du système proposé [124]

L'objectif de cette thèse de doctorat est d'utiliser l'ontologie du profil de l'apprenant pour améliorer la personnalisation des recommandations de contenu. Pour cela, nous avons rassemblé et traité les attributs de l'apprenant en utilisant des techniques de web sémantique afin d'en sélectionner les plus significatifs pour améliorer la performance de l'apprenant. Nous avons ensuite classé les apprenants en fonction de leurs capacités d'apprentissage, leur style d'apprentissage, leur score dans le test de positionnement et leur test de connaissance. Ces catégories sont utilisées pour fournir des recommandations de contenu personnalisées et adaptatives.

4.1 Perspective modulaire du système proposé

Cette approche propose un modèle d'apprentissage constructiviste, dans lequel les apprenants ont le rôle principal dans leur processus d'apprentissage.

Notre proposition combine : L'ontologie du profil de l'apprenant (Learner Profile Ontology : LPO) comme source principale des informations sur les apprenants, l'ontologie des ressources matérielles (Material Resource Ontology : MRO) qui sera utilisée par les apprenants dans leurs processus d'apprentissage et finalement l'ontologie de relation apprenant-cours (Learner-Course Relation Ontology : LCRO) pour combiner les apprenants avec les supports de cours. L'ontologie du profil de l'apprenant (LPO) expose toutes les informations pertinentes sur l'apprenant, notamment des données personnelles (nom complet, sexe, coordonnées, nom d'utilisateur, mot de passe, etc...), les informations académiques de l'apprenant, ainsi que ses prérequis. Toutes ces informations seront extraites via les informations collectées par le web sémantique social. L'ontologie de ressource matérielle (MRO) fournit des informations sur le matériel de cours ou le support utilisé par LMS dans la phase d'apprentissage comme : (type de cours, nom du cours, domaine du cours, ID du cours...), notre système utilise SKOS (simple knowledge organization system) [146] pour identifier le vocabulaire du domaine et la norme OWL pour représenter le contenu pédagogique. Notre proposition de personnalisation et d'adaptation des contenus est basée sur les caractéristiques cognitives de chaque apprenant. Pour cela, nous avons utilisé la modélisation de contenu d'apprentissage basé sur SWS (Social Web Semantic) qui utilise des algorithmes de web sémantique pour collecter des informations sur l'utilisateur. Ces informations sont ensuite modélisées selon les paramètres de l'ontologie LPO.

Nous avons également utilisé l'ontologie MRO pour réorganiser le contenu des ressources utilisées par les gestionnaires de l'application et le contenu du cours en fonction du niveau de difficulté. Enfin, nous avons utilisé l'ontologie LCRO pour adapter le contenu des ressources en fonction des performances de l'apprenant.

Grâce à LCRO, l'apprenant suit un cycle d'apprentissage qui consiste à réaliser une activité d'apprentissage, puis évaluer et enregistrer les scores pour déterminer la pertinence du contenu d'apprentissage. Nous utilisons les profils d'apprenants modélisés dans LPO pour diffuser du contenu d'apprentissage personnalisé. La section 4.2 détaille les concepts utilisés pour construire ces profils d'apprenants dans LPO.

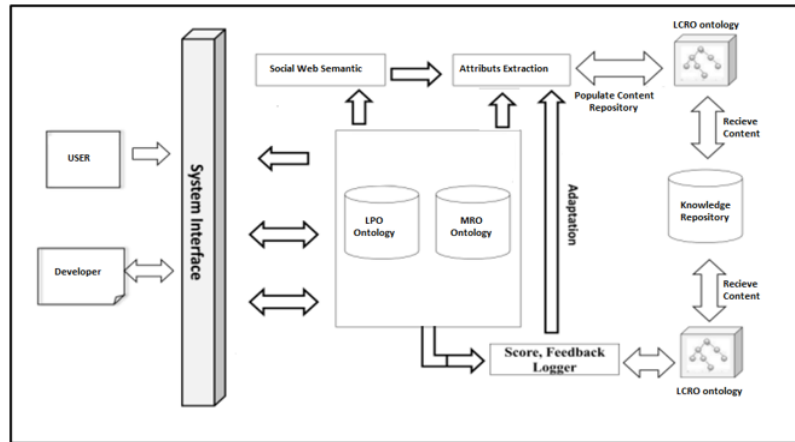


FIGURE 28 : Perspective modulaire du système proposé [124]

4.2 Perspective en couche du système proposé

La perspective en couches, comme illustré dans la figure 29, donne un aperçu de la mise en œuvre de l'architecture proposée dans les couches technologiques actuelles et les stratégies de développement utilisées pour la diffusion efficace de contenus d'apprentissage pour faciliter la tâche des apprenants.

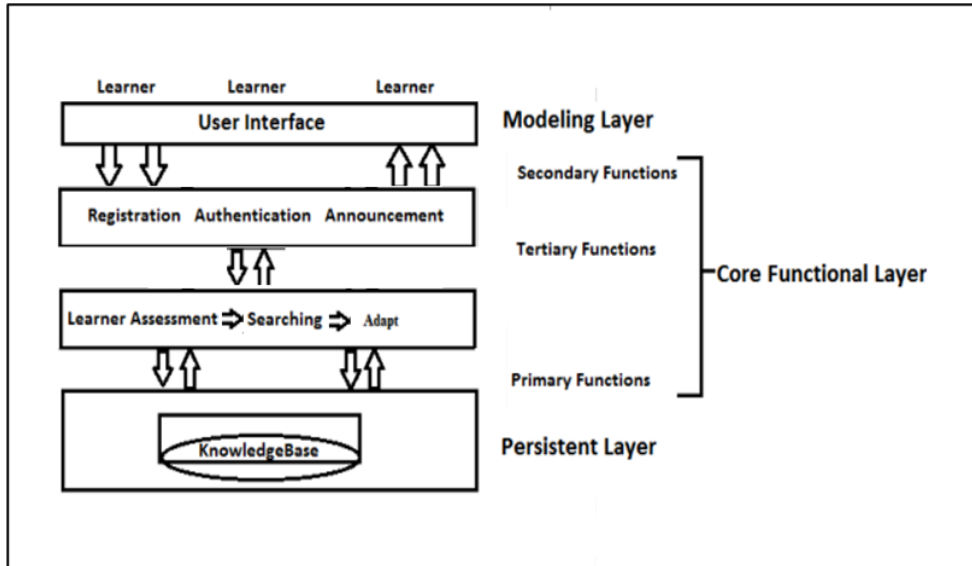


FIGURE 29 : Perspective en couche du système proposé [124]

L'environnement proposé comporte trois couches principales, à savoir (1) la couche de modélisation, (2) la couche fonctionnelle de base et (3) la couche persistante. Chacune de ces couches est subdivisée en "unité" qui facilite la division,

la gestion et la maintenance des cycles d'apprentissage. Chacune de ces couches est élaborées dans ce qui suit :

Couche de modélisation : cette couche fournit un espace d'interaction aux apprenants ainsi que pour les enseignants afin d'effectuer différentes tâches telles que l'inscription au cours, la visualisation du contenu et la gestion du cours. Du point de vue technologique, l'interface comprend une combinaison de pages HTML/JSP interagissant avec la couche fonctionnelle centrale par le biais de Servlets Java qui, à son tour, effectue des opérations de gestion des données.

Couche fonctionnelle : cette couche couvre la partie principale de la mise en œuvre du système proposé. Du point de vue technologique, cette couche fonctionne avec le modèle de conception DTO (Data Transfer Object) basé sur la technologie SPARQL (Protocol and RDF Query Language). La couche fonctionnelle incorpore trois unités :

- Unité des fonctions primaires : cette unité comprend deux unités de fonctions primaires, à savoir la phase d'enregistrement et l'authentification. La phase d'enregistrement permet aux utilisateurs de créer un compte ou de s'inscrire sur le système, tandis que l'authentification vérifie l'identité des apprenants lorsqu'ils se connectent au système.
- Unité des fonctions secondaires : cette unité est responsable de la gestion des caractéristiques/profils des apprenants, comme par exemple le score, le Feedback, la traçabilité du fichier Log. Cette unité comprend les exigences fonctionnelles ayant un rôle clé dans l'efficacité du système par le biais des aspects de personnalisation et adaptabilité. L'adaptabilité et la personnalisation réalisées par la catégorisation des concepts de profil de l'apprenant dans la section "LPO".
- Couche persistante : cette couche comprend des objets pour le stockage des instances ontologiques, telles que JENA (JENA un logiciel open source pour le web Sémantique), utilisé pour le stockage des contenus d'apprentissage et des profils des apprenants, etc. JDBC, utilisé pour se connecter à la couche persistante, et SPARQL pour la manipulation des données, c'est-à-dire la récupération, la création, la mise à jour et la suppression des informations vers et depuis les ontologies créées. Chaque module comprend

des classes correspondront à des unités fonctionnelles. La communication entre les différents modules se fait par le biais de la programmation appelée VOs (Virtual Objects) ou DTO (Data Transfer Objects). Afin de gérer ce code, on a utilisé le modèle de conception DAO (Data Access Object) qui fonctionne en trois couches, à savoir l'interface, la couche logique et la couche d'accès aux données pour l'interaction avec la couche de base de données.

Après avoir compris l'architecture du système proposé, la première étape consiste à identifier les attributs de l'apprenant pour une catégorisation appropriée de sa nature. La section suivante « L'ontologie du profil de l'apprenant (Learner Profile Ontology : LPO) » présente de manière exhaustive les attributs utilisés dans notre système.

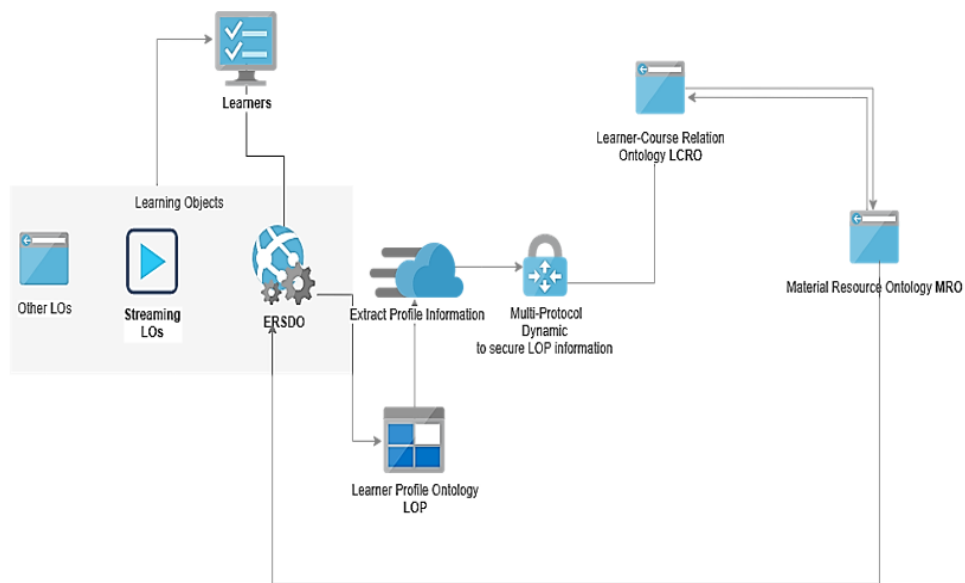


FIGURE 30 : Architecture du système de recommandation e-learning basé sur l'ontologie dynamique (ERSDO) [124]

4.3 L'ontologie du profil de l'apprenant (Learner Profile Ontology : LPO)

La première étape consiste à établir le profil des apprenants dans une ontologie afin de bénéficier de la technologie du web 3.0. Dans cette optique, deux

aspects seront utilisés, à savoir le format dans lequel ces attributs de profil sont stockés et la qualité de chaque attribut de profil. Et puisque notre objectif s'attire plus vers la construction d'une ontologie dynamique, nous avons opté pour le web sémantique, qui nous donne la possibilité de la mise à jour dynamique des profils grâce à la puissance de ses algorithmes en matière de la modification et de la mise à jour. Dans l'ERSDO, l'apprenant se caractérise par une forte liaison avec l'ontologie LPO. Les performances des techniques d'apprentissage automatique utilisées pour la catégorisation de ces profils dépendent largement de la qualité de l'ensemble de données utilisées. A cet égard, le web sémantique social apparaît comme une solution efficace qui poursuit la traçabilité quotidienne des apprenants dans les différents domaines du web. Comme mentionné précédemment, le web sémantique social contient de grandes quantités de données fournies par les utilisateurs sous forme de texte, de statut, commentaires, hashtags, etc...

Grâce aux algorithmes de web sémantique social, les apprenants sont catégorisés sous forme des attributs bien structurés en fonction de leur profil et de leurs différentes activités dans le web. Par exemple, un commentaire bien rédigé en anglais sur un sujet dans les réseaux sociaux, pourrait être considéré comme un barème du jugement linguistique et sémantique. Bien entendu, si le commentaire a eu lieu sur un sujet technique, l'apprenant aura deux attributs, l'un sur la qualité de la langue utilisée, et l'autre sur la pertinence de l'idée partagée avec les internautes. La modélisation du profil de l'apprenant consiste à acquérir les attributs des apprenants à partir de différentes sources qui sont utilisées pour catégoriser les profils en fonction de leur niveau d'expertise et de leurs compétences. Ces catégories d'apprenants sont utilisées pour recommander, par la suite, des contenus d'apprentissage à l'apprenant.

4.3.1 Attributs de learner Profil (l'apprenant)

Les attributs de l'apprenant contribuent efficacement dans le processus de la catégorisation des apprenants, cela est bien démontré lors de l'expérimentation. Le processus d'élaboration du profil de l'apprenant et les techniques d'acquisition de ces attributs sont présentés ci-dessous par ordre d'importance :

- **Test de positionnement** : le test de positionnement fournit une base pour prédire la capacité d'un individu à acquérir certaines connaissances

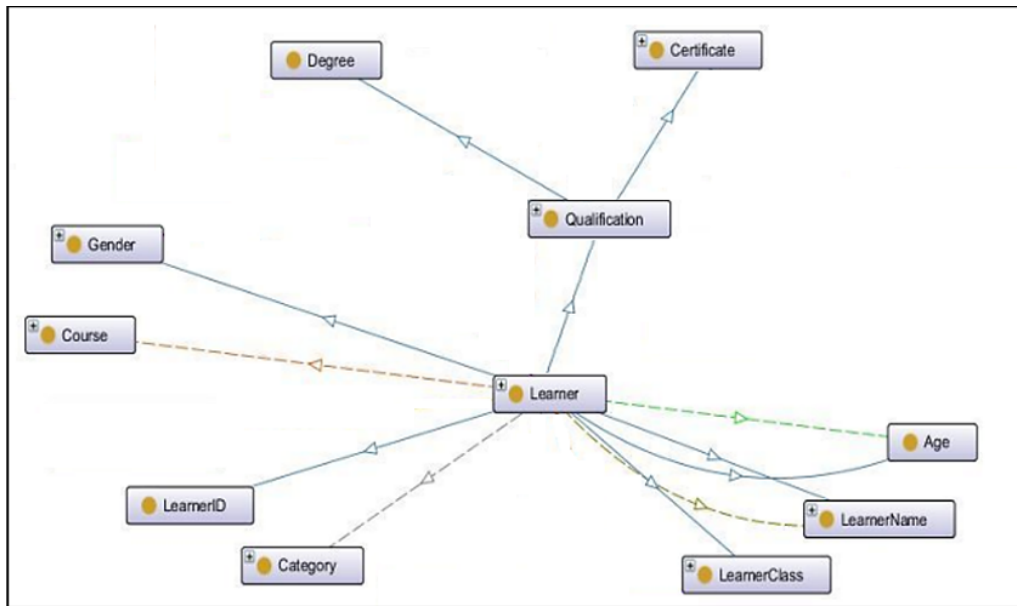


FIGURE 31 : Ontologie de l'apprenant (LPO)

et compétences. Il permet également de prédire le potentiel d'un individu grâce aux résultats obtenus. Chaque test couvre différents domaines tels que l'aptitude numérique, les compétences analytiques, le raisonnement cognitif et le raisonnement verbal.

- **Pré-test** : le pré-test a été développé pour les apprenants afin d'identifier leurs niveaux en tenant compte de leurs connaissances préalables et de leur score.
- **Style d'apprentissage** : le style d'apprentissage est une démarche permettant d'apprécier la méthode pédagogique adaptée à l'apprenant et qui lui convient. Ce jugement repose sur la description de l'attitude et du comportement d'apprentissage selon des règles bien précises. Par exemple, le modèle de Felder-Silverman [147] utilisé pour mesurer le style d'apprentissage de l'apprenant selon quatre dimensions qu'on peut les présenter comme suit :

1. Actif/réfléchi
2. Sensationnel/intuitif
3. Visuel/verbal

4. Séquentiel/global

- **Questionnaire** : Les apprenants reçoivent un questionnaire qui leur permet d'évaluer leur comportement d'apprentissage. Chaque apprenant reçoit une série de questions basées sur des pratiques standards pour mesurer le style de l'apprenant. Cela a permis d'avoir un aperçu de l'instinct de l'apprenant pour identifier le moyen d'enseignement le plus adapté ainsi que son impact en termes de performance.
- **Qualification** : la qualification de l'apprenant s'intervient comme mesure permettant de décider des compétences cognitives de l'apprenant. Chose donnant liberté à examiner si l'apprenant a suivi aux cours des niveaux précédents.
- **L'âge** : l'âge de l'apprenant est également un facteur important pour déterminer le degré de sa capacité à saisir les idées et les concepts. Ce volet est utilisé pour classer si les apprenants ont un apprentissage lent ou rapide.
- **La localité/l'origine** : la localité des apprenants, s'est également révélée être un facteur ayant un impact sur la performance académique de l'apprenant et l'obtention de son diplôme. Chose poussant vers la réflexion selon laquelle, chacune des villes est étiquetée avec un niveau particulier (villes métropolitaines, semi-urbaines et sous-développées).
- **Compétences** : est un attribut permettant de savoir les centres d'intérêts de l'apprenant ainsi que ses qualités linguistiques.

En vue de la quantification de ces attributs, on va utiliser deux indices, le percentile et l'indice de Stanine.

Le percentile est le pourcentage de candidats qui se situent en dessous d'un score brut particulier. Par exemple, un percentile de 65 signifie que l'apprenant à un meilleur score par rapport au pourcentage 65 de la population de sa catégorie.

Stanine est une série de notes représentée sur une échelle de 0 à 9, cette dernière a été adoptée comme une mesure pour évaluer et classer l'apprenant en fonction du score obtenu. Les rubriques du modèle Stanine commencent par un

niveau très bas (V. Low), inférieure à la moyenne (Below Avg) en tant que la moyenne supérieure à la moyenne (Above Avg) est très élevée (V. High).

	V. Low	Below Avg.		Average			Above Avg.		V. High
Stanine	1	2	3	4	5	6	7	8	9
Percentile	<= 4	5-11	12-22	23-40	41-58	59-77	78-89	90-95	>96

TABLE 6 le percentile et l'indice de Stanine.

4.3.2 Jeu de données du profil des apprenants

Une fois que les attributs du profil des apprenants sont définis autant qu'ils sont indiqués dans la section précédente, ces attributs du profil de l'apprenant seront modélisés dans une ontologie appelée LPO. Dans cette tendance ontologique, chaque attribut sera défini comme une classe caractérisée par les propriétés de cette dernière et qui seront présentées en utilisant OWL. La représentation du profil de l'apprenant sous la forme d'une base de connaissances ontologique peut fournir une représentation optimale de l'information avec une cohérence entre les concepts en réduisant la redondance avec une capacité de recherche d'information intelligente. La figure 8 présente un extrait graphique de notre instance d'ontologie correspondant aux concepts, aux propriétés et aux individus.

4.3.3 Prétraitement des attributs de l'apprenant

Tous les attributs de profils des apprenants ne sont pas forcément expressifs et significatifs dans la catégorisation des apprenants. Ces attributs peuvent ne pas contribuer à la précision de la classification et par la suite provoquent un sur-ajustement. Par ailleurs, un mécanisme de prétraitement a été utilisé pour éliminer les attributs les moins importants de l'ensemble de données.

Le prétraitement de ces enregistrements a été effectué en fonction des métadonnées des attributs comprenant les valeurs des données avec leurs catégories correspondantes. Les caractéristiques ayant un impact maximal sur la classification d'un apprenant, dans une certaine catégorie, seront suggérées dans cette

phase. Le nombre de caractéristiques présentées et celles qui ont été sélectionnées dans le tableau suivant.

Type d'éléments	Ensemble de caractéristiques liées	Caractéristiques sélectionnées par WEKA
Académique	Course, research, experience, preTestScore, class, project, PreReqTest.	Course, PreTestScore, PreReqTest
Démographique	LearnerID, name, age, qualification, region	LearnerID, name, age
Comportemental	LearnerStyle, LearnerAptitude, capacity	LearnerStyle, LearnerAptitude
Spécialisation	LearnerThesis, learnerSupervisor, area of research, publication	LearnerThesis, publication

TABLE 7 Le nombre de caractéristiques présentées et celles qui ont été sélectionnées par WEKA

Les attributs suggérés après le prétraitement des données sont représentés numériquement sur une échelle de 1 à 10. Cette représentation numérique offre deux avantages : d'une part, la facilitation de la mesure des attributs de l'apprenant et d'autre part, la favorisation de l'utilisation des techniques d'apprentissage car ces techniques fonctionnent plus correctement sur les données numériques.

4.4 L'architecture de l'ontologie des ressources matérielles (Material Resource Ontology : MRO)

Dans l'ontologie des ressources matérielles MRO, les contenus d'apprentissage ont été modélisés de façon où chaque concept fait référence à une leçon, un topic, un livre ou un contenu (voir figure 32), et chaque concept admet des sous-concepts de manière hiérarchique. Par exemple, le concept "contents" admet six sous-concepts catégorisés de la manière suivante : lecture, assignement, quiz, exams, exercices, prérequis. De plus, les concepts présentés dans la MRO sont étiquetés avec des niveaux de difficultés, ce qui rend son utilisation plus pratique pour une recommandation adaptée et alignée sur le niveau cognitif de l'apprenant.

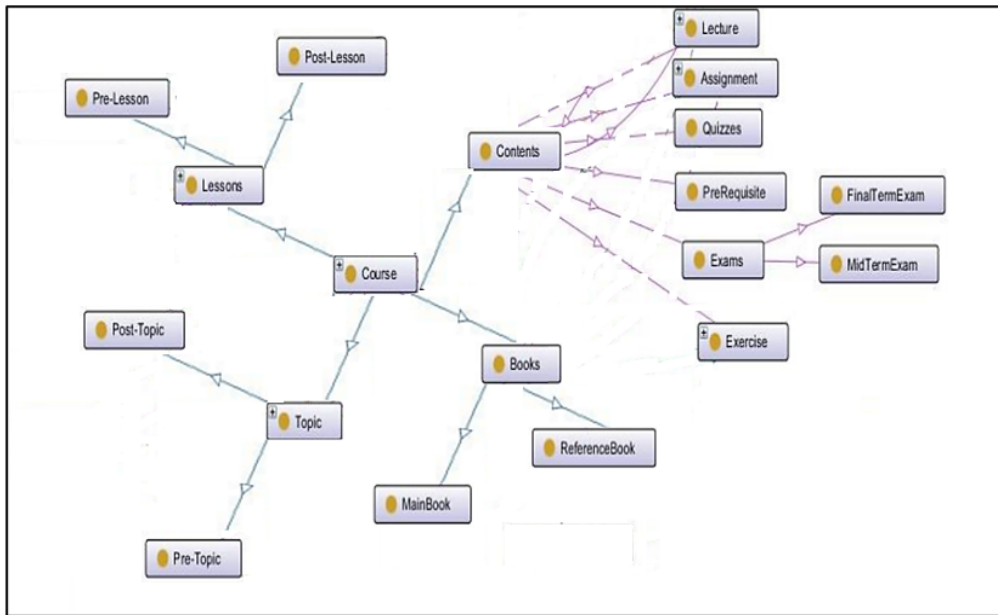


FIGURE 32 : L'architecture de l'ontologie des ressources matérielles

4.4.1 L'ontologie de ressource matérielle (MRO)

La génération des ressources est une méthode d'adaptation bien établie dans les systèmes d'apprentissage intelligents. Sa fonction principale contribue à la formalisation du matériel d'apprentissage individualisé pour chaque apprenant, cette formalisation sélectionne dynamiquement la méthode d'apprentissage optimale et convenable à chaque étape du processus d'apprentissage. Cette méthode permettra à l'apprenant d'atteindre son objectif d'apprentissage.

Dans notre architecture, les ressources d'apprentissage en ligne, se composent de modules d'apprentissage et de modules de test. La création de ressource est le processus qui consiste à sélectionner des modules de cours et à les agencer de manière appropriée pour le groupe cible ou les apprenants individuels. Il s'agit de l'un des domaines de recherche les plus intéressants liés aux systèmes d'apprentissage à distance.

4.4.2 Méthodes pour formaliser des ressources de formation

La formalisation de cours d'apprentissage en ligne dans les systèmes d'apprentissage intelligents comprend la planification du contenu ainsi que sa diffusion. La planification du matériel d'apprentissage (contenu) consiste à générer, organiser

et sélectionner ces contenus sur la base des connaissances actuelles de l'apprenant et à estimer un plan déterminant le moment où il faut le re-planifier (adapter un plan existant) ou générer un nouveau plan. La planification du contenu, également appelée stratégie d'apprentissage, fait référence à la sélection d'activités et d'interactions qui aideront l'apprenant à atteindre les objectifs d'apprentissage. A ce propos, Les stratégies d'apprentissage les plus courantes sont : l'apprentissage par l'exemple, l'apprentissage par la lecture de textes, l'apprentissage par la pratique et la simulation.

4.4.3 Ontologie dynamique de relation apprenant-cours (Learner-Course Relation Ontology : LCRO))

Dans l'objectif de recommander les contenus d'apprentissage aux apprenants, la LCRO calcule la similarité entre le nouvel apprenant et d'autres apprenants au profil similaire qui ont déjà consulté les plates-formes d'apprentissage. Dans l'approche que nous proposons, chaque apprenant envoie des requêtes demandant des matériaux de ressource spécifiques avec des propriétés particulières. Chaque demande transformée en sous-ensemble de caractéristiques formant deux groupes, les attributs du profil de l'apprenant et les attributs de la ressource matérielle. Si l'apprenant, qui demande l'information, est nouvellement inscrit, le système de recommandation E-learning avec ontologie dynamique (ERSDO) fournit dynamiquement toute l'information extraite de la demande envoyée à l'ontologie prédéfinie LPO et sécurise cette information avec le dynamique multi-protocole de sécurité. Cette approche fonctionne principalement à base de trois modules : le modèle de l'apprenant, le modèle du contenu d'apprentissage et le recommandeur basé sur la connaissance. Une explication complète de la représentation ontologique de ce modèle a été illustrée dans les sections précédentes.

Dans notre architecture, LCRO a principalement deux fonctions : il met à jour, de manière continue et systématique, toutes les ressources d'information des plateformes d'apprentissage en ligne à base du modèle web sémantique, puis il fournit aux apprenants des documents recommandés selon des mesures de similarités. Pour mettre un stockage dynamique de toutes les ressources d'information, LCRO utilise une technique de regroupement pour former des sous-groupes ayant des propriétés similaires. Dans l'étape de recommandation, LCRO calcule les similarités en fonction de l'apprenant et des ontologies de cours.

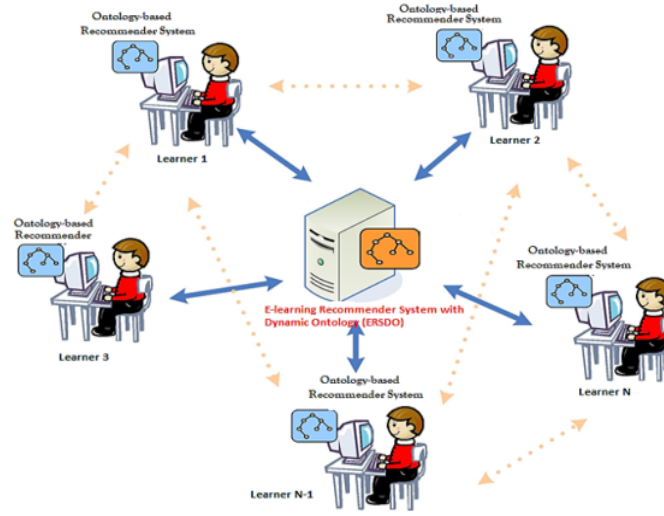


FIGURE 33 : Architecture de l'ontologie dynamique de l'apprentissage en ligne [124]

4.4.4 Similarité de LCRO

Pour calculer la similarité entre les profils des apprenants et les ressources pédagogiques, il existe plusieurs mesures. La mesure couramment éprouvée et que nous avons utilisée dans notre test est la similarité en cosinus [148]. Nous utilisons l'équation de similarité du cosinus pour calculer les scores de similarité entre deux vecteurs des attributs. Dans notre cas, le premier vecteur est issu des apprenants et le second vecteur est issu de la matrice des informations stockées dans l'ontologie LCRO :

$$\cos((\mathbf{w}_c), (\mathbf{w}_s)) = \frac{(\mathbf{w}_c) \cdot (\mathbf{w}_s)}{\|\mathbf{w}_c\| \cdot \|\mathbf{w}_s\|} \quad (\text{Chapitre III.1})$$

4.4.5 Algorithme de recommandation utilisé par LCRO

Le module de recommandation comprend la génération de la liste des cours recommandés à l'apprenant cible. Nous utilisons l'algorithme suivant basé sur l'ontologie LPO et l'ontologie MRO.

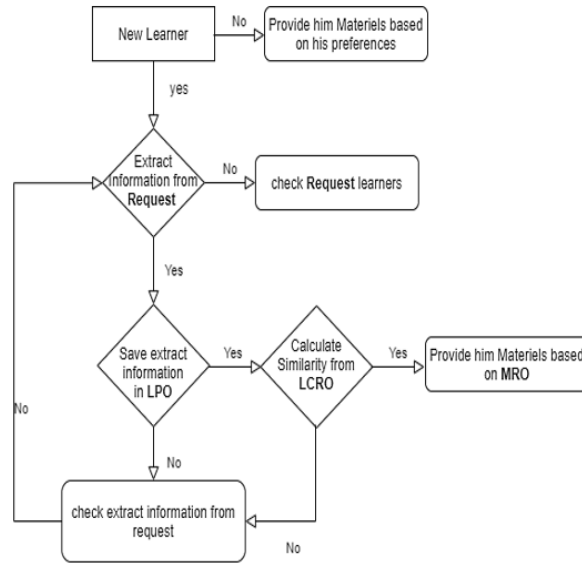


FIGURE 34 : le flux de travail de notre système de recommandation d'ontologie proposé [124]

5 Résultats expérimentaux

Dans cette section, nous détaillons les détails des résultats obtenus par la mise en œuvre de l'architecture que nous proposons. Notre architecture a été conçue pour utiliser Moodle, une plateforme de supports de cours bien connue et largement utilisée. Moodle fournit un ensemble complet d'outils et de fonctionnalités qui peuvent faciliter un enseignement et un apprentissage efficaces. Le choix de Moodle comme plateforme pour notre architecture repose sur plusieurs facteurs, notamment sa flexibilité, sa facilité d'utilisation et sa compatibilité avec divers systèmes de gestion de l'apprentissage.

Pour tester l'efficacité de l'architecture que nous proposons, nous avons mené des expériences avec des étudiants de premier cycle de l'Université Sidi Mohamed Ben Abdellah (USMBA). Nous avons choisi cette population en fonction de sa pertinence par rapport au sujet de recherche, ainsi que de son accessibilité. L'USMBA est un établissement d'enseignement supérieur très respecté, avec une population étudiante diversifiée, ce qui en fait un lieu idéal pour mener nos expériences.

Les expériences ont été menées pendant une période déterminée, au cours

Algorithme 1 : Top N Recommandations en utilisant LCRO

Begin

Input

Set of courses objects $CO = \{i1, i2, i3, \dots, in\}$ in MRO

Set of attributes learners $L = \{L1, i2, i3, \dots, ik\}$ in LPO

Output

top N recommendations

Method

for each $i \in CO, k \in L$, **do**

1 : Identify domain vocabulary using SKOS

2 : Use clustering technique to form sub-clusters with similar properties

FROM LPO

3 : Compute ontological similarity for each sub-clusters.

4 : Generate top N recommendation for target learner

End for

End

FIGURE 35 : Algorithme 1 : Top N Recommandations en utilisant LCRO

de laquelle les étudiants ont eu accès aux supports de cours via Moodle. Nous avons recueilli des données sur plusieurs paramètres, notamment l'engagement des étudiants, les taux de participation et les résultats d'apprentissage. Nous avons également mené des enquêtes et des entretiens pour recueillir les commentaires des étudiants sur leur expérience avec les supports de cours et la plateforme Moodle.

Les résultats de nos expériences ont montré que l'architecture proposée était efficace pour améliorer l'engagement des étudiants, les taux de participation et les résultats d'apprentissage. Les étudiants ont rapporté des expériences positives avec le matériel de cours et la plateforme Moodle, indiquant que notre architecture

était facile à utiliser et efficace pour faciliter leur apprentissage. Ces résultats sont importants car ils démontrent le potentiel de l'architecture que nous proposons pour améliorer la qualité de l'enseignement et l'expérience d'apprentissage des étudiants.

5.1 Analyse du jeu de données utilisée par LCRO

Pour assurer la validité de notre approche, notre expérience nécessite un grand ensemble de données avec des contributions de nombreux participants. Afin de répondre à cette exigence, nous avons choisi d'utiliser Moodle comme plateforme pour notre processus expérimental. Moodle, un système de gestion d'apprentissage open-source, offre une série d'avantages qui en font un choix approprié pour notre recherche.

L'un des avantages de l'utilisation de Moodle pour notre expérience est la possibilité de recueillir des données complètes sur les interactions des étudiants. En utilisant Moodle, nous pouvons suivre les différentes activités des étudiants, comme les cours qu'ils visitent, la façon dont ils remplissent leurs quiz et leurs devoirs, et d'autres paramètres. Ces informations détaillées sont cruciales pour évaluer l'efficacité de notre approche et s'assurer que les données collectées sont de haute qualité.

De plus, la nature open-source de Moodle nous permet d'intégrer nos algorithmes de manière transparente dans sa programmation. Cette intégration nous permet de créer un ensemble de données personnalisé, adapté aux besoins spécifiques de notre expérience. Pour générer notre ensemble de données, nous avons collecté des informations auprès de trois sources différentes : la faculté des sciences humaines de Dhar El Mahraz, la faculté des sciences humaines de Sais-Fes et l'ensemble de données de Coursera.

Le jeu de données de la faculté des sciences humaines de Dhar El Mahraz contient 11 départements, chacun comprenant entre 50 et 150 cours de soutien. Chaque département a son propre répertoire qui comprend tous les fichiers appartenant à ce département. De même, les ressources de la Faculté des Humanités, Sais-Fes sont répertoriées dans sept départements, chaque cours ayant son propre répertoire contenant tous les fichiers pertinents. Le jeu de données Coursera com-

prend principalement 890 ressources, que nous avons intégrées dans notre jeu de données Moodle.

En incorporant les données de ces diverses sources dans notre jeu de données Moodle, nous pouvons nous assurer que notre approche est validée dans une série de contextes différents. Cette approche nous permet de collecter des données de haute qualité qui sont plus représentatives de la population étudiante générale, ce qui conduit à des résultats plus précis et plus fiables.

Sources	Number Of de- part- ments	Number Of courses	Departments
FACULTY OF ART AND HU- MAN SCIENCES DHAR EL MAHRAZ	11	2378	Department of : Arabic Language and Literature, French Language and Li- terature, English Language and Literature, Hispanic Language and Literature, German Language and Lite- rature, Islamic Studies, His- tory, Geography, Philoso- phy, Sociology, Psychology.
FACULTY OF ART AND HU- MAN SCIENCES SAIS	7	1569	Department of : Arabic Language and Literature, French Language and Li- terature, English Language and Literature, Hispanic Language and Literature, German Language and Li- terature, Islamic Studies, Geography.
COURSERA	6 catego- ries	890	A variety of courses

TABLE 8 Ensemble de données propriétés du diplôme BA, années 2020/2021

5.2 Processus expérimental

L'objectif principal de ce travail est d'évaluer l'efficacité de notre algorithme dans la phase de recommandation d'un système d'apprentissage en ligne. Plus précisément, nous cherchons à déterminer dans quelle mesure notre algorithme peut recommander des objets d'apprentissage en fonction des préférences de nos apprenants, afin d'améliorer leur expérience d'apprentissage globale.

Pour atteindre cet objectif, nous avons mené deux expériences sur des serveurs différents (table 2), en utilisant le même jeu de données. Le but de ces expériences était d'évaluer le nombre de consultations du même jeu de données dans les deux serveurs, avec l'hypothèse que si le système recommande un contenu approprié aux apprenants, le nombre de consultations augmentera, ainsi que la durée de chaque consultation.

Dans le premier serveur, nous avons utilisé ERSDO, tandis que le second serveur a employé le filtrage de contenu classique à des fins de recommandation. Notre analyse était basée sur le nombre d'apprenants ayant accédé aux objets d'apprentissage recommandés et sur la durée de leurs consultations. Si le nombre d'accès était élevé et que la durée des consultations était significativement plus longue, cela indiquerait que notre approche a réussi à fournir un contenu approprié au public cible.

En comparant les résultats des deux expériences, nous avons pu évaluer l'efficacité de notre approche. Nous avons analysé le nombre de consultations de chaque objet d'apprentissage, ainsi que la durée de chaque consultation, afin de déterminer si notre algorithme était capable de faire des recommandations efficaces basées sur les préférences de nos apprenants. Si les données montraient que notre algorithme était capable de fournir un contenu pertinent et attrayant aux apprenants, cela indiquerait que notre approche a réussi à améliorer l'expérience d'apprentissage globale. Globalement, nos expériences ont été conçues pour tester l'efficacité de notre algorithme dans un contexte d'apprentissage en ligne. En évaluant le nombre de consultations et la durée de chaque consultation, nous avons pu tirer des conclusions sur l'efficacité de notre approche et sur sa capacité à recommander des objets d'apprentissage en fonction des préférences de nos apprenants.

5.3 Précision du classificateur

Afin d'évaluer l'efficacité de notre système ERSDO par rapport au filtrage classique basé sur le contenu, nous avons utilisé la mesure F1 comme principale mesure. Cette mesure est largement utilisée dans l'apprentissage automatique et la recherche d'informations pour évaluer la précision et les performances globales d'un système. La mesure F1 est une mesure combinée qui intègre à la fois la précision et le rappel, fournissant une évaluation plus complète de l'efficacité d'un système. La mesure F1 est calculée à l'aide de la formule suivante [149] :

$$\text{F1 measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (\text{Chapitre III.2})$$

Où la précision est le nombre de recommandations positives vraies divisé par le nombre total de recommandations faites, et le rappel est le nombre de recommandations positives vraies divisé par le nombre total d'éléments pertinents.

En utilisant la mesure F1, nous avons pu évaluer les performances de notre système ERSDO par rapport au filtrage classique basé sur le contenu d'une manière plus précise et plus complète. Cela nous a permis de déterminer quel système était le plus efficace pour recommander un contenu pertinent et intéressant à nos apprenants. Les résultats présentés dans le tableau 3 montrent que notre algorithme de recommandation ERSDO a surpassé l'approche de filtrage classique en termes de mesure F1. La mesure F1 était plus élevée pour notre système ERSDO proposé, ce qui indique qu'il était capable de fournir des recommandations de meilleure qualité à nos apprenants.

La mesure F1 est une mesure importante pour évaluer les performances des algorithmes de recommandation car elle prend en compte à la fois la précision et le rappel. Un score F1 plus élevé indique que l'algorithme a été en mesure de fournir des recommandations plus précises et plus pertinentes aux utilisateurs.

Les résultats du tableau 3 indiquent que notre système ERSDO a été capable de fournir un score F1 plus élevé par rapport à l'algorithme basé sur le filtrage classique. Cela suggère que notre approche est plus efficace pour recommander un contenu pertinent et attrayant à nos apprenants.

Taille des données par objet d'apprentissage	Nombre d'apprenants ayant utilisé des objets basés sur ERSDO (Server 1)	Nombre d'apprenants ayant utilisé des objets basés sur CONTENT-BASED (Server 2)
Petite taille	8789	6457
Taille moyenne	5687	4987
Grande taille	6192	2678

TABLE 9 Nombre des apprenants par serveur

Size of data by learning objects	ERSDO			CONTENT_BASED		
	P	R	F1	P	R	F1
[10-20[	0,439	0,391	0,41	0,357	0,389	0,371
[20-30[	0,481	0,318	0,382	0,297	0,341	0,318
[30-40[	0,452	0,395	0,421	0,317	0,429	0,369
[40-50[	0,39	0,43	0,409	0,26	0,394	0,313
[50-60[	0,58	0,42	0,487	0,31	0,47	0,379
[60-70[	0,61	0,567	0,597	0,411	0,393	0,401
[70-80[	0,691	0,68	0,687	0,599	0,549	0,582

TABLE 10 Precision, Recall et F1-measure de ERESDO et CONTENT BASED

5.4 Discussion

Les méthodes de recommandation conventionnelles utilisées ne prennent pas en compte les similarités entre les apprenants ayant les mêmes centres d'intérêt. Aussi, elles ne tiennent pas en compte certains problèmes techniques tels que la sparsité ou le démarrage à froid. ERSDO en tant qu'un système de recommandation vise à surmonter tous ces problèmes, y compris l'extraction de données à partir de sources multiples, en établissant une connexion dynamique entre les profils des utilisateurs et les ressources d'apprentissage, ce qui permet de réduire

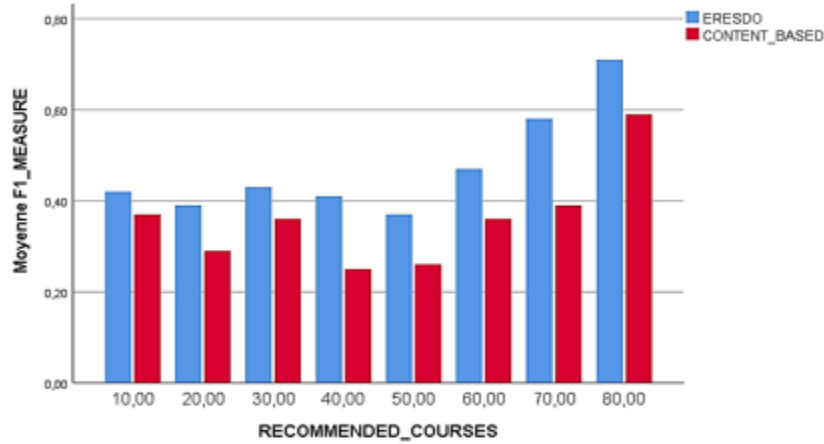


FIGURE 36 : La précision de la recommandation de l'approche ERESDO par rapport à l'approche CONTENT-BASED [124]

la surcharge d'informations et de préparer des recommandations correctes.

De plus, il fournit des cours adaptés qui conviennent au maximum aux besoins des apprenants. La figure 6 présente une comparaison entre l'approche ERESDO et l'approche traditionnelle basée sur le contenu en utilisant la mesure F1 ; les expériences ont montré que notre approche donne de meilleurs résultats en termes de performance et de précision en ce sens où le filtrage traditionnel est basé sur le contenu, car ce dernier présente de nombreuses limitations dans la phase de mise en correspondance de multiples ressources hétérogènes. L'intégration de l'ontologie dynamique, dans le processus de recommandation, améliore la qualité des recommandations et atténue les inconvénients des méthodes traditionnelles. De plus, l'utilisation de l'ontologie dans notre approche nous a permis d'incorporer le style d'apprentissage de l'apprenant et son niveau de connaissance dans le processus de recommandation, ce qui leur permet de fournir des recommandations plus personnalisées.

6 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle approche des systèmes de recommandation pour l'apprentissage en ligne, basée sur une ontologie dynamique. L'ontologie se compose de trois ontologies principales : l'ontologie de l'apprenant, l'ontologie du cours et l'ontologie du cours-apprenant, chacune d'entre

elles contenant plusieurs sous-modules. Ces ontologies sont mises à jour dynamiquement en utilisant les techniques du web sémantique et du web sémantique social, et sont intégrées dans le processus de recommandation. L'un des principaux avantages de notre approche est qu'elle traite efficacement le problème du démarrage à froid et de la dispersion, qui sont des défis courants dans les systèmes de recommandation E-learning. En incorporant une ontologie dynamique, nous sommes en mesure de faire des recommandations personnalisées qui prennent en compte les besoins et intérêts uniques de chaque apprenant. Les résultats expérimentaux présentés dans ce chapitre montrent que notre approche surpasse les approches traditionnelles basées sur le contenu en termes de performance et de précision.

Cela souligne l'importance des approches basées sur les ontologies dans les systèmes de recommandation E-learning, en particulier lorsqu'il s'agit de fournir aux apprenants les meilleures recommandations possibles. Enfin, nous avons remarqué que la classification des ressources d'apprentissage joue un rôle essentiel pour garantir la qualité de nos recommandations. En classant les ressources en fonction de leur contenu et d'autres facteurs pertinents, nous pouvons nous assurer que les apprenants reçoivent des recommandations adaptées à leurs besoins et intérêts spécifiques.

Dans l'ensemble, nous pensons que l'approche que nous proposons représente une avancée significative dans le développement des systèmes de recommandation pour l'apprentissage en ligne, et nous sommes impatients d'explorer davantage le potentiel des ontologies dynamiques dans ce contexte. Dans le chapitre suivant, nous nous focalisons sur notre deuxième contribution, qui porte sur la classification des objets d'apprentissage à l'aide d'algorithmes d'apprentissage automatique et d'outils intelligents. Cette contribution vise à permettre l'intégration de notre nouveau système de recommandation dans les systèmes de gestion de l'apprentissage (LMS), en fournissant des recommandations plus personnalisées à chaque apprenant.

Pour atteindre cet objectif, nous explorons diverses techniques d'apprentissage automatique, telles que la classification et les arbres de décision, qui peuvent être utilisés pour classer les objets d'apprentissage en fonction de leurs attributs et caractéristiques. Nous étudions également l'utilisation d'outils intelligents, tels

que le traitement du langage naturel de metadata pour extraire des informations pertinentes des objets d'apprentissage et améliorer la précision du processus de classification.

L'intégration de notre système de recommandation dans les LMS impliquera le développement d'un plugin capable de communiquer avec la base de données du LMS et de récupérer des informations pertinentes sur les apprenants et leurs préférences. Ces informations seront ensuite utilisées par le système de recommandation pour générer des recommandations personnalisées pour chaque apprenant, en tenant compte de leur style d'apprentissage, de leur niveau de connaissance et de leurs intérêts.

En combinant la puissance des algorithmes d'apprentissage automatique et des outils intelligents avec notre système de recommandation dynamique basé sur l'ontologie, nous visons à fournir un environnement d'apprentissage en ligne plus efficace et efficient, capable de répondre aux divers besoins d'apprentissage des étudiants. Dans le prochain chapitre, nous abordons les détails techniques de notre approche de classification et démontrons comment elle peut être intégrée dans les LMS pour améliorer la qualité des recommandations d'apprentissage en ligne. Dans le chapitre suivant, nous explorons une autre approche pour la recommandation en e-learning : la recommandation basée sur la classification multi-labels des objets d'apprentissage. Cette approche consiste à classer les objets d'apprentissage en fonction de leurs caractéristiques et à utiliser cette classification pour recommander des objets d'apprentissage pertinents à l'utilisateur. Nous examinons les principes fondamentaux de la classification multi-labels, les technologies utilisées, les avantages et les limites de cette approche. Nous présentons également les résultats de nos expériences pour évaluer la performance de cette approche. Ces expériences ont été menées en utilisant différentes mesures d'évaluation, telles que la précision, le rappel et la F-mesure. Les résultats obtenus nous permettront de mieux comprendre la performance de cette approche et de déterminer son utilité pour la recommandation en e-learning. Enfin, Nous discutons les différentes étapes de la mise en œuvre, des outils nécessaires, et des aspects techniques relatifs à l'implémentation d'un système de recommandation basé sur la classification multi-labels des objets d'apprentissage.

Chapitre IV

Implémentation d'un système de recommandation basé sur la Classification multi-labels des objets d'apprentissage

Sommaire

1	Introduction	135
2	Preliminaires	137
3	L'apprentissage automatique	154
4	Travaux relatifs aux objets d'apprentissage dans les systèmes d'apprentissage en ligne	161
5	Notre approche	164
6	Conclusion	172

1 Introduction

Dans le domaine de l'apprentissage en ligne, l'essor de l'e-pédagogie a eu un impact considérable sur les performances des systèmes d'apprentissage en ligne. Cette émergence a conduit à une augmentation du partage et de la création de ressources connues sous le nom d'objets d'apprentissage, qui peuvent être facilement explorés par les apprenants. Le terme "objets d'apprentissage" a été mentionné pour la première fois dans le domaine des systèmes d'apprentissage par Wayne Olgins en 1994 [150], et depuis, d'autres termes, tels que "objets de contenu" et "objets de connaissance", ont été proposés. En 2002, l'IEEE Learning Technology

Standards Committee a défini un objet d'apprentissage comme "toute entité, numérique ou autre, qui peut être utilisée pour l'apprentissage, l'éducation ou la formation"[39]. Cependant, la majorité des environnements d'apprentissage en ligne ne sont toujours pas en mesure d'exploiter pleinement la puissance de ces objets d'apprentissage, ce qui entrave leurs multiples utilisations et réutilisations dans différents contextes. Pour relever ce défi, nous proposons l'incorporation de systèmes de classification dans le web-mining pour maximiser l'exploration du contenu de ces objets. La classification des objets d'apprentissage peut également aider les apprenants à identifier les informations les plus pertinentes et à sélectionner les ressources les plus appropriées. La classification des objets d'apprentissage est un processus qui implique l'utilisation d'algorithmes d'apprentissage automatique, tels que la classification bayésienne et la machine à vecteur de support, pour structurer et regrouper les objets.

L'objectif de ce chapitre est double. Premièrement, il cherche à établir un modèle théorique pour la classification des objets d'apprentissage en combinant une approche de classification multi-labels [151] avec la logique floue [152]. L'approche de classification multi-label considère la présence de plusieurs étiquettes ou catégories pour un seul objet d'apprentissage, tandis que la méthode de la logique floue tient compte de l'imprécision et de l'incertitude dans le processus de classification. Deuxièmement, le chapitre vise la proposition d'une nouvelle méthode de partage d'objets basée sur la classification utilisant des techniques de web-mining. En intégrant les théories de classification aux techniques de web-mining, les performances des objets d'apprentissage peuvent être améliorées et rendues plus accessibles aux apprenants.

Le processus de sélection des caractéristiques dans la classification des objets d'apprentissage vise à déterminer les caractéristiques les plus importantes et à éliminer de l'ensemble de données qui ne sont pas nécessaires. Cela permet d'améliorer la précision et l'efficacité globales du processus de classification. L'ensemble de données utilisé dans ce processus peut être hautement dimensionnel, ce qui signifie qu'il comporte un grand nombre de caractéristiques, et par la suite il est essentiel d'éliminer les caractéristiques redondantes pour s'assurer que les résultats ne sont pas biaisés par des informations non pertinentes. Une fois les caractéristiques pertinentes identifiées, l'accent est mis sur la catégorisa-

tion des objets d'apprentissage à l'aide de la classification multi-labels (MLC) et des algorithmes Fuzzy C-Means (FCM). L'algorithme FCM est une technique de regroupement qui utilise la mesure de la distance euclidienne pour améliorer la précision du processus de classification. L'utilisation du MLC garantit que chaque objet d'apprentissage sera identifié par plusieurs étiquettes ou catégories, ce qui permet une catégorisation plus nuancée et plus précise. Pour valider l'efficacité de l'approche proposée, une série d'expériences a été menée à l'aide d'ensembles de données du monde réel. Les résultats de ces études prouvent que l'utilisation de FCM en combinaison avec MLC conduit à des résultats de classification améliorés par rapport aux méthodes traditionnelles. Le chapitre est structuré de manière à permettre une bonne compréhension du processus. Dans la section 1, les principaux concepts et définitions utilisés dans le travail sont expliqués en détail. La section 2 présente un résumé des méthodes utilisées pour classer les objets d'apprentissage, tandis que la section 3 décrit l'approche proposée. Enfin, la section 4 présente les résultats expérimentaux et les conclusions, ce qui permet de mieux comprendre la viabilité de l'approche proposée.

2 Préliminaires

La croissance rapide des systèmes d'apprentissage à distance et la popularité croissante des cours en ligne ouverts et massifs ont stimulé le développement de nouvelles technologies dans l'apprentissage en ligne. Avant de plonger dans la recherche pour cette étude, il est impératif d'exposer les concepts cruciaux qui sous-tendent notre classification des objets d'apprentissage. Dans le premier point de cette section, nous nous penchons sur la définition et les caractéristiques des objets d'apprentissage, afin de bien comprendre ce qu'ils sont et pourquoi ils sont importants dans l'apprentissage en ligne. Ensuite, nous examinons dans le deuxième point l'approche de la classification multi-labels (MLC), en explorant son importance dans le domaine des théories de classification. Dans le troisième point, nous nous concentrons sur la méthode Fuzzy C-Means (FCM), une technique largement utilisée pour le clustering et la classification. Nous examinons en détail ses performances, y compris ses forces et ses limites. Le dernier point met en évidence l'importance des théories de l'exploration des données du Web (WDM) dans le domaine de la classification. Le WDM joue un rôle crucial dans

le processus de collecte, d'analyse et d'extraction d'informations significatives à partir de grands ensembles de données, et son importance dans l'apprentissage en ligne ne peut être surestimée. .

2.1 Objets d'apprentissage

Le concept d'objets d'apprentissage a été introduit à la fin des années 1990 pour répondre au besoin croissant de matériel pédagogique numérique flexible et adaptable. L'idée était de créer des unités de contenu autonomes et modulaires pouvant être facilement assemblées dans des programmes ou dans des cours éducatifs plus importants, et réutilisées dans de multiples contextes d'apprentissage [153]. Au début des années 2000, avec la généralisation de l'internet et les progrès technologiques, le développement et l'utilisation des objets d'apprentissage ont pris de l'ampleur. La possibilité de partager et de distribuer des objets d'apprentissage en ligne a permis aux éducateurs, aux experts en la matière et aux organisations d'accéder à des ressources pédagogiques de haute qualité créées par d'autres et de les utiliser [154].

Aujourd'hui, les objets d'apprentissage sont largement utilisés dans diverses formes d'apprentissage en ligne et d'enseignement en ligne, notamment les cours en ligne ouverts et massifs (MOOC) [154]. Parmi les exemples d'objets d'apprentissage figurent des présentations multimédias, des simulations interactives, des quiz et des évaluations, et même des cours en ligne complets. La polyvalence et l'adaptabilité des objets d'apprentissage en font un outil essentiel pour soutenir l'apprentissage et le développement à l'ère numérique. En créant et en partageant ces ressources, les éducateurs et les experts en la matière peuvent aider les apprenants à accéder aux connaissances et aux compétences dont ils ont besoin pour réussir, où qu'ils soient et quels que soient leurs besoins d'apprentissage.

Les objets d'apprentissage présentent plusieurs caractéristiques essentielles qui en font un outil efficace pour soutenir l'apprentissage et le développement [45] :

- Modularité : les objets d'apprentissage sont des unités autonomes de contenu qui peuvent être combinées et assemblées dans des programmes sous forme des modules. Cette structure modulaire permet une personnalisation et une adaptation faciles à différents contextes et situations d'apprentissage.

- Réutilisabilité : les objets d'apprentissage peuvent être réutilisés et adaptés dans de multiples contextes d'apprentissage, ce qui permet aux éducateurs et aux experts en la matière d'accéder plus facilement à des ressources pédagogiques de haute qualité créée par d'autres et de les utiliser.
- Interopérabilité : les objets d'apprentissage peuvent être intégrés dans différents systèmes de gestion de l'apprentissage (LMS) et plateformes, ce qui permet d'utiliser et de partager ces ressources dans un large éventail d'environnements éducatifs.
- Adaptabilité : les objets d'apprentissage peuvent être personnalisés pour s'adapter à des contextes et à des situations d'apprentissage spécifiques, ce qui permet des expériences d'apprentissage personnalisées et efficaces.
- Accessibilité : la possibilité de distribuer des objets d'apprentissage en ligne permet aux apprenants d'accéder à des ressources pédagogiques de haute qualité de n'importe où et à tout moment.
- Évolutivité : la structure modulaire des objets d'apprentissage permet de faire évoluer les programmes et les cours éducatifs pour répondre aux besoins d'un grand nombre d'apprenants.
- Abordabilité : en créant et en partageant des objets d'apprentissage, les éducateurs et les experts en la matière peuvent contribuer à réduire les coûts associés à l'élaboration et à la diffusion de programmes et de cours éducatifs.

En termes pratiques, les objets d'apprentissage peuvent prendre de nombreuses formes et être utilisés dans un large éventail de contextes éducatifs. Par exemple, un objet d'apprentissage peut être une page Web avec un contenu pédagogique et des éléments multimédias, une simulation interactive qui permet aux apprenants d'explorer des concepts complexes, un test d'auto-évaluation pour évaluer la compréhension, ou un ensemble de directives ou de procédures pour aider les apprenants à accomplir une tâche spécifique.

Le terme "Objet d'apprentissage" a été formellement défini par le Learning Technology Standards Committee (LTSC) comme suit [155] :

" (...) un objet d'apprentissage est défini comme toute entité, numérique ou

non numérique, qui peut être utilisée pour l'apprentissage, l'éducation ou la formation. "

Avant la définition établie par le Learning Technology Standards Committee, David Wiley s'est penché sur la conception des objets d'apprentissage dans sa thèse de doctorat et a formulé sa propre définition [38] :

"toute ressource numérique qui peut être réutilisée pour soutenir l'apprentissage".

Le concept d'objets d'apprentissage est intégré dans les systèmes de gestion de contenu d'apprentissage (LCMS) comme un moyen d'organisation pour accéder au contenu éducatif. La manière dont les objets d'apprentissage sont structurés et utilisés dans une plate-forme d'apprentissage est décrite dans un modèle de contenu. Un exemple d'un tel modèle est illustré à la figure 1. Le modèle de contenu sert de plan directeur pour l'organisation et la gestion des objets d'apprentissage dans le LCMS, garantissant que l'information est stockée et récupérée d'une manière cohérente et efficace. L'utilisation d'un modèle de contenu permet la manipulation efficace et organisée de grandes quantités de contenu éducatif, rationalisant le processus d'accès et d'utilisation de l'information dans la plate-forme d'apprentissage.

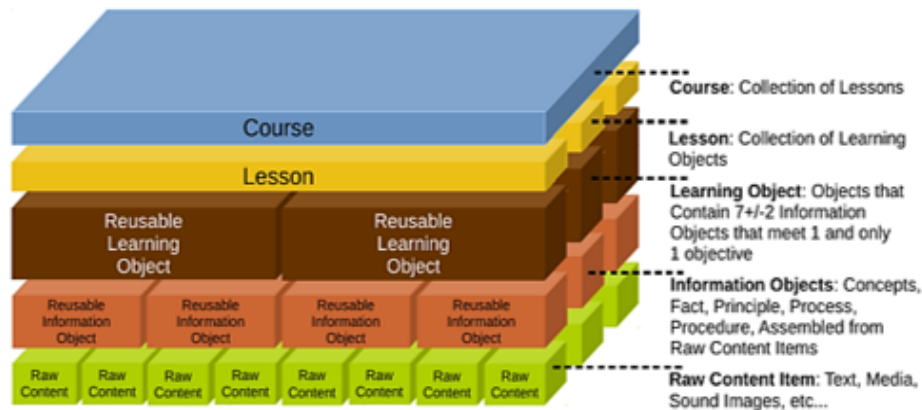


FIGURE 37 : : Représentation d'un modèle de contenu pour un LCMS avec des objets d'apprentissage réutilisables. [156]

Le modèle de contenu d'un système de gestion de l'apprentissage (LMS) est généralement composé de plusieurs couches, chacune servant un objectif différent.

Ces couches sont conçues pour garantir une utilisation efficace et efficiente des objets d'apprentissage au sein de la plateforme. Les différentes couches du modèle de contenu peuvent inclure [157] :

- Les données brutes (Raw Data) : cette couche est constituée des actifs numériques réels utilisés pour l'apprentissage, telle que le texte, les images, les fichiers audio et vidéo.
- Objet d'apprentissage réutilisable (Reusable Learning Object) : un objet d'apprentissage réutilisable (RLO) est une ressource numérique qui a été conditionnée et formatée de manière à pouvoir être facilement réutilisée et réaffectée dans différents contextes d'apprentissage. Un RLO peut être composé d'un ou plusieurs éléments de données brutes et est conçu pour être indépendant du contexte.
- Leçon (Lesson) : une leçon est une collection d'objets d'apprentissage réutilisables qui ont été combinés pour créer une expérience d'apprentissage spécifique. Une leçon peut inclure du texte, des images, des vidéos et d'autres médias, ainsi que des évaluations et des éléments interactifs.
- Cours (Course) : un cours est une collection plus large de leçons qui sont conçues pour fournir une expérience d'apprentissage complète. Les cours peuvent être autonomes ou faire partie d'un programme éducatif plus vaste.
- Programme d'études (Curriculum) : un programme d'études est une collection de cours qui sont organisés et structurés pour fournir un programme éducatif complet. Le programme d'études peut être considéré comme la couche supérieure du modèle de contenu, représentant le programme éducatif global.

Il est important de noter que la structure exacte du modèle de contenu peut varier d'un LMS à l'autre. Cependant, l'objectif fondamental du modèle de contenu vise à fournir un cadre flexible et évolutif pour l'organisation et la diffusion du contenu pédagogique. La figure 38 illustre ce principe.

Dans leur travail, Baumgartner et Kalz [158] ont proposé une représentation d'un objet d'apprentissage composé de deux sous-objets : un sous-objet didactique et un sous-objet technique. Le sous-objet didactique représente le contenu

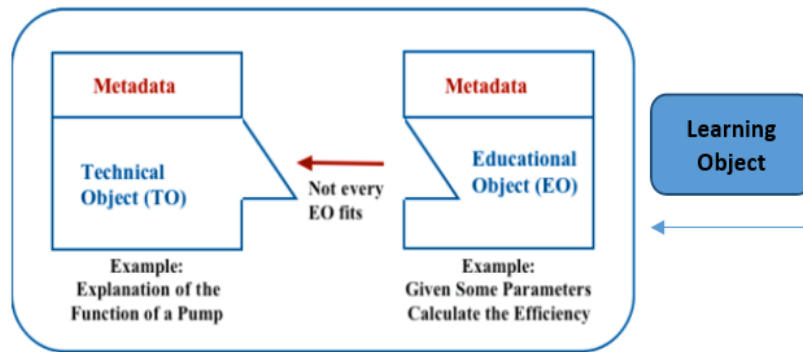


FIGURE 38 : Représentation d'un objet d'apprentissage composé d'un sous-objet didactique et d'un sous-objet technique [158]

éducatif de l'objet d'apprentissage, il est responsable de l'enseignement réel du matériel d'apprentissage. Il comprend des informations sur les objectifs d'apprentissage, le processus d'apprentissage et l'évaluation des résultats d'apprentissage. D'autre part, le sous-objet technique représente les aspects techniques de l'objet d'apprentissage, tels que le format et la structure du matériel d'apprentissage, les métadonnées qui décrivent l'objet d'apprentissage, et les exigences techniques pour accéder et utiliser ces objets. La division d'un objet d'apprentissage en ces deux sous-objets est basée sur l'idée qu'il devrait y avoir une séparation claire entre le contenu pédagogique et les aspects techniques. Cette séparation facilite la réutilisation, l'adaptation et la mise à jour de l'objet d'apprentissage, ainsi que son partage entre différents systèmes et plateformes éducatifs. En outre, elle permet le développement d'objets d'apprentissage spécifiquement conçus pour différents types d'apprenants, de styles d'apprentissage et d'environnements éducatifs. Par conséquent, la représentation composée d'un sous-objet didactique et d'un sous-objet technique proposée fournit un cadre utile pour le développement et la gestion d'objets d'apprentissage qui sont flexibles, adaptables et accessibles à tous les apprenants.

Le Learning objet Metadata (LOM) est aujourd'hui le modèle le plus répandu pour représenter les objets d'apprentissage au sein d'un système. Ces métadonnées aident, d'une part, à rechercher des contenus et des formats de documents spécifiques et, d'autre part, à évaluer si certains programmes d'apprentissage sont bien adaptés à des besoins des apprenants. Pour souligner l'importance et l'utilité des métadonnées pour les objets d'apprentissages et de savoir si elles sont

vraiment toutes nécessaires. Ochoa et al. [159] ont mené une enquête pour savoir quels éléments de données du modèle LOM sont les plus utilisés et à quelle fréquence. Pour ce faire, ils ont collecté et analysé plus d'un demi-million d'enregistrements de métadonnées provenant d'une grande collection de métadonnées d'objets d'apprentissage utilisant le standard LOM. Sur les 86 champs, seulement 43 sont répertoriés.

2.2 La classification multi-Labels (MLC)

La classification multi-Labels (MLC) est une approche d'apprentissage automatique qui attribue plusieurs étiquettes à une seule instance. Elle est utilisée dans divers domaines tels que la classification d'images, la classification de textes et les systèmes de recommandation. L'approche MLC a gagné une grande popularité ces dernières années car elle permet une représentation et une analyse plus complètes des données, par rapport à la classification traditionnelle à étiquette unique [151].

L'approche MLC a été largement étudiée et plusieurs avancées ont été réalisées dans ce domaine. En 2005, les auteurs Baumgartner et Kalz [158] ont introduit le concept d'une approche MLC basée sur un arbre de décision. En 2015, plusieurs études ont été menées en utilisant différentes ressources et propriétés pour la MLC. En 2019, les auteurs ont introduit un nouveau concept de classification des objets d'apprentissage basé sur la mémoire à long terme (LSTM) et l'approche de classification Random Forest. Ces avancées dans la MLC ont permis d'obtenir des résultats de classification meilleurs et plus précis dans une variété d'applications.

Le schéma de la classification multi-labels se compose généralement de plusieurs étapes, notamment [160] :

- Le prétraitement des données : cette étape consiste à nettoyer, transformer et organiser les données afin qu'elles soient prêtes pour le processus de classification. Il peut s'agir de traiter les valeurs manquantes, de normaliser les données et de convertir les variables catégorielles en variables numériques.
- Extraction de caractéristiques : à ce stade, les caractéristiques les plus pertinentes des données sont sélectionnées pour être utilisées comme données

d'entrée du modèle de classification. Il peut s'agir de sélectionner les variables les plus pertinentes ou de réduire la dimensionnalité des données à l'aide de techniques telles que l'analyse en composantes principales (ACP) ou l'analyse discriminante linéaire (ADL).

- Sélection du modèle : à ce stade, l'algorithme de classification multi-labels le plus approprié est sélectionné en fonction des caractéristiques des données et du problème à résoudre. Les algorithmes couramment utilisés dans la classification multi-labels comprennent les arbres de décision, les machines à vecteurs de support (SVM), Naive Bayes et les réseaux neuronaux.
- Entraînement du modèle : le modèle sélectionné est entraîné en utilisant les données prétraitées, dans le but d'apprendre les relations entre les caractéristiques et les étiquettes.
- Évaluation du modèle : cette étape consiste à évaluer les performances du modèle formé à l'aide de mesures telles que l'exactitude, la précision, le rappel et le score F1.
- Déploiement : si le modèle est performant, il peut être déployé dans des applications du monde réel. Dans le contexte de la classification multi-labels, cela pourrait impliquer l'intégration du modèle dans une plateforme d'apprentissage en ligne pour classer et organiser les ressources pédagogiques.

2.3 Types de Classification Multi-label

Il existe plusieurs types de classification multi-labels[161][162], notamment :

1. Pertinence binaire : la pertinence binaire est une approche simple où un classificateur binaire distinct est formé pour chaque étiquette. Par exemple, considérons un scénario dans lequel nous devons classer des articles dans plusieurs catégories telles que "sports", "politique", "divertissement" et "technologie". Un classificateur binaire est formé pour chaque catégorie afin de prédire si un article appartient ou non à cette catégorie. Il y aura donc quatre classificateurs distincts dans ce cas. L'avantage de cette méthode est sa simplicité, mais elle ne prend pas en compte les dépendances entre les étiquettes.

2. Chaînes de classificateurs : les chaînes de classificateurs sont une méthode qui enchaîne les classificateurs binaires de telle sorte que la sortie d'un classificateur est utilisée comme entrée pour le classificateur suivant. Par exemple, considérons le même scénario que ci-dessus, mais cette fois, nous utiliserons des chaînes de classificateurs. Le premier classificateur sera entraîné à prédire si un article appartient ou non à la catégorie "sports". Le deuxième classificateur prendra la sortie du premier classificateur, ainsi que les caractéristiques de l'article, comme entrée et prédira s'il appartient ou non à la catégorie "politique". Ce processus se poursuit pour les autres catégories. Cette méthode prend en compte les dépendances entre les étiquettes et peut donner lieu à une amélioration des performances par rapport à la pertinence binaire.
3. Label Powerset : label Powerset transforme le problème multi-étiquettes en un problème multi-classes en considérant toutes les combinaisons d'étiquettes possibles pour un échantillon. Par exemple, considérons le même scénario que ci-dessus. L'approche Label Powerset prendrait en compte toutes les combinaisons possibles des quatre catégories, telles que "sport", "politique", "divertissement" et "technologie", "sport" et "politique", "sport" et "divertissement", etc. Ces combinaisons sont traitées comme des classes distinctes, et un classificateur unique est entraîné à prédire la classe (combinaison d'étiquettes) pour chaque échantillon.
4. Adaptive Boosting (AdaBoost.MH) : adaptive Boosting (AdaBoost.MH) est une extension de l'algorithme AdaBoost spécialement conçue pour la classification multi-label. Il adapte les poids des instances en fonction de leur mauvaise classification dans les tours précédents, et cette adaptation est faite d'une manière qui prend en compte les dépendances entre les étiquettes. Par exemple, considérons le même scénario que ci-dessus. Dans ce cas, l'algorithme AdaBoost. MH prendrait en compte les corrélations entre les quatre catégories et ajusterait les poids des instances en conséquence.
5. Arbres de décision : les arbres de décision sont des modèles arborescents qui peuvent traiter des problèmes de classification à étiquettes multiples en créant plusieurs branches pour chaque étiquette. Par exemple, considérons le même scénario que ci-dessus. Un arbre de décision peut être utilisé

pour prédire les catégories de chaque article. À chaque nœud de l'arbre, une décision est prise en fonction des caractéristiques de l'article, et la prédiction finale est le résultat du suivi des branches jusqu'aux feuilles, qui représentent les étiquettes de classe.

6. Random Forest : random Forest est un ensemble d'arbres de décision et peut être utilisé pour la classification multi-label. Chaque arbre de l'ensemble fait une prédiction, et la prédiction finale est le résultat de la combinaison des prédictions de tous les arbres. Par exemple, considérons le même scénario que ci-dessus. Dans ce cas, une forêt aléatoire serait formée pour prédire les catégories de chaque article. Les prédictions de plusieurs arbres de décision sont combinées pour obtenir la prédiction finale.
7. Réseaux neuronaux : les approches basées sur l'apprentissage profond, telles que les réseaux neuronaux convolutionnels (CNN) et les réseaux neuronaux récurrents (RNN), peuvent également être utilisées pour la classification multi-label.

2.4 Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM) est un algorithme de regroupement de données qui est utilisé pour partitionner un ensemble de données en plusieurs groupes (ou clusters) sur la base de leur similarité. Le FCM est une version floue du populaire algorithme de clustering K-Means et est utilisé lorsqu'il y a une incertitude dans les données ou lorsqu'un partitionnement précis n'est pas possible [163].

Dans le FCM, chaque point de données est associé à un ensemble de valeurs d'appartenance qui reflètent son degré d'appartenance à chacun des clusters. Les valeurs d'appartenance sont déterminées en minimisant une fonction objective qui quantifie la différence entre les points de données et les centroïdes des clusters. La fonction objective utilisée dans FCM est la fonction objective fuzzy c-means, qui est basée sur le carré de la distance euclidienne entre chaque point de données et les centres des clusters[164].

L'algorithme FCM commence par initialiser de manière aléatoire les centres des clusters, puis met à jour les valeurs d'appartenance de chaque point de données à chaque itération jusqu'à convergence. Le processus de mise à jour des

valeurs d'appartenance et des centroïdes de cluster se poursuit jusqu'à ce que la fonction objective atteigne un minimum ou un nombre maximal d'itérations. Le modèle FCM a été largement utilisé dans de nombreuses applications, telles que la segmentation d'images, la classification de textes et la reconnaissance de formes. Il s'est avéré être un algorithme robuste et efficace pour les problèmes de classification multi-labels où il y a de l'incertitude dans les données ou lorsqu'un partitionnement précis n'est pas possible [165].

L'un des principaux avantages du FCM par rapport aux autres algorithmes de clustering est sa capacité à gérer l'incertitude des données, qui est souvent présente dans les applications du monde réel. De plus, le FCM permet le chevauchement des clusters, ce qui le rend plus adapté aux problèmes de classification multi-label où un seul point de données peut appartenir à plusieurs clusters. L'algorithme fonctionne en plusieurs étapes [165], notamment :

- L'initialisation : l'étape initiale consiste à sélectionner aléatoirement k centres de cluster, où k est le nombre de clusters souhaités. Les centres de grappes représentent les estimations initiales de l'emplacement des grappes.
- Calcul de l'appartenance : dans cette étape, les valeurs d'appartenance de chaque point de données à chacun des k clusters sont calculées. Les valeurs d'appartenance sont calculées en fonction de la distance entre le point de données et le centre de la grappe, et le degré d'appartenance de chaque point de données est représenté par une matrice d'appartenance.
- Calcul du centroïde de la grappe : l'étape suivante consiste à calculer les centroïdes de la grappe, qui sont mis à jour en fonction des valeurs d'appartenance. Les centroïdes sont mis à jour en prenant la moyenne de tous les points de données dans le cluster, où les poids de chaque point de données sont déterminés par sa valeur d'appartenance.
- Mise à jour de l'appartenance : après la mise à jour des centroïdes de la grappe, les valeurs d'appartenance de chaque point de données sont mises à jour sur la base des centroïdes nouvellement calculés.
- Vérification de la convergence : dans cette étape, l'algorithme vérifie si les changements dans les valeurs d'appartenance sont dans un niveau de

tolérance spécifié. Si les changements sont dans la tolérance, l'algorithme converge et se termine, sinon les étapes 2 à 4 sont répétées jusqu'à ce que la convergence soit atteinte.

- Affectation des clusters : une fois que l'algorithme converge, chaque point de données est affecté à la grappe ayant la valeur d'appartenance la plus élevée.

Ce cycle de mise à jour des centroïdes de cluster, des valeurs d'appartenance et de vérification de la convergence se poursuit jusqu'à ce que l'algorithme converge vers une solution. La solution représente les affectations finales des clusters et les centroïdes des clusters, qui peuvent alors être utilisés pour une analyse ou une prise de décision ultérieure.

2.4.1 Les types de FCM

- FCM standard : il s'agit de la forme originale de l'algorithme FCM, où l'objectif est de minimiser la fonction objectif qui consiste en la somme des distances au carré entre les points de données et les centres de cluster. Par exemple, dans le cadre de la segmentation de la clientèle, une entreprise peut recueillir des données sur l'historique des achats des clients et utiliser l'algorithme FCM standard pour regrouper les clients en différents groupes en fonction de leur comportement.
- FCM modifié : il s'agit d'une variante de l'algorithme FCM standard qui incorpore un terme de régularisation pour éviter les problèmes de singularité dans la fonction objective. Par exemple, dans la segmentation d'images, un ensemble de données d'imagerie médicale peut contenir certaines images avec des niveaux de bruit élevés. Pour surmonter les limites du FCM standard dans ces scénarios, le FCM modifié peut être utilisé avec un terme de régularisation pour éviter les problèmes de singularité et fournir de meilleurs résultats.
- FCM possibiliste : cette variante de l'algorithme FCM introduit une vision possibiliste du clustering, où les valeurs d'appartenance sont interprétées comme la possibilité ou la probabilité qu'un point de données appartienne

à un cluster particulier. Par exemple, dans la segmentation de marché, une entreprise peut utiliser le FCM possibiliste pour regrouper ses clients en fonction de leur probabilité d'acheter un produit particulier. Dans ce cas, les valeurs d'appartenance représenteraient la probabilité qu'un client appartienne à un groupe de produits particulier.

- FCM spatial : cette variante de l'algorithme FCM prend en compte les relations spatiales entre les points de données, et est particulièrement utile pour la segmentation d'images et la reconnaissance de formes. Par exemple, dans la segmentation d'images, le FCM spatial peut être utilisé pour segmenter une image en différentes régions sur la base des relations spatiales entre les pixels.
- FCM robuste : cette variante de l'algorithme FCM est conçue pour être robuste aux valeurs aberrantes et aux données bruyantes, en incorporant des techniques d'estimation robustes dans la fonction objective. Par exemple, dans le cadre de la segmentation de la clientèle, une entreprise peut avoir certains clients dont le comportement est inhabituel et qui s'écartent de la norme. Le modèle FCM robuste peut être utilisé pour regrouper ces clients en fonction de leur comportement, même en présence de données aberrantes et bruyantes.
- FCM multi-objectif : cette variante de l'algorithme FCM incorpore plusieurs objectifs dans la fonction objectif, permettant l'optimisation simultanée de plusieurs critères. Par exemple, dans le cas de la segmentation de la clientèle, une entreprise peut vouloir optimiser simultanément la taille des groupes et la similarité des clients au sein de chaque groupe. La méthode FCM multi-objectifs peut être utilisée pour trouver un compromis entre ces multiples objectifs.
- FCM hybride : cette variante de l'algorithme FCM combine les forces de différentes variantes de FCM pour surmonter les limites d'une seule approche. Par exemple, dans la segmentation d'images, une approche FCM hybride pourrait incorporer à la fois des informations spatiales et des techniques d'estimation robustes pour fournir de meilleurs résultats que l'une ou l'autre approche seule.

2.4.2 Fuzzy C-Means (FCM) et les objets d'apprentissages

Fuzzy C-Means (FCM) est un algorithme de regroupement populaire qui peut être appliqué dans le contexte de la classification des objets. Cet algorithme affecte chaque point de données à plusieurs clusters avec un degré d'appartenance, qui représente la similarité entre le point de données et les centres des clusters. Dans la classification d'objets, le FCM peut être utilisé pour identifier les affectations de clusters pour un ensemble d'objets sur la base de leurs caractéristiques, telles que la couleur, la texture, la forme, le contenu, etc.

Les centres de regroupement servent de prototypes ou de modèles des objets dans les différentes classes, et les nouveaux objets peuvent être classés en déterminant leur degré d'appartenance à ces centres de regroupement. La classification est effectuée en assignant le nouvel objet à la classe ayant le degré d'appartenance le plus élevé.

La méthode FCM a été appliquée avec succès à diverses tâches de classification d'objets, notamment la classification d'images, la classification de textes et la bioinformatique. Dans la classification d'images, par exemple, la FCM a été utilisée pour classer les objets dans différentes classes en fonction de leurs caractéristiques visuelles, telles que la couleur, la texture et la forme [166]. Dans la classification de textes, la FCM a été utilisée pour classer les documents dans différentes catégories en fonction de leur contenu et de leurs caractéristiques. En bioinformatique, la FCM a été utilisée pour classer les données d'expression génétique en différentes catégories fonctionnelles.

Cependant, il est crucial de comprendre que la précision des classifications produites par le FCM dépend de la qualité des résultats du clustering. Cette qualité est influencée par des facteurs tels que le nombre de clusters, les conditions initiales et la métrique de similarité utilisée. Par conséquent, une expérimentation et une validation approfondies sont essentielles pour garantir que le classificateur basé sur le FCM fournit des résultats précis et fiables.

2.5 Data Mining

2.5.1 Définition

Le data mining fait référence au processus d'analyse de quantités massives de données sous différents angles pour identifier les relations entre les entités qui les composent et les transformer en informations exploitables. En français, ce procédé porte différents noms : exploration de données fouille de données Ou extraction des connaissances à partir de données. Le data mining est une branche de la science des données qui combine des méthodes statistiques et mathématiques avec des systèmes d'apprentissage automatique appliquées sur des bases de données. Le concept de data mining a émergé dans les années 1980 et 1990 – avec l'avènement de la gestion de bases de données et des techniques de Machine Learning [167] .

Les termes analyse de données et exploration de données sont souvent confondus. Cependant, l'analyse de données peut être comprise comme une sous-catégorie de l'exploration de données qui se focalise sur :

Nettoyage des données brutes,

Identification des modèles,

Création des modèles.

Par contre l'analyse des données, est la partie de l'exploration de données qui s'intéresse plus sur l'obtention de l'information à partir des données. Leur objectif est d'appliquer l'analyse statistique pour rendre les données plus lisibles.

2.5.2 Origine du concept Data mining

L'une des premières traces du concept de data mining remonte à 1936. C'est cette année-là que le mathématicien et cryptographe britannique Alan Turing a proposé l'idée d'une machine capable d'effectuer l'analyse et l'exploration des données similaires à ceux des ordinateurs modernes[168]. Ensuite, dans les années 1980, le data mining a commencé avoir l'essor et à trouver l'existence. Il tire son nom du fait que les data scientists de l'époque comparaient la recherche d'informations précieuses dans de grandes bases de données avec l'extraction de minerai. Les deux processus nécessitent de passer au crible de grandes quantités de matériaux pour trouver des articles de grande valeur.

2.5.3 Méthodes d'exploration de données

Un certain nombre de méthodes sont utilisées dans l'exploration de données. Selon le spécialiste de l'intégration de données Talend, les éléments suivants sont parmi les plus utilisés[169] :

- Nettoyage de données : avant que les données puissent être analysées et traitées, les erreurs doivent être identifiées et supprimées dans le cadre du nettoyage et de la préparation des données.
- L'exploration de données : l'exploration de données utilise souvent l'intelligence artificielle pour des tâches liées à la planification, à l'apprentissage, au raisonnement et à la résolution de problèmes.
- Le clustering : le clustering est utilisé pour diviser un ensemble de données en classes significatives et pour comprendre la structure des données.
- La classification : dans le cadre de la classification des données, les objets du jeu de données sont liés à des catégories ou classes cibles. L'objectif est de les prédire le plus précisément possible.
- L'analyse de données : l'analyse des données extrait des informations des données.
- Le stockage de données : l'entreposage de données est la base de la plupart des processus d'exploration de données. Un entrepôt de données est une collection de données métiers.
- L'apprentissage automatique : avec l'aide de l'apprentissage automatique, la recherche de modèles dans les données est automatisée.
- L'analyse de régression : est appliquée à un ensemble de données spécifique pour prédire certaines variables (telles que les chiffres de vente, la température ou le cours des actions).

2.5.4 Data mining - Processus

Les spécialistes du data mining obtiennent généralement des résultats fiables en suivant un processus structuré et réitérable en six étapes :

Le modèle de processus en six étapes « Cross Industry Standard Process for

Data Mining » (CRISP-DM) a été publié en 1999.

L'objectif : standardiser les processus d'exploration de données dans tous les secteurs. Les six phases du CRISP-DM sont les suivantes [170] :

1. Compréhension de l'entreprise : cette phase consiste à comprendre les objectifs, les exigences et la portée du projet. Il est divisé en quatre tâches qui sont :
 - Détermination des objectifs en comprenant la motivation des parties prenantes ;
 - Evaluation de la situation pour déterminer la disponibilité des ressources, les besoins du projet, les risques et les imprévus ;
 - Définition des succès d'un point de vue technique.
 - Élaboration des plans détaillés pour les technologies et les outils.
2. Compréhension des données : la phase suivante consiste à identifier, collecter et analyser les ensembles de données nécessaires pour atteindre les objectifs du projet. Cette phase comprend également quatre tâches :
 - La collecte des données initiales ;
 - La description de données ;
 - L'examen de données ;
 - La vérification de la qualité des données.
3. Préparation des données : la troisième phase, souvent la plus longue, d'un projet comprend cinq tâches :
 - Sélectionner des ensembles de données et documenter les raisons de l'inclusion ou de l'exclusion ;
 - Nettoyer les données ;
 - Dérivation de nouveaux attributs à partir de données existantes pour construire de nouvelles données ;
 - Intégrer et formater des données provenant de différentes sources.

4. Modelage : la création de modèles de données implique quatre tâches :
 - La sélection des techniques de modélisation ;
 - La création des conceptions de test ;
 - La création des modèles ;
 - L'évaluation des modèles.
5. Évaluation : bien que la phase de modélisation comprenne l'évaluation du modèle technique, cette phase consiste à déterminer quel modèle répond le mieux aux besoins de l'entreprise. Il a trois tâches :
 - Évaluation des résultats ;
 - Processus d'examen ;
 - définition des prochaines étapes.
 - Déploiement : La phase finale consiste à mettre le modèle

en pratique. Quatre tâches sont envisagées ici :

- élaboration et documentation de plan de déploiement du modèle ;
- Création de plan de surveillance et de maintenance ;
- Préparation d'un rapport final ;
- Examiner le projet.

Avec Analytics Solutions Unified Method for Data Mining (ASUM-DM), IBM a publié une extension en 2015 qui s'appuie sur CRISP-DM. Le modèle de Big Blue complète la phase de déploiement avec la collaboration, le contrôle de version, la sécurité et la conformité.

3 L'apprentissage automatique

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle qui se concentre sur le développement d'algorithmes et de modèles statistiques permettant aux ordinateurs d'apprendre à partir de données sans être explicitement programmés. L'objectif principal de l'apprentissage automatique est de

développer des algorithmes capables d'améliorer automatiquement leurs performances avec l'expérience [171].

Les algorithmes d'apprentissage automatique peuvent être classés en trois grandes catégories : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement [171].

3.1 L'apprentissage supervisé

L'apprentissage supervisé est un type d'apprentissage automatique où l'algorithme est entraîné sur un ensemble de données étiquetées, ce qui signifie que les caractéristiques d'entrée et la sortie souhaitée sont fournies. L'objectif de l'apprentissage supervisé est de construire un modèle capable de généraliser à partir des données d'apprentissage pour faire des prédictions précises sur de nouvelles données non vues [172].

Dans l'apprentissage supervisé, les données d'apprentissage sont constituées de paires de caractéristiques d'entrée et de l'étiquette de sortie correspondante. Par exemple, dans la classification d'images, les caractéristiques d'entrée peuvent être les valeurs des pixels d'une image, tandis que l'étiquette de sortie peut être l'objet présent dans l'image (par exemple, chien, chat, etc.). L'algorithme utilise les données d'apprentissage pour apprendre la correspondance entre les caractéristiques d'entrée et l'étiquette de sortie.

Une fois le modèle formé, il peut être utilisé pour faire des prédictions sur de nouvelles données en fournissant les caractéristiques d'entrée et en utilisant la correspondance apprise pour déterminer l'étiquette de sortie. La précision du modèle est généralement évaluée à l'aide d'un ensemble de données qui n'a pas été utilisé pour la formation, appelé données de test. Les algorithmes d'apprentissage supervisé peuvent être classés en deux groupes : régression et classification [173].

- Les algorithmes de régression sont utilisés lorsque la sortie est une valeur continue, telle que le prix d'une maison ou la température. Dans la régression, l'objectif est de trouver une fonction qui associe les caractéristiques d'entrée à la valeur de sortie continue. Les algorithmes de régression courants comprennent la régression linéaire, la régression polynomiale et les arbres de décision.

- Les algorithmes de classification, quant à eux, sont utilisés lorsque la sortie est une valeur discrète, comme une classification binaire (oui/non) ou une classification multi-classes (par exemple, différentes catégories d'objets dans une image). En classification, l'objectif est d'apprendre une frontière de décision qui sépare les différentes classes dans l'espace des caractéristiques. Les algorithmes de classification les plus courants sont les k-voisins les plus proches, les machines à vecteurs de support et les réseaux neuronaux.

Le choix de l'algorithme et la qualité des résultats dépendent de plusieurs facteurs, tels que la taille et la complexité de l'ensemble de données, le nombre de caractéristiques d'entrée et la relation entre les caractéristiques d'entrée et la sortie. Il est crucial de prétraiter et de nettoyer soigneusement les données avant d'entraîner le modèle, et d'évaluer régulièrement les performances du modèle pour s'assurer qu'il se généralise bien aux nouvelles données [173].

En conclusion, l'apprentissage supervisé est une technique puissante pour résoudre les problèmes dont le résultat souhaité est connu. Il permet de développer des modèles capables de faire des prédictions précises sur la base de caractéristiques d'entrée, et a été appliqué à un large éventail de problèmes dans divers domaines, tels que le traitement du langage naturel, la vision par ordinateur et les systèmes de recommandation.

3.2 Les algorithmes d'apprentissage non supervisé

L'apprentissage non supervisé est un type d'apprentissage automatique dans lequel l'algorithme ne reçoit pas de données d'entraînement étiquetées. Au lieu de cela, l'algorithme ne reçoit que les caractéristiques d'entrée et doit trouver des modèles ou une structure dans les données par lui-même. Contrairement à l'apprentissage supervisé, l'objectif de l'apprentissage non supervisé n'est pas de faire des prédictions, mais de découvrir la structure sous-jacente des données [174].

Les algorithmes d'apprentissage non supervisé peuvent être classés en deux catégories : le clustering et la réduction de la dimensionnalité [174].

- Les algorithmes de clustering visent à répartir les données en groupes (également appelés clusters) sur la base de la similarité. L'objectif est de trouver

des groupes de points de données similaires qui peuvent être utilisés pour une analyse ultérieure ou pour mieux comprendre la structure des données. Les algorithmes de clustering les plus courants sont les k-means, le clustering hiérarchique et le clustering basé sur la densité. Par exemple, dans la segmentation du marché, le clustering peut être utilisé pour regrouper les clients en fonction de leurs habitudes d'achat, ce qui peut ensuite être utilisé pour adapter les campagnes de marketing ou pour concevoir de nouveaux produits destinés à des segments spécifiques.

- Les algorithmes de réduction de la dimensionnalité, quant à eux, visent à représenter des données à haute dimension dans un espace à plus faible dimensionnalité tout en préservant autant que possible la structure et l'information. Cela peut être utile pour visualiser des données hautement dimensionnelles, réduire le bruit ou améliorer les performances des algorithmes d'apprentissage supervisé. Les algorithmes courants de réduction de la dimensionnalité comprennent l'analyse en composantes principales (ACP), le t-SNE et les autoencodeurs.

Il est important de noter que les algorithmes d'apprentissage non supervisé peuvent souvent produire plusieurs solutions qui sont valides et significatives, et le choix de l'algorithme et du nombre de clusters dépend de la nature des données et du problème à résoudre. Il est également courant d'utiliser l'apprentissage non supervisé comme étape de prétraitement de l'apprentissage supervisé, par exemple en réduisant la dimensionnalité des données avant de former un classificateur.

En conclusion, l'apprentissage non supervisé est une technique puissante pour découvrir la structure et les modèles dans les données où les informations étiquetées ne sont pas disponibles. Il a été appliqué à un large éventail de problèmes dans divers domaines, tels que le traitement des images, le traitement du langage naturel et la détection des anomalies. Les algorithmes d'apprentissage non supervisé sont un outil essentiel pour l'exploration et l'analyse des données, et ils jouent un rôle crucial dans de nombreux pipelines d'apprentissage automatique.

3.3 L'apprentissage par renforcement

L'apprentissage par renforcement est un type d'apprentissage automatique qui se concentre sur la formation des agents à prendre des décisions dans un en-

vironnement en maximisant un signal de récompense. Dans l'apprentissage par renforcement, un agent interagit avec un environnement, reçoit un retour d'information sous la forme de récompenses ou de pénalités, et met à jour sa stratégie de prise de décision en fonction de ce retour d'information [175].

L'idée centrale de l'apprentissage par renforcement est d'entraîner un agent à trouver une séquence d'actions qui maximise un signal de récompense au fil du temps. C'est ce qu'on appelle la fonction de récompense, qui définit l'objectif de l'agent. L'agent entreprend des actions dans l'environnement, et l'environnement répond par un signal de récompense qui indique la performance de l'agent par rapport à son objectif.

L'apprentissage par renforcement peut être utilisé pour résoudre un large éventail de problèmes, allant des jeux au contrôle de robots et de voitures autonomes. Voici quelques applications courantes de l'apprentissage par renforcement :

- Le jeu : L'apprentissage par renforcement a été utilisé pour développer des agents de jeu capables de battre des joueurs humains à des jeux complexes comme les échecs, le Go et le poker.
- La robotique : L'apprentissage par renforcement peut être utilisé pour entraîner des robots à effectuer des tâches, telles que la saisie d'objets ou la marche.
- Systèmes de contrôle : L'apprentissage par renforcement peut être utilisé pour contrôler des systèmes, tels que les réseaux électriques, les systèmes de circulation et les voitures à conduite autonome.

Les algorithmes d'apprentissage par renforcement peuvent être classés en trois grandes catégories : les méthodes basées sur la valeur, les méthodes basées sur la politique et les méthodes basées sur [176].

- Les méthodes basées sur la valeur estiment la récompense attendue à long terme pour chaque état ou action et utilisent cette information pour sélectionner les actions. L'algorithme basé sur la valeur le plus populaire est le Q-Learning, qui utilise une table Q pour stocker la récompense estimée pour chaque paire état-action.

- Les méthodes basées sur les politiques apprennent directement une politique, qui est une correspondance entre les états et les actions. Les méthodes basées sur la politique sont bien adaptées aux problèmes où l'espace d'état est vaste ou continu, et où la politique optimale n'est pas facilement représentée sous forme de tableau.
- Les méthodes basées sur un modèle apprennent un modèle de l'environnement, qui peut être utilisé pour simuler les effets des actions et faire des prédictions sur les récompenses futures. Les méthodes basées sur les modèles sont souvent plus efficaces en termes d'échantillonnage que les méthodes basées sur les valeurs et les politiques, mais elles nécessitent davantage de ressources informatiques.

En conclusion, l'apprentissage par renforcement est une technique puissante pour résoudre les problèmes où un agent doit apprendre à prendre des décisions dans un environnement en fonction d'un signal de récompense. L'apprentissage par renforcement a été appliqué à un large éventail de problèmes et a donné des résultats remarquables dans des domaines tels que la robotique, les jeux et les systèmes de contrôle. À mesure que les algorithmes d'apprentissage par renforcement continuent de progresser, ils sont susceptibles de jouer un rôle de plus en plus important dans l'intelligence artificielle et l'apprentissage automatique.

Dans le contexte de notre étude, l'algorithme Bayésien Naïf sera utilisé pour classer les objets d'apprentissage sur la base de leurs métadonnées. Les métadonnées peuvent inclure des informations telles que la date de création, l'auteur, le format, la taille, etc. En réduisant les dimensions des métadonnées et en utilisant l'algorithme Bayésien Naïf, notre tâche consiste à classer avec précision les objets d'apprentissage dans différentes catégories en fonction de leurs caractéristiques. Il convient de noter que les performances de l'algorithme bayésien naïf peuvent être affectées par le choix des caractéristiques, la présence de caractéristiques non pertinentes et la qualité des données. Il est donc important de sélectionner soigneusement les caractéristiques, de prétraiter les données et d'évaluer les performances de l'algorithme au moyen de techniques de validation appropriées afin de garantir l'obtention de résultats précis et fiables.

L'algorithme bayésien naïf fonctionne en faisant des prédictions basées sur les probabilités d'appartenance des caractéristiques à chaque classe. Il calcule la pro-

babilité de chaque classe en fonction des caractéristiques et choisit la classe ayant la probabilité la plus élevée comme prédiction. L'algorithme est appelé "naïf" car il part de l'hypothèse simplificatrice que les caractéristiques sont indépendantes les unes des autres, ce qui n'est pas toujours le cas dans les données du monde réel [177]. Malgré cette limitation, les algorithmes bayésiens naïfs restent un choix populaire pour la classification des textes et le filtrage du spam en raison de leur simplicité, de leur efficacité et de leurs bonnes performances. Dans notre cas, le Bayésien Naïf sera utilisé comme suit :

$$P(\text{class}|\text{object}) = \frac{P(\text{class}) \cdot P(\text{object}|\text{class})}{P(\text{object})} \quad (\text{Chapitre IV.1})$$

Où :

$P(\text{classe}|\text{LO})$: Désigne la probabilité que même un objet d'apprentissage O appartienne à une classe C.

$P(\text{LO})$: Décrit la probabilité qu'un objet d'apprentissage puisse exister.

$P(\text{classe})$: Calcule la probabilité dans l'ensemble du groupe sur la base du nombre total d'objets apprenants dans toutes les sections.

$P(\text{LO}|\text{classe})$: Représente la probabilité d'une classe spécifique d'apprentissage d'objet, et l'objet peut être modélisé comme des ensembles de termes, ainsi $p(\text{objet}|\text{classe})$ peut être imprimé comme :

$$p(\text{LO}|\text{class}) = \prod_i p(\text{LO}_i|\text{class}) \quad (\text{Chapitre IV.2})$$

Donc :

$$p(\text{class}|\text{LO}) = p(\text{class}) \prod_i p(\text{LO}_i|\text{class}) \quad (\text{Chapitre IV.3})$$

Où :

$P(\text{O}_i)$: La probabilité que l'objet i d'un objet d'apprentissage donné apparaisse dans une classe C, qui peut être déterminée comme suit :

$$p(\text{LO}_i|\text{class}) = \frac{T_{ct} + \lambda}{N_c + \lambda V} \quad (\text{Chapitre IV.4})$$

Où :

T_{ct} : Le nombre de fois où le nom du terme de l'objet apparaît dans la catégorie C

N_c : Le nombre de termes dans la catégorie C

V : L'échelle du vocabulaire de la table

λ : Une constante positive, normalement 1 ou 0,5 pour éviter une probabilité nulle.

4 Travaux relatifs aux objets d'apprentissage dans les systèmes d'apprentissage en ligne

Le domaine de la classification des objets d'apprentissage dans les systèmes d'apprentissage en ligne fait l'objet de recherches approfondies depuis plusieurs années. Les chercheurs ont utilisé diverses techniques de classification pour classer ses objets d'apprentissage, les rendant ainsi plus accessibles et plus faciles à rechercher.

Historiquement, en 2011, de multiples études ont été menées sur la classification des objets d'apprentissage en utilisant différentes ressources et propriétés [178]. Ces études visaient à identifier les méthodes les plus efficaces pour classer les objets d'apprentissage. Par exemple, en 2011, Batista et al [179] ont proposés un système qui permet la récupération et la classification des LO et fournit un soutien individualisé pour la sélection des matériaux d'apprentissage afin de faire le choix le plus approprié parmi de nombreuses alternatives. En plus de cette étude, d'autres articles de recherche significatifs sur la classification des objets d'apprentissage dans les systèmes d'apprentissage en ligne ont été publiés. Par exemple, en 2012, les auteurs López et al [180] ont publié un article intitulé "A model for multi-label classification and ranking of learning objects", qui proposait une approche qui utilise des méthodes de classification multi-label pour la recherche d'objets d'apprentissage étiquetés (LOs) par des métadonnées d'objets

d'apprentissage (LOM). Plus précisément, le modèle offre une méthodologie qui illustre la tâche de mapping multi-label des LOs en types de requêtes à travers un espace multi-label émergent, et qui peut améliorer le premier choix des apprenants ou des enseignants. En 2019, les auteurs ont introduit un nouveau concept de classification des objets d'apprentissage basé sur la mémoire à long terme (LSTM) et l'approche de classification Random Forest [181]. Dans cet article, les auteurs donnent un aperçu de l'application des méthodes d'exploration de données dans le processus d'apprentissage en ligne avec l'apprentissage basé sur le Web. Cependant, avec les paradigmes de classification multi-labels, seuls neuf articles de la littérature ont développé des LO dans les systèmes d'apprentissage en ligne.

Ces articles proposent une variété d'approches pour incorporer les corrélations caractéristiques-étiquettes dans les métadonnées des objets d'apprentissage afin d'améliorer simultanément l'accessibilité et la recommandation. Ces travaux se concentrent sur la manière de modifier les classifieurs formels à étiquette unique pour les étiquettes multiples. Par exemple, certaines études ont utilisé la catégorisation hiérarchique multi-labels des métadonnées avec des méthodes d'apprentissage automatique pour améliorer la recherche et le classement des ressources pédagogiques. Les [182] étudient la catégorisation hiérarchique multi-labels dans le contexte des systèmes de recommandation. Ils proposent une catégorisation hiérarchique multi-label des métadonnées avec une méthode d'apprentissage automatique pour améliorer la recherche et la classification des ressources éducatives. De plus, cette étude contribue aux recherches précédentes en fournissant un ensemble de données hiérarchiques multi-labels sur les objets d'apprentissage dans un format approprié.

Dans l'ensemble, la recherche dans ce domaine a contribué de manière significative au développement de méthodes plus efficaces et plus précises pour la classification des objets d'apprentissage dans les systèmes d'apprentissage en ligne. Ces méthodes sont cruciales pour améliorer l'accessibilité et la facilité de recherche des ressources éducatives, permettant aux apprenants de trouver plus facilement du matériel pédagogique pertinent.

Le tableau 11 présente une revue des articles de recherche significatifs qui ont adopté des méthodes de MLC avec le contenu des objets d'apprentissage en

utilisant les métadonnées des LO.

Référence	Contribution
González et al [183]	Classification automatique des objets d'apprentissage pour réduire le nombre de caractéristiques utilisées.
Batista et al. (2011) [179]	Un système de classification multi-label des objets d'apprentissage.
López et al (2012) [180]	Un modèle pour la classification multi-label et le classement des objets d'apprentissage
Carrillo et al (2013) [184]	Classification multi-label pour les systèmes de recommandation
Birzniece et al (2014) [185]	Application d'un système de classification interactif à la comparaison des cours d'études universitaires
Aldreies et al (2016) [186]	Évaluation comparative de quatre algorithmes de classification multi-label dans la classification des objets d'apprentissage
Selvan et al (2019)[187]	Classification multi-label des objets d'apprentissage à l'aide de techniques d'apprentissage profond (Deep Learning)
Zhao et al (2019) [181]	Une classification efficace des objets d'apprentissage basée sur la mémoire à long terme et la classification par forêt aléatoire
Rani et al (2020) [188]	Classification multi-label des objets d'apprentissage à l'aide d'algorithmes d'apprentissage automatique

TABLE 11 : Résumé des études significatives qui ont appliqué les méthodes de la CML sur les LO

5 Notre approche

5.1 La classification des objets d'apprentissages basée sur les techniques de Clustering et Feature Selection

Nous avons commencé par l'idée que les termes sémantiquement similaires sont utilisés dans des contextes similaires, ce qui nous a amenés à réfléchir à la manière de manipuler les ressources d'apprentissage en ligne telles que les cours, les vidéos et le matériel pédagogique de manière flexible. Nous nous sommes également demandé comment explorer et réutiliser efficacement ces ressources dans différents domaines.

Pour répondre à la première question, nous avons adopté la philosophie de l'apprentissage par l'objet, ce qui implique d'analyser et de schématiser chaque ressource selon la normalisation des objets d'apprentissage de l'IEEE, comme les métadonnées, le taxon, l'identifiant, etc. Pour explorer ces métadonnées, notre intérêt s'est porté sur les techniques de web mining et l'algorithme Fuzzy C-Means Clustering. Les techniques de web mining désignent l'ensemble des techniques visant à explorer, traiter et analyser de grandes masses d'informations liées à l'activité du web. L'algorithme Fuzzy C-Means Clustering divise les données numériques en clusters.

Pour approfondir cette idée, l'algorithme de clustering Fuzzy C-Means (FCM) est une technique utilisée pour regrouper des points de données similaires en clusters. Dans le contexte des objets d'apprentissage, nous pouvons considérer les métadonnées de chaque objet d'apprentissage comme un point de données, où les mots contenus dans les métadonnées représentent les caractéristiques du point de données. En appliquant le clustering FCM à ces caractéristiques de métadonnées, nous pouvons regrouper des objets d'apprentissage similaires sur la base de la similarité de leurs métadonnées. Cela nous permet de mieux organiser et classer les objets d'apprentissage, ce qui facilite leur recherche et leur réutilisation dans différents contextes. De plus, en utilisant le clustering FCM, nous pouvons calculer la similarité entre différents objets d'apprentissage, ce qui nous permet de recommander des objets d'apprentissage similaires aux utilisateurs en fonction de leurs interactions précédentes avec le système. Cela peut conduire à une

expérience d'apprentissage plus personnalisée et plus efficace pour l'utilisateur.

Pour répondre à la deuxième question concernant la réutilisation efficace des ressources dans différents contextes, nous avons exploré l'utilisation de la classification multi-labels. Cette approche permet de classer les ressources avec plusieurs étiquettes, ce qui permet de les réutiliser dans différents domaines. La méthode consiste à construire une structure arborescente hiérarchique de catégories, chaque catégorie représentant une étiquette. La structure arborescente permet une approche de classification multi-labels où chaque objet d'apprentissage peut être affecté à une ou plusieurs étiquettes en fonction de la catégorie à laquelle il appartient. Cette approche s'est avérée efficace pour améliorer la précision et l'efficacité de la classification et de la réutilisation des objets d'apprentissage dans différents domaines. Notre méthodologie est composée de deux étapes [189], comme indiqué dans la figure 39. Dans la première phase, connue sous le nom de phase d'apprentissage, notre algorithme génère la liste des objets d'apprentissage sur la base de l'extraction effectuée par les techniques de web mining et de la similarité exécutée par FCM. Dans la deuxième phase, notre algorithme classe les objets d'apprentissage listés dans la première phase selon le degré de similarité et les étiquette avec la classification Multi-label afin de réutiliser les objets et de proposer les meilleures recommandations aux apprenants. Avec cette approche, nous attendons deux objectifs : l'amélioration de la similarité des termes analysés et extraits par le web mining et l'optimisation des méta-recommandations.

La figure 39 illustre la conception que nous proposons pour classer les objets d'apprentissage à l'aide de l'algorithme de classification multi-label en combinaison avec les techniques Fuzzy C-Means et Web-Mining. Cette conception a été développée dans le but d'améliorer la recommandation d'objets d'apprentissage aux apprenants dans les systèmes d'apprentissage en ligne.

La première étape de la conception implique la collecte et l'organisation des objets d'apprentissage dans une base de connaissances, qui contient des métadonnées et d'autres informations pertinentes sur les objets. L'étape suivante consiste à utiliser des techniques de Web-Mining pour extraire des informations de la base

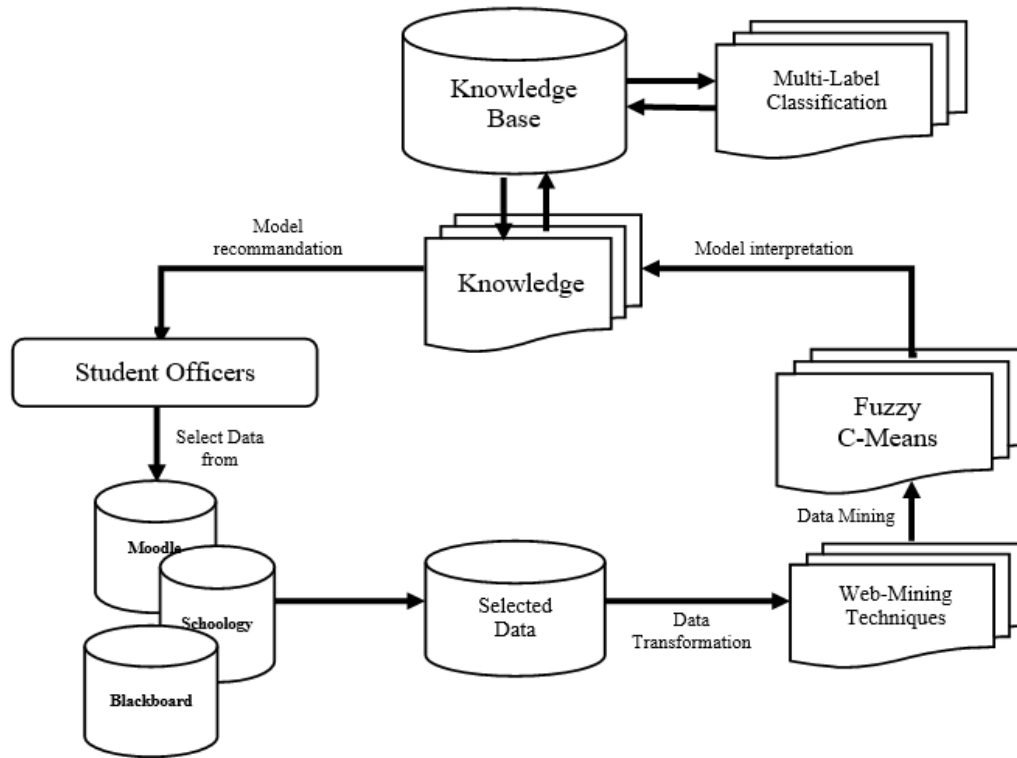


FIGURE 39 : Illustration de notre conception avec l'algorithme de classification multi-label combiné aux techniques Fuzzy C-Means et Web-Mining.[189]

de connaissances.

Une fois que les informations nécessaires ont été extraites, l'algorithme de classification multi-labels est appliqué pour attribuer plusieurs étiquettes à chaque objet d'apprentissage. Cette approche permet aux objets d'apprentissage d'être plus flexibles et réutilisables dans différents contextes.

L'algorithme Fuzzy C-Means est ensuite appliqué pour calculer la similarité entre les objets d'apprentissage, ce qui réduit la quantité de données en ignorant celles qui sont inappropriées. Il s'agit d'une étape cruciale du processus, car elle améliore les performances du système en filtrant les données non pertinentes.

Les résultats de la classification multi-labels et de Fuzzy C-Means sont combinés pour générer une liste d'objets d'apprentissage recommandés qui sont les plus pertinents pour les besoins de l'apprenant. L'expérience d'apprentissage est ainsi personnalisée et efficace, puisque les apprenants ne reçoivent que les objets

d'apprentissage les plus pertinents. La conception vise à améliorer la recommandation des objets d'apprentissage aux apprenants, dans le but ultime d'améliorer l'expérience d'apprentissage et les résultats de l'apprentissage.

5.2 Évaluation et discussion

5.2.1 Corpus utilisé

Dans l'étude, les objets d'apprentissage ont été sélectionnés en fonction de diverses catégories, telles que les mathématiques, l'informatique, la chimie, la physique et la biologie. Les objets d'apprentissage comprennent des présentations, des pratiques et des modèles conceptuels, qui sont différents types de ressources couramment utilisées dans l'apprentissage en ligne. Le tableau 12 fournit des descriptions détaillées des objets d'apprentissage utilisés dans l'étude, y compris le nom de la plateforme à partir de laquelle les données ont été collectées, le type d'objet d'apprentissage, la catégorie et une brève description. Ces informations sont utiles pour comprendre les types de ressources qui ont été analysées et comment elles ont été classées.

En sélectionnant des données provenant de plusieurs plateformes et d'une variété de catégories, l'étude a pu fournir une analyse complète des ressources d'apprentissage en ligne et de leur potentiel pour une exploration et une réutilisation efficace dans différents domaines.

Type de LO	Explication	Exemple
Objet de présentation	Transmettre des informations spécifiques sur un sujet précis.	Théories du système d'apprentissage en ligne
Objet de pratique	S'entraîner avec un Feedback	Quiz, exercices, QCM.
Modèle conceptuel	Résumé de certains sujets avec des concepts connexes	Représentation visuelle des phénomènes

TABLE 12 : Une description des objets d'apprentissage utilisés dans notre recherche

Les plateformes auprès desquelles les données sont collectées sont les ensembles de données Moodle, Blackboard et Schoology.

- Moodle est une plateforme d'apprentissage gratuite et open-source. Elle offre des solutions spécifiques à divers besoins éducatifs, comme Moodle App, Moodle Education, Moodle Net, Moodle Workplace, et des intégrations certifiées.
- Blackboard est l'un des noms les plus populaires dans le domaine de l'apprentissage numérique. La plate-forme fournit les fonctions de gestion de l'apprentissage de base. Elle peut également gérer les classes en ligne et mixtes[46].
- La plate-forme Schoology vise à fournir tous les outils dont les professeurs ont besoin pour concevoir des leçons, communiquer avec leurs étudiants et collaborer avec d'autres éducateurs [190].

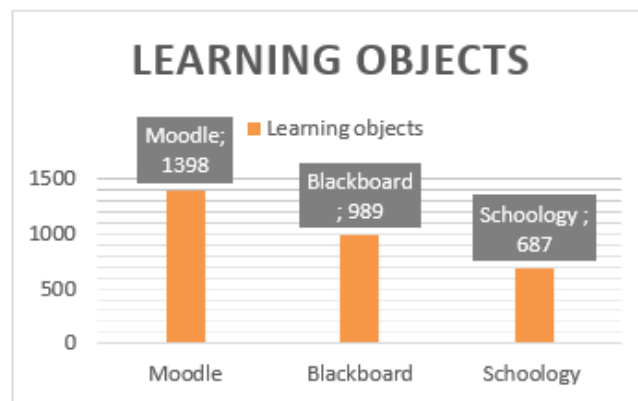


FIGURE 40 : le nombre d'objets d'apprentissage pour les ensembles de données Moodle, Blackboard et Schoology. [189]

5.2.2 Procédure expérimentale et mesure de la performance

Pour évaluer la performance de l'approche présentée dans notre étude, nous avons utilisé un ensemble d'indicateurs sous forme de règles mathématiques, tels que l'exactitude de la classification, la précision, le rappel et la mesure F1. Ces indicateurs sont couramment utilisés pour évaluer la performance de tout système proposé par rapport aux systèmes existants.

- La précision de classification est une mesure de la fréquence à laquelle le modèle prédit correctement la classe d'une instance. Elle est calculée comme

le rapport entre le nombre de prédictions correctes et le nombre total de prédictions. La précision est la fraction de vrais positifs (instances correctement prédites) parmi les instances prédites comme positives. Elle mesure la précision du modèle lorsqu'il prédit une classe positive. La précision est calculée comme le rapport entre les vrais positifs et la somme des vrais positifs et des faux positifs. La précision de la classification montre combien de prédictions sont correctes.

$$Accuracy = \frac{Number of correct predictions}{(Number of all predictions)} \quad (\text{Chapitre IV.5})$$

Dans certaines situations, la précision constitue une bonne mesure, alors que dans d'autres, elle semble insuffisante. Par exemple, une précision de prédiction de 94 indique que 94 échantillons sur 100 ont été anticipés avec succès sans que l'on sache au préalable quelles tâches sont correctement prédites. La précision mesure la qualité de notre modèle lorsque la prédiction est positive.

$$Precision = \frac{TP}{TP + FP} \quad (\text{Chapitre IV.6})$$

Avec :

TP (True positive) : Prédire qu'un objet positif est positif (ok).

FP (False positive) : Prédire qu'un objet négatif est positif (pas bon).

- Le rappel, également appelé sensibilité, est la fraction de vrais positifs parmi les instances qui appartiennent à la classe positive. Il mesure la capacité du modèle à identifier correctement les instances positives. Le rappel est calculé comme le rapport entre les vrais positifs et la somme des vrais positifs et des faux négatifs.

Le rappel mesure la capacité de notre modèle à prédire correctement les classes positives.

$$Recall = \frac{TP}{TP + FN} \quad (\text{Chapitre IV.7})$$

Avec : FN (Faux négatifs) : Prédire une classe positive comme négative

- La mesure F1 est la moyenne harmonique de la précision et du rappel. Elle combine la précision et le rappel en un seul score et fournit une évaluation plus équilibrée des performances du modèle.

La mesure F1 est calculée comme suit :

$$\text{F1 measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (\text{Chapitre IV.8})$$

Nous avons utilisé ces indicateurs pour évaluer les performances de notre approche proposée et la comparer aux approches existantes dans la littérature.

5.2.3 Réduction de la dimensionnalité

Le critère de sélection que nous avons mentionné vise principalement à identifier le modèle le plus approprié pour suggérer des objets d'apprentissage pertinents. À cet égard, le critère est utilisé pour déterminer le meilleur classificateur qui peut classer rapidement et précisément les objets d'apprentissage en fonction de leurs caractéristiques, telles que le sujet, le niveau de difficulté, le style d'apprentissage, etc. Pour atteindre cet objectif, les techniques de classification sont particulièrement utiles. Ces techniques permettent au système d'analyser de grandes quantités de données, d'ignorer les informations non pertinentes et de regrouper les objets d'apprentissage similaires. Ce faisant, le système peut identifier les objets d'apprentissage les plus appropriés pour chaque apprenant en fonction de ses besoins et préférences d'apprentissage individuels. Les avantages de l'utilisation d'un bon classificateur sont nombreux. Par exemple, il permet au système de traiter rapidement de grandes quantités de données, d'identifier les objets d'apprentissage les plus pertinents et de faire des recommandations aux apprenants en temps réel. Cela permet d'économiser du temps et des efforts, tant pour les apprenants que pour les instructeurs, et de s'assurer que les apprenants reçoivent le matériel d'apprentissage le plus approprié pour répondre à leurs besoins.

Dans notre approche, nous visons à évaluer l'efficacité de l'approche que nous proposons pour classer et suggérer des objets d'apprentissage. Pour ce faire, nous

prévoyons de comparer les performances de notre approche avec les algorithmes traditionnels d'apprentissage automatique tels que SVM et Naive Bayesian. Nous utiliserons la précision (PR), le rappel (RE) et la mesure F1 comme indicateurs de performance pour évaluer la précision de notre approche proposée. Ces indicateurs seront utilisés pour évaluer la performance de notre approche par rapport aux algorithmes d'apprentissage automatique traditionnels. Nous utiliserons Python comme plateforme logicielle pour notre évaluation. Python a un support intégré pour un large éventail d'algorithmes d'apprentissage automatique, y compris SVM et Naive Bayesian. Notre évaluation commencera par tester les performances des données à l'aide de SVM. Nous testerons ensuite les mêmes données en utilisant SVM combiné à la classification multi-label et FCM. De même, nous testerons les données à l'aide de Naive Bayesian (NB) et NB combiné à la classification multi-label et FCM. En comparant les performances de l'approche proposée avec celles des algorithmes traditionnels d'apprentissage automatique, nous espérons démontrer l'efficacité de notre approche pour classer et suggérer des objets d'apprentissage. Les résultats de notre évaluation pourraient potentiellement améliorer l'efficacité et la précision des systèmes de gestion de l'apprentissage, car notre approche permet des recommandations plus personnalisées et plus précises de matériel d'apprentissage pour les apprenants.

Les résultats de l'évaluation démontrent que l'approche proposée est plus performante pour la classification des objets d'apprentissage que les autres algorithmes d'apprentissage dans différents corpus. Les résultats des différents classificateurs utilisés dans l'évaluation sont présentés dans le tableau 13. Les résultats calculés des différents classificateurs confirment que les résultats du SVM combiné au clustering en logique floue et à la classification multi-label pour la réduction des données sont généralement meilleurs que le NB et ses variantes. Cela montre que l'approche proposée peut réduire efficacement les données et améliorer la précision de la classification et de la suggestion des objets d'apprentissage.

DataSet	Moodle			Blackboard			Schoolary		
	PR	RE	F1	PR	RE	F1	PR	RE	F1
SVM	90,03	53,04	87,5	94,52	67,87	89,1	89,45	51,87	89,01
NB	84,19	45,90	77,3	83,22	58,43	79,5	87,16	49,67	81,23
SVM + Fuzzy C-Means+ multi-label	92,07	58,81	91,4	95,52	72,87	90,01	89,89	57,43	90,98
NB+ Fuzzy C-Means+ multi-label	87,37	49,56	83,1	85,01	64,94	79,9	87,34	51,25	90,01

TABLE 13 : Comparaison de différentes méthodes d'apprentissage automatique en utilisant les méthodes FCM et MLC

6 Conclusion

Ce chapitre aborde le problème de la classification des objets d'apprentissage à l'aide de la classification multi-labels et des Fuzzy C-Means afin d'améliorer leur recommandation aux apprenants et de faciliter leur manipulation et leur partage par les systèmes d'apprentissage en ligne. Pour ce faire, nous avons d'abord organisé toutes les ressources d'apprentissage en ligne selon la norme IEEE Learning Object Standard et utilisé des techniques d'exploration du Web pour les explorer, en nous concentrant sur leurs métadonnées. L'utilisation de la classification multi-label dans le système d'apprentissage en ligne vise à rendre tous les objets d'apprentissage réutilisables dans différents contextes. En outre, cette approche peut améliorer les performances des n-recommandations les plus importantes. La technique Fuzzy C-Means est principalement utilisée pour calculer la similarité entre les objets d'apprentissage et réduire les données massives en ignorant celles qui sont inappropriées. Pour évaluer l'efficacité de l'approche que nous proposons, nous avons utilisé des sources de données provenant de trois des meilleurs systèmes de plateforme au monde (Moodle, Blackboard et Schoology). Nous avons ensuite utilisé la précision, le rappel et la mesure F1, les indicateurs les plus connus dans le domaine de la classification et de la recommandation,

pour évaluer les performances de l'approche proposée.

Nous avons utilisé des bibliothèques préprogrammées Python pour comparer la mise en œuvre expérimentale d'algorithmes d'apprentissage automatique traditionnels tels que SVM et Naive Bayesian avec notre approche sur les données proposées. Les résultats de l'étude ont montré que notre approche proposée est plus performante que les techniques traditionnelles pour la classification des objets d'apprentissage.

Dans l'ensemble, notre approche proposée a le potentiel d'améliorer l'efficacité et la précision des systèmes d'apprentissage en ligne en fournissant des recommandations plus personnalisées et plus précises aux apprenants. En utilisant la classification multi-label et Fuzzy C-Means pour la réduction des données, notre approche est capable de gérer plus efficacement les grandes quantités de données que l'on trouve généralement dans les systèmes de gestion de l'apprentissage.

Conclusion générale et perspectives

Sommaire

1	Synthèse	174
2	Bilan des travaux	176
3	Perspectives	178

1 Synthèse

Au cours de cette recherche, nous avons exploré les différentes problématiques liées aux systèmes d'apprentissage en ligne avec l'objectif principal d'améliorer la qualité de l'apprentissage et de recommandation. Pour y parvenir, nous avons étudié plusieurs aspects techniques relatifs aux différents intervenants impliqués dans le processus d'apprentissage en ligne.

Sur le plan technique, Nous avons constaté que la plupart des systèmes d'apprentissage en ligne se basent sur une méthodologie de recommandation pré-établie, qui ne prend pas en compte les changements de l'environnement et les évolutions de l'utilisateur en temps réel. Pour remédier à cette problématique, nous avons proposé une nouvelle approche de recommandation basée sur l'ontologie et la classification. L'ontologie est une technique qui permet de décrire de manière formelle et structurée les connaissances d'un domaine particulier. Nous

avons utilisé cette technique pour créer une ontologie de l'apprentissage qui permet de décrire les concepts, les relations et les dépendances entre les différents éléments du domaine de l'apprentissage en ligne. La classification, quant à elle, permet de classer les éléments du domaine selon des critères prédéfinis. Nous avons utilisé cette technique pour classer les différents profils d'utilisateurs en fonction de leurs caractéristiques, de leurs préférences et de leurs besoins en matière d'apprentissage. En combinant ces deux techniques, nous avons créé un système de recommandation dynamique qui peut adapter ses suggestions en temps réel en fonction de l'évolution de l'environnement et de l'utilisateur, ainsi que les connaissances acquises par l'apprenant au fil du temps.

Un autre aspect que nous avons exploré dans cette thèse concerne la question de la personnalisation de l'apprentissage en ligne en fonction des caractéristiques techniques du système. Nous avons constaté que les systèmes d'apprentissage en ligne actuels sont souvent conçus selon une approche "taille unique", cela signifie que les systèmes d'apprentissage sont conçus pour répondre aux besoins d'un grand nombre d'apprenants sans prendre en compte les différences individuelles en termes de styles d'apprentissage, de niveaux de connaissances et d'expériences antérieures. Cette approche peut être inefficace pour certains apprenants qui ont des besoins ou des préférences spécifiques qui ne sont pas pris en compte dans le système d'apprentissage. Cette problématique est étroitement liée à la question des objets d'apprentissage, qui sont des éléments clés des systèmes d'apprentissage en ligne et qui peuvent avoir un impact significatif sur l'efficacité de l'apprentissage.

Dans ce contexte, nous avons entamé des recherches pour mieux comprendre les caractéristiques des objets d'apprentissage et leur rôle dans le processus d'apprentissage en ligne. Nous avons constaté que la manipulation des objets d'apprentissage peut influencer le processus de recommandation de bout en bout dans les systèmes d'apprentissage en ligne. Pour remédier à cette lacune, nous avons proposé des techniques basées sur l'analyse des données d'apprentissage pour identifier les caractéristiques clés des objets d'apprentissage, ce qui permet de mieux comprendre les besoins et les préférences des apprenants et de personnaliser l'apprentissage en ligne en conséquence.

2 Bilan des travaux

Au cours de notre parcours de recherche dans le cadre de notre thèse de doctorat, nous avons effectué plusieurs travaux portant sur différents aspects de l'apprentissage en ligne. Cela s'est traduit par la publication de quatre articles scientifiques indexés Scopus Q1, à savoir :

- "ERSDO : E-learning Recommender System based on Dynamic Ontology", publié dans "Education and Information Technologies", présente un système de recommandation pour l'apprentissage en ligne basé sur une ontologie dynamique. En utilisant des techniques d'exploration de données via des ontologies dynamiques, ce système propose des recommandations personnalisées aux apprenants en fonction de leur profil et de leurs préférences. Les résultats de cette recherche ont montré que l'utilisation de l'ontologie dynamique permet d'obtenir des recommandations plus pertinentes pour les utilisateurs, améliorant ainsi l'efficacité de l'apprentissage en ligne.
- "Multi-Label Classification of Learning Objects Using Clustering Algorithms Based on Feature Selection", publié par "International Journal of Emerging Technologies in Learning", aborde la classification des objets d'apprentissage en utilisant des algorithmes de regroupement et de sélection des caractéristiques. Dans cet article, nous avons proposé un modèle de classification qui permet de catégoriser les objets d'apprentissage en fonction de leurs caractéristiques communes. Les résultats de cette recherche ont montré que ce modèle de classification peut être efficacement utilisé pour améliorer la qualité de l'apprentissage, en fournissant des objets d'apprentissage plus pertinents pour les apprenants.
- "Fuzzy C-Means and Web Mining Methods for Learning Object Analysis" publié dans le journal "Emerging Science Journal". Dans cette contribution, nous avons proposé une méthode de traitement des données d'apprentissage en utilisant des algorithmes de clustering, logique floue et des techniques d'analyse de données Web. Cette nouvelle combinaison a permis d'identifier les caractéristiques clés des objets d'apprentissage pour une meilleure compréhension des besoins et des préférences des apprenants.

- "New Perspective of Learning objects in e-learning system", publié dans "International Journal of Information and Learning Technology", dans cette recherche nous avons présenté une nouvelle perspective sur les objets d'apprentissage dans les systèmes d'apprentissage en ligne. Nous avons souligné l'importance de prendre en compte l'aspect social de l'apprentissage en ligne en intégrant des outils de collaboration dans la conception des objets d'apprentissage.

Chacun de ces articles a exploré une problématique spécifique de l'apprentissage en ligne, tels que la recommandation d'objets d'apprentissage, la classification d'objets d'apprentissage, l'analyse de données Web et les perspectives novatrices de l'apprentissage en ligne. Nos résultats ont permis de proposer des solutions innovantes pour améliorer la qualité des systèmes d'apprentissage en ligne.

En outre, nous avons mené un travail d'investigation sur un article indexé Scopus Q4, qui a exploré la problématique d'adaptation des systèmes d'apprentissage en ligne dans les environnements universitaires. Cet article, intitulé :

" Adapting student learning to the e-learning environment", dans lequel nous avons proposé une réflexion sur la manière dont les enseignants peuvent adapter les environnements d'apprentissage en ligne pour mieux répondre aux besoins des étudiants. Nous avons notamment discuté la nécessité de fournir des feedbacks réguliers pour favoriser l'engagement et la motivation des étudiants. Cette contribution a pour objectif d'améliorer la qualité de l'expérience d'apprentissage en ligne et de favoriser la réussite des étudiants.

En plus de ces articles, nous avons présenté cinq communications orales lors de conférences indexées Springer. Ces présentations portaient sur des sujets tels que la sécurité des systèmes d'apprentissage en ligne, les méthodes de recommandation de cours, l'utilisation du Big data dans l'apprentissage en ligne, l'adaptation de l'apprentissage des étudiants à l'environnement en ligne et la manière de façonner l'apprentissage des étudiants pour un environnement d'apprentissage spécifique.

En résumé, notre parcours de recherche a abouti à des réalisations significatives dans le domaine de l'apprentissage en ligne, avec des contributions im-

portantes dans la conception de systèmes de recommandation personnalisés, la classification des objets d'apprentissage, l'analyse de données Web et la nouvelle perspective sur les objets d'apprentissage. Ces réalisations ont été présentées lors de conférences prestigieuses et témoignent de notre engagement dans la recherche scientifique.

3 Perspectives

Comme perspectives de nos travaux, nous envisageons d'élargir notre champ d'investigation pour inclure d'autres types d'objets d'apprentissage et d'autres aspects de la personnalisation de l'apprentissage en ligne, tels que la personnalisation des évaluations et des feedbacks. Nous prévoyons également d'explorer des techniques d'apprentissage automatique plus avancées pour améliorer les performances de nos approches de recommandation personnalisée. En outre, nous nous intéressons à l'impact de l'utilisation de nos approches de personnalisation sur la motivation, l'engagement et la rétention des apprenants. Ainsi, nous prévoyons d'explorer d'autres méthodes d'analyse de données pour améliorer la précision et la pertinence des recommandations personnalisées, d'approfondir notre compréhension des facteurs qui influencent l'apprentissage en ligne, tels que les interactions sociales, les émotions et la motivation, et d'élargir notre étude à d'autres domaines de l'apprentissage en ligne, tels que l'apprentissage collaboratif et l'apprentissage mobile. Enfin, nous envisageons de collaborer avec des acteurs de l'éducation et des entreprises pour mettre en œuvre des solutions concrètes basées sur nos travaux de recherche et contribuer ainsi à l'amélioration de l'efficacité et de la qualité de l'apprentissage en ligne.

Annexes

Exemple d'implémentation de l'approche proposée ERSDO

Dans cet exemple ERSDO crée une ontologie simple qui comprend des concepts tels que "cours vidéo", "quiz pratique", "article", "algèbre", "calcul" et "géométrie". cette ontologie développe une recommandation simple qui recommande des ressources basées sur les interactions de l'utilisateur avec l'ensemble de données.

Dans cet exemple, chaque table possède une clé primaire (PK), qui est un identifiant unique pour chaque enregistrement de la table. Les clés étrangères (FK) établissent les relations entre les tables :

- La table Mathématiques possède des clés étrangères vers les tables Vidéo-Lecture, Pratique et Article afin de relier les ressources à la matière mathématique.
- La table QuizAttempt possède une clé étrangère vers la table Practice pour relier les tentatives de quiz à la ressource pratique correspondante.
- Les tables VideoView et ArticleView possèdent des clés étrangères vers les tables VideoLecture et Article, respectivement, pour relier les vues à la ressource correspondante.

```
CREATE TABLE VideoLecture (  
    id INT PRIMARY KEY,  
    title VARCHAR(255),  
    length INT,  
    url VARCHAR(255),  
    description VARCHAR(255),  
    views INT,  
    likes INT,  
    dislikes INT  
);  
  
CREATE TABLE Practice (  
    id INT PRIMARY KEY,  
    title VARCHAR(255),  
    questions TEXT  
);  
  
CREATE TABLE Article (  
    id INT PRIMARY KEY,  
    title VARCHAR(255),  
    content TEXT  
);
```

FIGURE 41 : Création de la base de donnée

```
CREATE TABLE QuizAttempt (  
  id INT PRIMARY KEY,  
  timestamp DATETIME,  
  score INT,  
  quiz_id INT,  
  user_id INT,  
  FOREIGN KEY (quiz_id) REFERENCES Practice(id),  
  FOREIGN KEY (user_id) REFERENCES User(id)  
);  
  
CREATE TABLE User (  
  id INT PRIMARY KEY,  
  name VARCHAR(255),  
  email VARCHAR(255)  
);
```

```
CREATE TABLE VideoView (  
    id INT PRIMARY KEY,  
    timestamp DATETIME,  
    video_lecture_id INT,  
    user_id INT,  
    FOREIGN KEY (video_lecture_id) REFERENCES VideoLecture(id),  
    FOREIGN KEY (user_id) REFERENCES User(id)  
);  
  
CREATE TABLE ArticleView (  
    id INT PRIMARY KEY,  
    timestamp DATETIME,  
    article_id INT,  
    user_id INT,  
    FOREIGN KEY (article_id) REFERENCES Article(id),  
    FOREIGN KEY (user_id) REFERENCES User(id)  
);  
  
CREATE TABLE Mathematics (  
    id INT PRIMARY KEY,  
    lectures TEXT,  
    practices TEXT,  
    articles TEXT  
);
```

Les tables VideoView et ArticleView possèdent des clés étrangères vers les tables VideoLecture et Article, respectivement, pour relier les vues à la ressource correspondante.

Notez que la table "Mathématiques" n'est reliée à aucune autre table par une relation de clé étrangère. Au lieu de cela, elle contient trois colonnes, "lectures", "pratiques" et "articles", qui sont des listes d'identifiants correspondant aux enregistrements associés dans les tables VidéoLecture, Pratique et Article, respectivement. Cette conception permet de retrouver facilement toutes les ressources associées à un sujet particulier (dans ce cas, les mathématiques) sans avoir

à joindre plusieurs tables. Toutefois, elle nécessite un traitement supplémentaire pour extraire les données pertinentes de ces colonnes.

1 création d'une ontologie pour un domaine lié aux ressources d'apprentissage en ligne des mathématiques.

1. Définir l'espace de noms de l'ontologie et importer toutes les ontologies externes nécessaires :

```
@prefix : <http://example.com/math-ontology#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix owl: <http://www.w3.org/2002/07/owl#> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
  
@base <http://example.com/math-ontology> .  
  
<> rdf:type owl:Ontology ;  
    owl:imports <http://www.w3.org/2002/07/owl#> .
```

FIGURE 42 : Définition de l'espace de noms de l'ontologie et importation de toutes les ontologies externes nécessaires

2. Définir les classes de l'ontologie :

```
:Resource rdf:type owl:Class ;  
    rdfs:label "E-Learning Resource" .  
  
:Lecture rdf:type owl:Class ;  
    rdfs:subClassOf :Resource ;  
    rdfs:label "Video Lecture" .  
  
:Practice rdf:type owl:Class ;  
    rdfs:subClassOf :Resource ;  
    rdfs:label "Practice" .  
  
:Article rdf:type owl:Class ;  
    rdfs:subClassOf :Resource ;  
    rdfs:label "Article" .
```

FIGURE 43 : Définition des classes de l'ontologie

3. Définir les propriétés de l'ontologie :

```
:hasTitle rdf:type owl:ObjectProperty ;  
    rdfs:domain :Resource ;  
    rdfs:range xsd:string ;  
    rdfs:label "has title" .  
  
:hasLength rdf:type owl:DatatypeProperty ;  
    rdfs:domain :Lecture ;  
    rdfs:range xsd:int ;  
    rdfs:label "has length" .  
  
:hasQuestions rdf:type owl:DatatypeProperty ;  
    rdfs:domain :Practice ;  
    rdfs:range xsd:string ;  
    rdfs:label "has questions" .  
  
:hasContent rdf:type owl:DatatypeProperty ;  
    rdfs:domain :Article ;  
    rdfs:range xsd:string ;  
    rdfs:label "has content" .
```

FIGURE 44 : Définition des propriétés de l'ontologie

4. Définir les individus dans l'ontologie :

```
:Lecture1 rdf:type :Lecture ;
    :hasTitle "Introduction to Algebra" ;
    :hasLength 3600 ;
    :hasURL "https://example.com/lecture1" ;
    :hasDescription "This video lecture covers the basics of algebra, including va
    :hasViews 1000 ;
    :hasLikes 50 ;
    :hasDislikes 10 .

:Practice1 rdf:type :Practice ;
    :hasTitle "Algebraic Expressions Practice Set 1" ;
    :hasQuestions "1. Simplify  $2x + 3y - 4x - 5y$ . 2. Evaluate  $3x^2 + 2x - 1$  for x

:Article1 rdf:type :Article ;
    :hasTitle "An Introduction to Calculus" ;
    :hasContent "Calculus is a branch of mathematics that deals with rates of chan
```

FIGURE 45 : Définition des individus dans l'ontologie

5. Définir la hiérarchie des classes dans l'ontologie :

```
:Lecture rdfs:subClassOf :Resource .
:Practice rdfs:subClassOf :Resource .
:Article rdfs:subClassOf :Resource .
```

FIGURE 46 : la hiérarchie des classes dans l'ontologie

Il ne s'agit là que d'un exemple basique de création d'une ontologie pour un domaine lié aux ressources d'apprentissage en ligne des mathématiques.

2 Création des profils d'utilisateurs en fonction de leur interaction avec les ressources d'apprentissage en ligne :

```
import pandas as pd

# Load the user interaction data
user_data = pd.read_csv('user_data.csv')

# Group the interactions by user ID and resource type
user_resource_counts = user_data.groupby(['user_id', 'resource_type']).size().reset_index()

# Pivot the data to create a user-resource interaction matrix
user_resource_matrix = user_resource_counts.pivot(index='user_id', columns='resource_type', values='size')

# Normalize the interaction matrix to create user profiles
user_profiles = user_resource_matrix.div(user_resource_matrix.sum(axis=1), axis=0)

# Print the user profiles
print(user_profiles)
```

FIGURE 47 : Création des profils d'utilisateurs en fonction de leur interaction avec les ressources d'apprentissage en ligne

Dans cet exemple, "user-data.csv" est un fichier contenant des données sur les interactions des utilisateurs avec les ressources d'apprentissage en ligne. Chaque ligne du fichier représente une interaction unique et comprend l'identifiant de l'utilisateur, l'identifiant de la ressource, le type de ressource et l'horodatage.

Le code regroupe les interactions par identifiant d'utilisateur et par type de ressource, puis fait pivoter les données pour créer une matrice dans laquelle chaque ligne représente un utilisateur et chaque colonne un type de ressource. Les valeurs de la matrice représentent le nombre d'interactions que chaque utilisateur a eues avec chaque type de ressource.

3 Génération des recommandations pour un utilisateur en fonction de son profil et de l'ontologie :

```
import pandas as pd
import owlready2 as owl

# Load the user profile data
user_profiles = pd.read_csv('user_profiles.csv', index_col=0)

# Load the ontology
ontology = owl.get_ontology('ontology.owl').load()

# Get the classes and individuals from the ontology
classes = list(ontology.classes())
individuals = list(ontology.individuals())

# Create a dictionary to store the similarity scores
similarities = {}
```

```
# Calculate the similarity score between the user profile and each individual in t
for individual in individuals:
    similarity = 0
    for resource_type, weight in user_profiles.loc[0].iteritems():
        if weight > 0:
            class_ = ontology.search_one(iri='*#' + resource_type)
            if class_ in individual.is_a:
                similarity += weight
    similarities[individual.name] = similarity

# Sort the similarities dictionary in descending order
sorted_similarities = sorted(similarities.items(), key=lambda x: x[1], reverse=True)

# Get the top N recommendations
N = 10
recommendations = []
for i in range(N):
    if i >= len(sorted_similarities):
        break
    recommendation = sorted_similarities[i][0]
    recommendations.append(recommendation)

# Print the recommendations
print(recommendations)
```

FIGURE 48 : Génération des recommandations pour un utilisateur en fonction de son profil et de l'ontologie

Dans cet exemple, `user-profiles.csv` est un fichier contenant des profils d'utilisateurs générés à l'aide du code de l'exemple précédent. L'ontologie est chargée à l'aide de la bibliothèque `owlready2`, et les classes et les individus sont extraits de l'ontologie.

Le code calcule ensuite le score de similarité entre le profil de l'utilisateur et chaque individu de l'ontologie. Le score de similarité est calculé en additionnant les poids des types de ressources qui sont partagés entre le profil de l'utilisateur et l'individu. Les similarités sont stockées dans un dictionnaire, avec le nom de

l'individu comme clé et le score de similarité comme valeur.

Le dictionnaire est trié par ordre décroissant sur la base des scores de similarité, et les N meilleures recommandations sont sélectionnées. Les recommandations sont simplement les noms des personnes ayant les scores de similarité les plus élevés.

4 La liste des recommandations

```
import pandas as pd
import owlready2 as owl

# Load the user profile data
user_profiles = pd.read_csv('user_profiles.csv', index_col=0)

# Load the ontology
ontology = owl.get_ontology('ontology.owl').load()

# Get the classes and individuals from the ontology
classes = list(ontology.classes())
individuals = list(ontology.individuals())

# Create a dictionary to store the similarity scores
similarities = {}

# Calculate the similarity score between the user profile and each individual in the
for individual in individuals:
    similarity = 0
    for resource_type, weight in user_profiles.loc[0].iteritems():
        if weight > 0:
            class_ = ontology.search_one(iri='*#' + resource_type)
            if class_ in individual.is_a:
                similarity += weight
    similarities[individual.name] = similarity

# Sort the similarities dictionary in descending order
sorted_similarities = sorted(similarities.items(), key=lambda x: x[1], reverse=True)
```

```
# Get the top N recommendations
N = 10
recommendations = []
for i in range(N):
    if i >= len(sorted_similarities):
        break
    recommendation = sorted_similarities[i][0]
    recommendations.append(recommendation)

# Deliver the recommendations to the user
print("Here are your recommended e-learning resources:")
for recommendation in recommendations:
    individual = ontology.search_one(iri='*#' + recommendation)
    print("- " + individual.label[0])
```

FIGURE 49 : La liste des recommandations

Dans cet exemple, le code est similaire à celui de l'exemple précédent, avec l'ajout d'une simple interface de ligne de commande pour fournir les recommandations à l'utilisateur. Le code calcule d'abord les recommandations en utilisant la même méthode que précédemment, puis les imprime sur la console.

La sortie est formatée comme une simple liste de noms de ressources, précédée d'un message indiquant qu'il s'agit des ressources d'apprentissage en ligne recommandées par l'utilisateur.

Productions scientifiques

1 Publications

- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **ERSDO : Elearning recommender system based on dynamic ontology**. Education and Information Technologies, 2022, vol .27, no 6, p. 7549–7561.
- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **classification of learning objects using clustering algorithms based on feature selection**. International Journal of Emerging Technologies in Learning (iJET), 2022, vol. 17, no 20.
- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **Multi-Label Classification of Learning Objects in E-Learning Environment Based on Fuzzy C-Means Algorithm and Web Mining Techniques**. Emerging Science Journal, 2023, vol. 7, no 3.
- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **Adapting Student’s Learning to University Distance Learning : A Real Case Study**. Mathematical Statistician and Engineering Applications, 2023, vol. 71, no 4, p. 11050–11058.

2 Communications

- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **Big Data in E-learning : Literature Review and Challenges**. In : Kacprzyk, J., Ba-

las, V.E., Ezziyyani, M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2020). Advances in Intelligent Systems and Computing, vol 1417. Springer.

- Meryem Amane, Karima Aissaoui, et Mohammed Berrada. **A Multi-agent and Content-Based Course Recommender System for University E-learning Platforms**. In : Motahhir, S., Bossoufi, B. (eds) Digital Technologies and Applications. ICDTA 2021. Lecture Notes in Networks and Systems, vol 211. Springer.
- Karima Aissaoui, Meryem Amane, Mohammed Berrada et Mohammed Amine Madani .**A New Framework to Secure Cloud Based e-Learning Systems**. In Bennani, S., Lakhrissi, Y., Khaissidi, G., Mansouri, A., Khamli-chi, Y. (eds) WITS 2020. Lecture Notes in Electrical Engineering, vol 745. Springer, Singapore.
- Meryem Amane, Karima Aissaoui, et Mohammed Berrada **Adapting Students' Learning to the E-learning Environment : A Case Study of the Faculty of Letters and Human Sciences Dhar El Mahraz Fez**. In : The 5 International Conference on Recent Trends in Multi-Disciplinary Research (ICRTMDR-22) , 09 et 10 décembre 2022.
- Meryem Amane, Karima Aissaoui, et Mohammed Berrada **Shaping Students' Learning to a specific Learning Environment**. In : The International Conference on Digital Technologies and Applications, Janvier 27-28, 2023, Fez, Morocco.

Bibliographie

- [1] Linda Harasim. Learning theories : The role of epistemology, science, and technology. In Michael J Spector, Barbara B Lockee, and Marcus D. Childress, editors, *Learning, Design, and Technology : An International Compendium of Theory, Research, Practice, and Policy*, pages 1–39. Springer International Publishing, 2018.
- [2] Karima AISSAOUI. Collaboration des plateformes x-learning : Services et sécurité, 2017.
- [3] Steve Wheeler. e-learning and digital learning. In Norbert M. Seel, editor, *Encyclopedia of the Sciences of Learning*, pages 1109–1111. Springer US, 2012.
- [4] Paul Nicholson. A history of e-learning. In Baltasar Fernández-Manjón, Juan Manuel Sánchez-Pérez, Juan Antonio Gómez-Pulido, Miguel Angel Vega-Rodríguez, and José Bravo-Rodríguez, editors, *Computers and Education : E-Learning, From Theory to Practice*, pages 1–11. Springer Netherlands, 2007.
- [5] John Lee. History of computing in education. *International Federation for Information Processing Digital Library ; History of Computing in Education* ;, 145, 2006.
- [6] Andrea Santo-Sabato and Marta Vernaleone. From the first generation of distance learning to personal learning environments : An overall look. In

-
- Giovanni Vincenti, Alberto Bucciero, and Carlos Vaz de Carvalho, editors, *E-Learning, E-Education, and Online Training*, pages 155–158. Springer International Publishing, 2014.
- [7] Donald L Alderman, Lola Rhea Appel, and Richard T Murphy. PLATO and TICCIT : An evaluation of CAI in the community college. *Educational Technology*, 18(4) :40–45, 1978.
 - [8] Iuliana Dobre. Evaluation of student’s knowledge—an overview of the testing and assessment tools for e-assessment. pages 199–205, 2011.
 - [9] David R. Woolley. PLATO : The emergence of online community. In Judy Malloy, editor, *Social Media Archeology and Poetics*, page 0. The MIT Press, 2016.
 - [10] Carl Whithaus. The development of early computer-assisted writing instruction (1960–1978) : The double logic of media and tools. *Computers and the Humanities*, 38(2) :149–162, 2004.
 - [11] Donald L Alderman. Evaluation of the TICCIT computer-assisted instructional system in the community college. *ACM SIGCUE Outlook*, 13(3) :5–17, 1979. Publisher : ACM New York, NY, USA.
 - [12] Gail E Hawisher, Cynthia L Selfe, Charles Moran, and Paul J LeBlanc. *Computers and the teaching of writing in American higher education, 1979-1994 : A history*. Greenwood Publishing Group, 1996.
 - [13] Emilio Esposito and Michele Mastroianni. Information technology and personal computers : the relational life cycle. *Technovation*, 22(1) :41–50, 2002.
 - [14] Jonathan Nicholas Dron. Achieving self-organisation in network-based learning environments. 2002. Publisher : University of Brighton Brighton.
 - [15] Martin Ebner. E-learning 2.0= e-learning 1.0+ web 2.0 ? pages 1235–1239, 2007.
 - [16] Manuela Aparicio, Fernando Bacao, and Tiago Oliveira. An e-learning theoretical framework. *An e-learning theoretical framework*, (1) :292–307, 2016.

-
- [17] Mark Nichols. A theory for eLearning. *Journal of Educational Technology & Society*, 6(2) :1–10, 2003.
- [18] Charmaine Brooks and Susan Gibson. Professional learning in a digital age/l'apprentissage professionnel à l'ère numérique. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 38(2), 2012. Publisher : Canadian Network for Innovation in Education.
- [19] Sarah Guri-Rosenblit. 'distance education'and 'e-learning' : Not the same thing. *Higher education*, 49(4) :467–493, 2005.
- [20] Zlatko Bezovski and Subitcha Poorani. The evolution of e-learning and new trends. volume 6, pages 50–57. *Iiste*, 2016. Issue : 3.
- [21] Dongil Shin, En Sup Yoon, Kyung Yong Lee, and Euy Soo Lee. A web-based, interactive virtual laboratory system for unit operations and process systems engineering education : issues, design and implementation. *Computers & chemical engineering*, 26(2) :319–330, 2002.
- [22] Albert KW Wu and Chi Hong Leung. Evaluating learning behavior of web-based training (WBT) using web log. pages 736–737, 2002.
- [23] Cansu Cigdem Aydin and Guzin Tirkes. Open source learning management systems in e-learning and moodle. pages 593–600, 2010.
- [24] Jared Keengwe and David Georgina. The digital course training workshop for online learning and teaching. *Education and Information Technologies*, 17(4) :365–379, 2012.
- [25] Cary Roseth, Mete Akcaoglu, and Andrea Zellner. Blending synchronous face-to-face and computer-supported cooperative learning in a hybrid doctoral seminar. *TechTrends*, 57(3) :54–59, 2013.
- [26] Joseph D Novak. Meaningful learning : The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science education*, 86(4) :548–571, 2002.

-
- [27] William Watson and Sunnie Lee Watson. An argument for clarity : What are learning management systems, what are they not, and what should they become. 2007.
- [28] Priyavahani Subramanian, Nursyafeeka Zainuddin, Salem Alatawi, Tahmineh Javabdeh, and A Hussin. A study of comparison between moodle and blackboard based on case studies for better LMS. *Journal of Information Systems Research and Innovation*, 6 :26–33, 2014.
- [29] Suman Ninoriya, Pramila Chawan, and B Meshram. CMS, LMS and LCMS for eLearning. *Int J Comput Sci*, 8, 2011.
- [30] Morten Flate Paulsen. Online education systems : Discussion and definition of terms. *NKI distance education*, 202 :1–8, 2002.
- [31] Sultan Hammad Alshammari, M Bilal Ali, and Mohd Shafie Rosli. LMS, CMS and LCMS : The confusion among them. *Science International*, 30(3) :455–459, 2018.
- [32] Brit Toven-Lindsey, Robert A Rhoads, and Jennifer Berdan Lozano. Virtually unlimited classrooms : Pedagogical practices in massive open online courses. *The internet and higher education*, 24 :1–12, 2015.
- [33] Jane E Klobas, Bruce Mackintosh, and Jamie Murphy. The anatomy of MOOCs. In *Massive open online courses*, pages 1–22. Routledge, 2014.
- [34] Félix Buendía García and Antonio Hervás Jorge. Evaluating e-learning platforms through SCORM specifications. 2006.
- [35] Félix Buendía and Antonio Hervás. An evaluation framework for e-learning platforms based on educational standard specifications. pages 184–186, 2006.
- [36] Ioannis Kazanidis and Maya Satratzemi. *Towards the integration of adaptive educational systems with SCORM standard and authoring toolkits*. INTECH Open Access Publisher, 2009.
- [37] Edward R. Jones. Implications of SCORM and emerging e-learning standards on engineering education. In *2002 GSW Proceedings*, page 38744. ASEE Conferences, 2022-03.

-
- [38] David A Wiley. The learning objects literature. In *Handbook of research on educational communications and technology*, pages 345–353. Routledge, 2008.
- [39] Andrés Chiappe Laverde, Yasbley Segovia Cifuentes, and Helda Yadira Rincón Rodríguez. Toward an instructional design model based on learning objects. *Educational Technology Research and Development*, 55(6) :671–681, 2007.
- [40] Ana Casali, Claudia Deco, Agustín Romano, and Guillermo Tome. An assistant for loading learning object metadata : An ontology based approach. 2013.
- [41] Manuel Caeiro, Luis Anido, and Martin Llamas. A critical analysis of IMS learning design. In *Designing for Change in Networked Learning Environments*, pages 363–367. Springer, 2003.
- [42] David Griffiths, Josep Blat, Rocío Garcia, Hubert Vogten, and KL Kwong. Learning design tools. In *Learning design*, pages 109–135. Springer, 2005.
- [43] 1edtech learning design information model, 2003.
- [44] Barr Avron. IEEE learning technology standards committee (LTSC).
- [45] Carlo De Medio, Carla Limongelli, Filippo Sciarrone, and Marco Temperini. MoodleREC : A recommendation system for creating courses using the moodle e-learning platform. *Computers in Human Behavior*, 104 :106168, 2020.
- [46] Blackboard Inc. https://en.wikipedia.org/wiki/Blackboard_Learn. Dernier accès : 2016-11-09.
- [47] Chamilo. <https://chamilo.org/fr/>. Dernier accès : 2023-04-08.
- [48] Rajiv Ramnath and A. Kuriakose. Mobile technologies for teaching and learning. pages 645–662. 2015.
- [49] Yeonjeong Park. A pedagogical framework for mobile learning : Categorizing educational applications of mobile technologies into four types. *International Review of Research in Open and Distance Learning*, 12, 2011.

-
- [50] Chien Yu, Sang Joon Lee, and Ewing Carlos. Mobile learning : Trends, issues, and challenges in teaching and learning. pages 60–87. 2014.
 - [51] Mike Sharples and Daniel Spikol. Mobile learning. In *Technology Enhanced Learning : Research Themes*. 2017. Journal Abbreviation : Technology Enhanced Learning : Research Themes.
 - [52] Mike Sharples. The design of personal mobile technologies for lifelong learning. *Computers & Education*, 34(3) :177–193, 2000.
 - [53] Tereska Karran, Kambiz Madani, and George R Ribeiro-Justo. Self-learning and adaptive systems : The CODA approach. In *Software Defined Radio : Architectures, Systems and Functions*. John Wiley & Sons Ltd., 2003.
 - [54] José Sánchez-Prieto, Susana Olmos, and Francisco García-Peñalvo. *Mobile learning : Tendencies and lines of research*. 2013-11-14. Journal Abbreviation : ACM International Conference Proceeding Series Pages : 480 Publication Title : ACM International Conference Proceeding Series.
 - [55] Peng Gao, Jingyi Li, and Shuai Liu. An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. *Mobile Networks and Applications*, 26(5) :2123–2126, 2021.
 - [56] Azlinah Mohamed, Maryam Khanian Najafabadi, Yap Bee Wah, Ezzatul Akmal Kamaru Zaman, and Ruhaila Maskat. The state of the art and taxonomy of big data analytics : view from new big data framework. *Artificial Intelligence Review*, 53 :989–1037, 2020.
 - [57] Nada Elgendy and Ahmed Elragal. Big data analytics : a literature review paper. pages 214–227. Springer, 2014.
 - [58] Statista. Statista. <https://www.statista.com>. Dernier accès : 2022-11-09.
 - [59] Banica Logica and Radulescu Magdalena. Using big data in the academic environment. *Procedia Economics and Finance*, 33 :277–286, 2015.
 - [60] Elena Shoikova, Roumen Nikolov, Eugenia Kovatcheva, Boyan Jekov, and Lyubomir Gotsev. Big data framework overview. 55 :22–34, 2020.

-
- [61] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. The rise of “big data” on cloud computing : Review and open research issues. *Information systems*, 47 :98–115, 2015.
- [62] Michael Minelli, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics : emerging business intelligence and analytic trends for today’s businesses*, volume 578. John Wiley & Sons, 2013.
- [63] Fuad Bajaber, Radwa Elshaw, Omar Batarfi, Abdulrahman Altalhi, Ahmed Barnawi, and Sherif Sakr. Big data 2.0 processing systems : Taxonomy and open challenges. *Journal of Grid Computing*, 14 :379–405, 2016.
- [64] Lu Tan and Neng Wang. Future internet : The internet of things. volume 5, pages V5–376, 2010.
- [65] Xiaomeng Su. Introduction to big data. *IFUD1123*, 2017.
- [66] Stephen Kaisler, Frank Armour, J Alberto Espinosa, and William Money. Big data : Issues and challenges moving forward. pages 995–1004, 2013.
- [67] Grzegorz Arkit, Silva Robak, and Aleksandra Arkit. Applying e-learning systems for big data education. *Information Systems in Management*, 7(2) :85–96, 2018.
- [68] Lizhe Wang, Jie Tao, Rajiv Ranjan, Holger Marten, Achim Streit, Jingying Chen, and Dan Chen. G-hadoop : MapReduce across distributed data centers for data-intensive computing. *Future Generation Computer Systems*, 29(3) :739–750, 2013.
- [69] Meryem Amame, Karima Aissaoui, and Mohammed Berrada. Big data in e-learning : Literature review and challenges. *Advanced Intelligent Systems for Sustainable Development (AI2SD’2020) Volume 1*, pages 89–102, 2022.
- [70] Jeffrey Dean and Sanjay Ghemawat. Mapreduce : a flexible data processing tool. *Communications of the ACM*, 53(1) :72–77, 2010.
- [71] Salman Salloum, Ruslan Dautov, Xiaojun Chen, Patrick Xiaogang Peng, and Joshua Zhexue Huang. Big data analytics on apache spark. *International Journal of Data Science and Analytics*, 1 :145–164, 2016.

-
- [72] Alexander Shkapsky, Mohan Yang, Matteo Interlandi, Hsuan Chiu, Tyson Condie, and Carlo Zaniolo. Big data analytics with datalog queries on spark. pages 1135–1149, 2016.
- [73] Vinod Kumar Vavilapalli, Arun C Murthy, Chris Douglas, Sharad Agarwal, Mahadev Konar, Robert Evans, Thomas Graves, Jason Lowe, Hitesh Shah, and Siddharth Seth. Apache hadoop yarn : Yet another resource negotiator. pages 1–16, 2013.
- [74] Aniket Wakde, Purvesh Shende, Sudarshan Waydande, Shravani Uttarwar, and Ganesh Deshmukh. Comparative analysis of hadoop tools and spark technology. pages 1–4, 2018.
- [75] Liang-qiu Meng. Application of big data in higher education. In *Proceedings of the 2nd International Conference on Teaching and Computational Science*, pages 215–216. Atlantis Press, 2014.
- [76] Ben Daniel and Russell Butson. Foundations of big data and analytics in higher education. pages 39–47, 2014.
- [77] Tristan Denley. How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment*, 9 :61–69, 2014.
- [78] Katrina Sin and Loganathan Muthu. Application of big data in education data mining and learning analytics—a literature review. *ICTACT journal on soft computing*, 5(4) :1035–1049, 2015.
- [79] Shakeel Ahmed and Hemant Kumar Mehta. On applying big data and cloud computing for quality improvement in higher education. *Proc. of Int’l Cont. on Advances in Big Data Analytics. Np, nd Web*, 2015.
- [80] Ahmad Ashraf, Hazem El-Bakry, M Samir Abd El-razek, and Yehia El-Mashad. Handling big data in e-learning. *International Journal of Advanced Research in Computer Science & Technology*, 3(1) :47–51, 2015.
- [81] Ben Daniel. Big data and analytics in higher education : Opportunities and challenges. *British journal of educational technology*, 46(5) :904–920, 2015.

-
- [82] Zijiao Tang. On study of application of big data and cloud computing technology in smart campus. volume 100, page 012026. IOP Publishing, 2017. Issue : 1.
 - [83] Marouane Birjali, Abderrahim Beni-Hssane, and Mohammed Erritali. A novel adaptive e-learning model based on big data by using competence-based knowledge and social learner activities. *Applied Soft Computing*, 69 :14–32, 2018.
 - [84] Yi Li and Xiaoning Zhai. Review and prospect of modern education using big data. *Procedia Computer Science*, 129 :341–347, 2018.
 - [85] Mohammad Shorfuzzaman, M Shamim Hossain, Amril Nazir, Ghulam Muhammad, and Atif Alamri. Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Computers in Human behavior*, 92 :578–588, 2019.
 - [86] Aleksandra Nikolovska, Aleksandar Velinov, Stojance Spasov, and Zoran Zdravev. Framework for big data analytics of moodle data using hadoop in the cloud. 2018.
 - [87] Ruth Chweya, Samuel-Soma M Ajibade, Abba Kyari Buba, and Moveh Samuel. IoT and big data technologies : opportunities and challenges for higher learning. *International Journal of Recent Technology and Engineering (IJRTE)*, 9 (2), 909, 913, 2020.
 - [88] Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3) :56–58, 1997. Publisher : ACM New York, NY, USA.
 - [89] Robin Burke, Alexander Felfernig, and Mehmet H Göker. Recommender systems : An overview. *Ai Magazine*, 32(3) :13–18, 2011.
 - [90] Nicholas J. Belkin and W. Bruce Croft. Information filtering and information retrieval : Two sides of the same coin ? *Commun. ACM*, 35(12) :29–38, 1992. Place : New York, NY, USA Publisher : Association for Computing Machinery.
 - [91] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM*,

-
- 35(12) :61–70, 1992. Place : New York, NY, USA Publisher : Association for Computing Machinery.
- [92] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens : An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94*, pages 175–186. Association for Computing Machinery, 1994. event-place : Chapel Hill, North Carolina, USA.
- [93] Upendra Shardanand and Pattie Maes. Social information filtering : Algorithms for automating “word of mouth”. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '95*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995. event-place : Denver, Colorado, USA.
- [94] Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '95*, pages 194–201. ACM Press/Addison-Wesley Publishing Co., 1995. event-place : Denver, Colorado, USA.
- [95] F. Maxwell Harper and Joseph A. Konstan. The MovieLens datasets : History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), 2015. Place : New York, NY, USA Publisher : Association for Computing Machinery.
- [96] Hyeyoung Ko, Suyeon Lee, Yoonseo Park, and Anna Choi. A survey of recommendation systems : recommendation models, techniques, and application fields. *Electronics*, 11(1) :141, 2022.
- [97] Bam Bahadur Sinha and R Dhanalakshmi. Evolution of recommender paradigm optimization over time. *Journal of King Saud University-Computer and Information Sciences*, 34(4) :1047–1059, 2022.
- [98] s. M. Mahdi Seyednezhad, Kailey Cozart, John Bowllan, and Anthony Smith. *A Review on Recommendation Systems : Context-aware to Social-based*. 2018.
- [99] Negre Elsa. Les systèmes de recommandation : une catégorisation. 2018.

-
- [100] Idir Benouaret. Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels, 2017.
- [101] Sujoy Bag, Sri Krishna Kumar, and Manoj Kumar Tiwari. An efficient recommendation generation using relevant jaccard similarity. *Information Sciences*, 483 :53–64, 2019.
- [102] Ailin Zhu, Zexi Hua, Yu Shi, Yongchuan Tang, and Lingwei Miao. An improved k-means algorithm based on evidence distance. *Entropy*, 23(11) :1550, 2021.
- [103] G. Linden, B. Smith, and J. York. Amazon.com recommendations : item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1) :76–80, 2003.
- [104] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8) :30–37, 2009.
- [105] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys ’16, page 191–198, New York, NY, USA, 2016. Association for Computing Machinery.
- [106] Casper Hansen, Christian Hansen, Lucas Maystre, Rishabh Mehrotra, Brian Brost, Federico Tomasi, and Mounia Lalmas. Contextual and sequential user embeddings for large-scale music recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems*, pages 53–62, 2020.
- [107] Yilena Pérez-Almaguer, Raciél Yera, Ahmad A Alzahrani, and Luis Martínez. Content-based group recommender systems : A general taxonomy and further improvements. *Expert Systems with Applications*, 184 :115444, 2021.
- [108] Ruisheng Zhang, Qi-dong Liu, and Jia-Xuan Wei. Collaborative filtering for recommender systems. pages 301–308, 2014.
- [109] Lamis Al Hassanieh, Chadi Abou Jaoudeh, Jacques Bou Abdo, and Jacques Demerjian. Similarity measures for collaborative filtering recommender systems. pages 1–5, 2018.
- [110] Charu C Aggarwal and Charu C Aggarwal. Model-based collaborative filtering. *Recommender systems : the textbook*, pages 71–138, 2016.

-
- [111] Folasade Isinkaye, Yetunde Folajimi, and Bolanle Ojokoh. Recommendation systems : Principles, methods and evaluation. *Egyptian Informatics Journal*, 16, 2015.
 - [112] Robin Burke and Robin. *Hybrid Web Recommender Systems*, volume 4321. 2007-01-01. Journal Abbreviation : LNCS Publication Title : LNCS.
 - [113] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8) :30–37, 2009.
 - [114] Ya Zhang, Weiyuan Chen, and Zibin Yin. Collaborative filtering with social regularization for TV program recommendation. *Knowledge-Based Systems*, 54 :310–317, 2013-12-01.
 - [115] John K Tarus, Zhendong Niu, and Ghulam Mustafa. Knowledge-based recommendation : a review of ontology-based recommender systems for e-learning. *Artificial intelligence review*, 50 :21–48, 2018.
 - [116] Housseem Chemingui. Configuration guidée des lignes de produits basée sur les traces de processus, 2021-06-30.
 - [117] A. Jorge Morais, Eugénio Oliveira, and Alípio Mário Jorge. A multi-agent recommender system. In Sigeru Omatu, Juan F. De Paz Santana, Sara Rodríguez González, Jose M. Molina, Ana M. Bernardos, and Juan M. Corchado Rodríguez, editors, *Distributed Computing and Artificial Intelligence*, pages 281–288. Springer Berlin Heidelberg, 2012.
 - [118] Fatma Siala. A multi-agent recommender system using social networks. In *Proceedings of the 1st International Conference on Digital Tools & Uses Congress*, DTUC '18. Association for Computing Machinery, 2018. event-place : Paris, France.
 - [119] Punam Bedi, Sumit Kumar Agarwal, Vinita Jindal, and Richa. MARST : Multi-agent recommender system for e-tourism using reputation based collaborative filtering. In Aastha Madaan, Shinji Kikuchi, and Subhash Bhalla, editors, *Databases in Networked Information Systems*, pages 189–201. Springer International Publishing, 2014.

-
- [120] B. Twardowski and D. Ryzko. IoT and context-aware mobile recommendations using multi-agent systems. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 1, pages 33–40, 2015. Journal Abbreviation : 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT).
- [121] L. Yinggang and T. Xiangrong. Social recommendation system based on multi-agent deep reinforcement learning. In *2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS)*, pages 371–377, 2022. Journal Abbreviation : 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS).
- [122] Meryem Amame, Karima Aissaoui, and Mohammed Berrada. A multi-agent and content-based course recommender system for university e-learning platforms. In Saad Motahhir and Badre Bossoufi, editors, *Digital Technologies and Applications*, pages 663–672. Springer International Publishing, 2021.
- [123] Dhomas Hatta Fudholi, Wenny Rahayu, and Eric Pardede. A data-driven approach toward building dynamic ontology. pages 223–232. Springer, 2013.
- [124] Meryem Amame, Karima Aissaoui, and Mohammed Berrada. ERSDO : E-learning recommender system based on dynamic ontology. *Education and Information Technologies*, 27(6) :7549–7561, 2022.
- [125] Sarma Cakula and Maija Sedleniece. Development of a personalized e-learning model using methods of ontology. *Procedia Computer Science*, 26 :113–120, 2013.
- [126] Pragya Dwivedi and Kamal K Bharadwaj. Effective trust-aware e-learning recommender system based on learning styles and knowledge levels. *Journal of Educational Technology & Society*, 16(4) :201–216, 2013.
- [127] Michael Raschick, Donald E Maypole, and Priscilla A Day. Improving field education through kolb learning theory. *Journal of Social Work Education*, 34(1) :31–42, 1998.

-
- [128] Tim Berners-Lee and James Hendler. Publishing on the semantic web. *Nature*, 410(6832) :1023–1024, 2001. Publisher : Nature Publishing Group UK London.
- [129] Aaron Swartz and James Hendler. The semantic web : A network of content for the digital city. 2001.
- [130] Steve Bratt. Semantic web and other w3c technologies to watch. *Talks at W3C, January*, 2007.
- [131] Carlo Torniai, Jelena Jovanovic, Dragan Gašević, Scott Bateman, and Marek Hatala. E-learning meets the social semantic web. pages 389–393, 2008.
- [132] Jean-Pierre Cahier, L'Hédi Zaher, and Manuel Zacklad. Information seeking in a " socio-semantic web " application. pages 91–95, 2007.
- [133] Tom Gruber. Collective knowledge systems : Where the social web meets the semantic web. *Journal of web semantics*, 6(1) :4–13, 2008.
- [134] Mikael Nilsson and Matthias Palmér. Semantic web meta-data for e-learning : Some architectural guidelines. 2002.
- [135] Nicola Guarino and Pierdaniele Giaretta. Ontologies and knowledge bases. *Towards very large knowledge bases*, pages 1–2, 1995. Publisher : IOS press Clifton, VA, USA.
- [136] Tom Gruber. What is an ontology. 1993.
- [137] Ian Niles and Adam Pease. Towards a standard upper ontology. pages 2–9, 2001.
- [138] Cassia Trojahn, Renata Vieira, Daniela Schmidt, Adam Pease, and Giancarlo Guizzardi. Foundational ontologies meet ontology matching : A survey. *Semantic Web*, 13(4) :685–704, 2022.
- [139] Deborah L McGuinness and Frank Van Harmelen. OWL web ontology language overview. *W3C recommendation*, 10(10) :2004, 2004.
- [140] Michael R Genesereth and Richard E Fikes. Knowledge interchange format-version 3.0 : reference manual. 1992.

-
- [141] Mohammed Essmat Ibrahim, Yanyan Yang, and David Ndzi. Using ontology for personalised course recommendation applications. pages 426–438. Springer, 2017.
- [142] Mohammed E Ibrahim, Yanyan Yang, David L Ndzi, Guangguang Yang, and Murtadha Al-Maliki. Ontology-based personalized course recommendation framework. *IEEE Access*, 7 :5180–5199, 2018.
- [143] Bouchra Bouihi and Mohamed Bahaj. Ontology and rule-based recommender system for e-learning applications. *International Journal of Emerging Technologies in Learning (Online)*, 14(15) :4, 2019. Publisher : International Association of Online Engineering (IAOE).
- [144] Kahina Rabahallah, Latifa Mahdaoui, and Faïçal Azouaou. MOOCs recommender system using ontology and memory-based collaborative filtering. pages 635–641, 2018.
- [145] Oluwatoyin Catherine Agbonifo and Motunrayo Akinsete. Development of an ontology-based personalised e-learning recommender system. *International Journal of Computer (IJC)*, 38(1) :102–112, 2020.
- [146] Sergio Miranda, Francesco Orciuoli, and Demetrios G Sampson. A SKOS-based framework for subject ontologies to improve learning experiences. *Computers in Human Behavior*, 61 :609–621, 2016.
- [147] Silvia Rita Viola, Sabine Graf, and Tommaso Leo. Analysis of felder-silverman index of learning styles by a data-driven statistical approach. pages 959–964, 2006.
- [148] Khairil Imran Ghauth and Nor Aniza Abdullah. Learning materials recommendation using good learners’ ratings and content-based filtering. *Educational technology research and development*, 58 :711–727, 2010.
- [149] M. Alghobiri. A comparative analysis of classification algorithms on diverse datasets. *Engineering, Technology & Applied Science Research*, 8(2) :2790–2795, Apr. 2018.
- [150] Yueqing Yin and Lei Fan. Trends of open educational resources in higher education. pages 146–156. Springer, 2011.

-
- [151] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification : An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3) :1–13, 2007.
- [152] Ouafae El Aissaoui, Yasser El Alami El Madani, Lahcen Oughdir, and Youssef El Alloui. A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. *Education and Information Technologies*, 24 :1943–1959, 2019.
- [153] Rory McGreal. *Online education using learning objects*. Psychology Press, 2004.
- [154] Susan E Metros and Kathleen Bennett. Learning objects in higher education. 2002.
- [155] Norm Friesen. Three objections to learning objects. In *Online education using learning objects*, pages 78–88. Routledge, 2004.
- [156] H Wayne Hodgins. The future of learning objects. *Educational Technology*, pages 49–54, 2006.
- [157] Rabiman Rabiman, Muhammad Nurtanto, and Nur Kholifah. Design and development e-learning system by learning management system (LMS) in vocational education. *Online Submission*, 9(1) :1059–1063, 2020.
- [158] Peter Baumgartner and Marco Kalz. *Wiederverwendung von Lernobjekten aus didaktischer Sicht*. 2005.
- [159] Xavier Ochoa, Joris Klerkx, Bram Vandeputte, and Erik Duval. On the use of learning object metadata : The GLOBE experience. pages 271–284. Springer, 2011.
- [160] André C. P. L. F. de Carvalho and Alex A. Freitas. A tutorial on multi-label classification techniques. In Ajith Abraham, Aboul-Ella Hassanien, and Václav Snášel, editors, *Foundations of Computational Intelligence Volume 5 : Function Approximation and Classification*, pages 177–195. Springer Berlin Heidelberg, 2009.
- [161] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine learning*, 85 :333–359, 2011.

-
- [162] Zhi-Hua Zhou. *Ensemble methods : foundations and algorithms*. CRC press, 2012.
- [163] Mahdi Hashemzadeh, Amin Golzari Oskouei, and Nacer Farajzadeh. New fuzzy c-means clustering method based on feature-weight and cluster-weight learning. *Applied Soft Computing*, 78 :324–345, 2019.
- [164] Yi Ding and Xian Fu. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing*, 188 :233–238, 2016.
- [165] R Suganya and R Shanthi. Fuzzy c-means algorithm-a review. *International Journal of Scientific and Research Publications*, 2(11) :1, 2012.
- [166] Jelili Oyelade, Itunuoluwa Isewon, Olufunke Oladipupo, Onyeka Emebo, Zacchaeus Omogbadegun, Olufemi Aromolaran, Efosa Uwoghiren, Damilare Olaniyan, and Obembe Olawole. Data clustering : Algorithms and its applications. pages 71–81, 2019.
- [167] Reena Hooda and Nasib S Gill. Evolution of data warehouse and data mining. *International Journal of Managment, IT and Engineering*, 2(5) :96–104, 2012. Publisher : International Journals of Multidisciplinary Research Academy.
- [168] S Barry Cooper and Jan Van Leeuwen. *Alan Turing : His work and impact*. Elsevier, 2013.
- [169] Krzysztof J Cios, Witold Pedrycz, and Roman W Swiniarski. *Data mining methods for knowledge discovery*, volume 458. Springer Science & Business Media, 2012.
- [170] Gonzalo Mariscal, Oscar Marban, and Covadonga Fernandez. A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2) :137–166, 2010. Publisher : Cambridge University Press.
- [171] Pariwat Ongsulee. Artificial intelligence, machine learning and deep learning. pages 1–6, 2017.

-
- [172] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. *Machine learning techniques for multimedia : case studies on organization and retrieval*, pages 21–49, 2008.
- [173] Pratap Chandra Sen, Mahimarnab Hajra, and Mitadru Ghosh. Supervised classification algorithms in machine learning : A survey and review. pages 99–111. Springer, 2020.
- [174] M Emre Celebi and Kemal Aydin. *Unsupervised learning algorithms*, volume 9. Springer, 2016.
- [175] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. Deep reinforcement learning : an overview. pages 426–440. Springer, 2018.
- [176] Hongming Zhang and Tianyang Yu. Taxonomy of reinforcement learning algorithms. *Deep Reinforcement Learning : Fundamentals, Research and Applications*, pages 125–133, 2020.
- [177] Geoffrey I Webb, Eamonn Keogh, and Risto Miikkulainen. Naïve bayes. *Encyclopedia of machine learning*, 15 :713–714, 2010.
- [178] Jeevamol Joy and Renumol Vemballivel Govinda Pillai. Review and classification of content recommenders in e-learning environment. *Journal of King Saud University-Computer and Information Sciences*, 34(9) :7670–7685, 2022.
- [179] Vivian F López Batista, Fernando Prieta Pintado, Ana Belén Gil, Sara Rodríguez, and María N Moreno. A system for multi-label classification of learning objects. pages 523–531. Springer, 2011.
- [180] Vivian F López, Fernando De La Prieta, Mitsunori Ogihara, and Ding Ding Wong. A model for multi-label classification and ranking of learning objects. *Expert Systems with Applications*, 39(10) :8878–8884, 2012.
- [181] X Zhao, W Cui, and T Dong. An effective learning object classification based on long-term memory and random forest classification. *iJET*, 14(20) :107–119, 2019.
- [182] Carlos N Silla and Alex A Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22 :31–72, 2011.

-
- [183] Pedro González, Eva Gibaja, Alfredo Zapata, Víctor H Menéndez, and Cristóbal Romero. Towards automatic classification of learning objects : Reducing the number of used features. pages 394–395, 2017.
- [184] Dolly Carrillo, Vivian F López, and María N Moreno. Multi-label classification for recommender systems. pages 181–188. Springer, 2013.
- [185] Ilze Birzniece, Peteris Rudzajs, Diana Kalibatiene, Olegas Vasilecas, and Edgars Rencis. Application of interactive classification system in university study course comparison. *Informatics in Education*, 14(1) :13–34, 2015.
- [186] Asma Aldrees and Azeddine Chikh. Comparative evaluation of four multi-label classification algorithms in classifying learning objects. *Computer Applications in Engineering Education*, 24(4) :651–660, 2016.
- [187] R Selvan and K Kulothungan. Multi-label classification of learning objects using deep learning techniques. *International Journal of Computer Science and Mobile Computing*, 8(6) :10–21, 2019.
- [188] M Rani and V Kumar. Multi-label classification of learning objects using machine learning algorithms. *International Journal of Intelligent Systems and Applications*, 12(7) :36–43, 2020.
- [189] Meryem Amane, Karima Aissaoui, and Mohammed Berrada. Multi-label classification of learning objects using clustering algorithms based on feature selection. *International Journal of Emerging Technologies in Learning*, 17(20), 2022.
- [190] Alvin S Sicat and MA Ed. Enhancing college students’ proficiency in business writing via schoology. *International Journal of Education and Research*, 3(1) :159–178, 2015.

